Methods in Molecular Biology 1303

Springer Protocols

Juan I. Castrillo
Stephen G. Oliver *Editors*

Systems Biology of Alzheimer's Disease

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Systems Biology of Alzheimer's Disease

Edited by

Juan I. Castrillo

Department of Biochemistry & Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK; Genetadi Biotech S.L. Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain

Stephen G. Oliver

Department of Biochemistry & Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK

Humana Press

Editors
Juan I. Castrillo
Department of Biochemistry &
Cambridge Systems Biology Centre
University of Cambridge
Cambridge, UK

Genetadi Biotech S.L. Parque
Tecnológico de Bizkaia
Derio, Bizkaia, Spain

Stephen G. Oliver
Department of Biochemistry &
Cambridge Systems Biology Centre
University of Cambridge
Cambridge, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-2626-8 ISBN 978-1-4939-2627-5 (eBook)
DOI 10.1007/978-1-4939-2627-5

Library of Congress Control Number: 2015943618

Springer New York Heidelberg Dordrecht London

Cover illustration: Photomicrograph of stained sections, neurons and dendrites in the parahippocampal cortex of a patient whose soma contains anti-paired helical filaments (PHF-tau$_{AT8}$) in the putative pre-tangle state. Image reproduced from (Merino-Serrais et al (2013) *Brain* 136, 1913). Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License.

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

Alzheimer's disease (AD), the most common form of dementia, currently affects over 44 million individuals worldwide; its incidence is set to almost double by 2030 and more than triple by 2050. The global cost of dementia was estimated in 2010 at USD 604 billion and is likely to rise. Strong support exists toward progress on risk reduction via protective and modifiable factors, early diagnosis, and timely intervention (Prince et al. 2011; WHO 2012; ADI 2014).

Alzheimer's and many other neurodegenerative disorders are multifactorial in nature, involving a combination of genomic, epigenomic, networks dynamic, and environmental factors. The interplay of disease mechanisms and homeostatic biological networks will underlie the time of onset and rate of progression of the disease, with a cascade of downstream effects resulting in a range of patient-specific phenotypes. The proper investigation of the complexity of the disease requires new integrative Systems Biology approaches, at both the experimental and computational level. The practical goals of such investigation include improved classification of risks and the characterization and detection of the first imbalances that underlie the onset of the disease, in the expectation that early diagnosis will enable tailored and timely interventions.

This book addresses such an integrated approach to Alzheimer's disease and comprises six parts: In *Part I* (Chapters 1 and 2), we present AD as a complex multifactorial disease, with intrinsic susceptibility and network dynamics, for which integrative Systems Biology experimental and computational approaches are necessary. In *Part II* (Chapters 3–10), an up-to-date view of relevant pathways and networks underlying AD is provided. In *Part III* (Chapters 11–16), a number of the main disease models recapitulating AD features, with latest studies and methods, are presented. In *Part IV*, "Experimental Systems Biology" (Chapters 17–25), we present a perspective on next-generation molecular and high-throughput methods for the study of AD susceptibility and the characterization of the pathways and networks underlying the disease. These are applicable not only to AD but also to other multifactorial diseases. In *Part V*, "Computational Systems Biology" (Chapters 26–30), the latest computational and integrative network biology approaches are presented. Finally, in *Part VI*, "AD in Practice. From Systems Biology to Systems Medicine" (Chapters 31–33), selected examples of studies and strategies toward earlier diagnosis and tailored therapeutic intervention are included.

This book is intended for postgraduate students, postdoctoral researchers, and experts in different fields with an interest in comprehensive Systems Biology strategies applicable to AD and other complex multifactorial diseases (including other neurodegenerative diseases and cancers). We aim to present Systems Biology, including both experimental and computational approaches, as a new strategy for the study of AD and other dynamic multifactorial diseases, with the hope and expectation that the results will translate into more effective diagnosis and treatment, and improved public health policies. We expect this book to complement other excellent volumes and monographs on AD that cover fundamental, physiological, or medical aspects of the disease.

Cambridge, UK

Juan I. Castrillo
Stephen G. Oliver

References

Prince M, Bryce R, Ferri C (2011) World Alzheimer Report 2011. The benefits of early diagnosis and intervention. Alzheimer's disease International (ADI) 2011. (http://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf)

World Health Organization (2012) Dementia: a public health priority. World Health Organization and Alzheimer's Disease International, Geneva. (http://www.who.int/mental_health/publications/dementia_report_2012/en/)

Alzheimer's Disease International (2014) World Alzheimer Report 2014: dementia and risk reduction (an analysis of protective and modifiable factors) (http://www.alz.co.uk/research/world-report-2014)

Contents

Contributors

Jesús Avila • *Centro de Biología Molecular Severo Ochoa CSIC-UAM, Universidad Autónoma de Madrid, Madrid, Spain; CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain*

Ziv Bar-Joseph • *Lane Center for Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*

Dan Bean • *Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK*

Ruth Benavides-Piccione • *CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain; Laboratorio Cajal de Circuitos Corticales (CTB), UPM, Campus Montegancedo and Instituto Cajal (CSIC), Madrid, Spain*

Bruno A. Benitez • *Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA*

Lars Bertram • *Platform for Genome Analytics, Institutes of Neurogenetics and Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany*

Anita Bhattacharyya • *Waisman Center, University of Wisconsin-Madison, Madison, WI, USA*

George F. Bonadurer III • *Waisman Center, University of Wisconsin-Madison, Madison, WI, USA*

Jeff Van den Brande • *Laboratory of Functional Biology, Yeast Biotechnology Group, KU Leuven, Heverlee, Belgium*

Carol Brayne • *Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK*

D. Allan Butterfield • *Department of Chemistry, Center of Membrane Sciences and Sanders Brown Center on Aging, University of Kentucky, Lexington, KY, USA*

Tine Bynens • *Laboratory of Functional Biology, Yeast Biotechnology Group, KU Leuven, Heverlee, Belgium*

Juan I. Castrillo • *Department of Biochemistry & Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK; Genetadi Biotech S.L. Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain*

Bei Jun Chen • *School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW, Australia*

Damian C. Crowther • *Neuroscience Innovative Medicines, MedImmune Limited, Cambridge, UK*

Carlos Cruchaga • *Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA*

Mar Cuadrado-Tejedor • *Cell and Molecular Neuropharmacology, Neurosciences Division, Center for Applied Medical Research, CIMA, University of Navarra, Pamplona, Spain; Department of Anatomy, School of Medicine, University of Navarra, Pamplona, Spain*

Raquel Cuadros • *Centro de Biología Molecular Severo Ochoa CSIC-UAM, Universidad Autónoma de Madrid, Madrid, Spain; CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain*
Christophe Cullin • *Institut de Biochimie et Génétique Cellulaires, CNRS UMR 5095, Université Bordeaux 2, Bordeaux, France*
Javier DeFelipe • *CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain; Laboratorio Cajal de Circuitos Corticales (CTB), UPM, Campus Montegancedo and Instituto Cajal (CSIC), Madrid, Spain*
Yuetiva Deming • *Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA*
Fabio Di Domenico • *Department of Biochemical Sciences, Sapienza University of Rome, Rome, Italy*
Bruno Dubois • *Département de Neurologie, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A), Pavillon François Lhermitte, Hôpital de la Salpêtrière, Université Pierre et Marie Curie, Paris, France*
Anne Eckert • *Neurobiology Laboratory for Brain Aging and Mental Health, Psychiatric University Clinics, University of Basel, Transfaculty Research Platform, Molecular & Cognitive Neurosciences, Basel, Switzerland*
Valur Emilsson • *Icelandic Heart Association, Kopavogur, Iceland*
Genevieve Evin • *Department of Pathology, The University of Melbourne, Parkville, VIC, Australia; Oxidation Biology Laboratory, Mental Health Research Institute, Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia*
Eugenio Fava • *German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
Giorgio Favrin • *Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK*
Gang Feng • *Biomedical Informatics Center, Northwestern University Clinical and Translational Sciences Institute, Chicago, IL, USA*
Vanessa Franssens • *Laboratory of Functional Biology, Yeast Biotechnology Group, KU Leuven, Heverlee, Belgium*
Francesco Garaci • *Department of Diagnostic Imaging, Molecular Imaging, Interventional Radiology, and Radiotherapy, University of Rome "Tor Vergata", Rome, Italy; IRCCS San Raffaele Pisana, Rome and San Raffaele Cassino, Cassino, Italy*
Ana García-Osta • *Cell and Molecular Neuropharmacology, Neurosciences Division, Center for Applied Medical Research, CIMA, University of Navarra, Pamplona, Spain*
Anthony Gitter • *Department of Biological Engineering, Microsoft Research, Cambridge, MA, USA; Massachusetts Institute of Technology, Cambridge, MA, USA*
Elena Gómez • *Departamento de Neuropatología y Banco de Tejidos, Fundación CIEN, Madrid, Spain*
Diego Gomez-Nicola • *CNS Inflammation Group, Centre for Biological Sciences, Southampton General Hospital, University of Southampton, Southampton, UK*
Amandine Grimm • *Neurobiology Laboratory for Brain Aging and Mental Health, Psychiatric University Clinics, University of Basel, Transfaculty Research Platform, Molecular & Cognitive Neurosciences, Basel, Switzerland*
Liang-Hao Guo • *Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany*

Harald Hampel • *Département de Neurologie, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A), Pavillon François Lhermitte, Hôpital de la Salpêtrière, Université Pierre et Marie Curie, Paris, France*

Xianlin Han • *Diabetes and Obesity Research Center, Sanford-Burnham Medical Research Institute|, Orlando, FL, USA*

Dominic J. Hare • *The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia; Elemental Bio-Imaging Facility, University of Technology Sydney, Broadway, NSW, Australia*

David Hecker • *German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*

Sally Hunter • *Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK*

Michael Janitz • *School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW, Australia*

Caroline Janitz • *University of Western Sydney, Penrith, NSW, Australia*

Sheng Chih Jin • *Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA*

Clemens F. Kaminski • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*

Gabriele S. Kaminski Schierle • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*

Huining Kang • *Department of Neurosciences, University of New Mexico, Albuquerque, NM, USA*

Zaven S. Khachaturian • *The Campaign to Prevent Alzheimer's Disease by 2020 (PAD2020), Potomac, MD, USA*

Masataka Kikuchi • *Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; Research Association for Biotechnology, Tokyo, Japan*

Ryozo Kuwano • *Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan*

Qiao-Xin Li • *Department of Pathology, The University of Melbourne, Parkville, VIC, Australia; National Dementia Diagnostics Laboratory, Mental Health Research Institute, Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia*

Jung Yeon Lim • *Department of Genetics, University of Cambridge, Cambridge, UK*

Simon Lin • *Biomedical Informatics Center, Northwestern University Clinical and Translational Sciences Institute, Chicago, IL, USA*

Simone Lista • *Department of Psychiatry, Psychotherapy and Psychosomatics, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany*

Frederick J. Livesey • *Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK; Gurdon Institute, University of Cambridge, Cambridge, UK*

Ian Macreadie • *School of Applied Sciences and Health Innovations Research Institute, RMIT University, Melbourne, VIC, Australia*

Steven Martin • *Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK*

Joseph C. Masdeu • *Nantz National Alzheimer Center and Neuroimaging, Houston Methodist Neurological Institute, Weill Medical College of Cornell University, Houston, TX, USA; Section on Integrative Neuroimaging, Clinical Brain Disorders Branch, NIMH, National Institutes of Health, Bethesda, MD, USA*
Miguel Medina • *CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain*
Hongkang Mei • *Informatics and Structure Biology, R&D China, GlaxoSmithKline, Shanghai, China*
Fiona M. Menzies • *Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK*
Paula Merino-Serráis • *CIBERNED (Centro de Investigación en Red sobre Enfermedades Neurodegenerativas), Madrid, Spain; Laboratorio Cajal de Circuitos Corticales (CTB), UPM, Campus Montegancedo and Instituto Cajal (CSIC), Madrid, Spain*
James D. Mills • *School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW, Australia*
Akinori Miyashita • *Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan*
Satoshi Mizuno • *Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan*
Steven Moore • *Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK; Gurdon Institute, University of Cambridge, Cambridge, UK*
Jun Nakaya • *Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan*
Soichi Ogishima • *Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan*
Stephen G. Oliver • *Department of Biochemistry & Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK*
Stanislav Ott • *Department of Genetics, University of Cambridge, Cambridge, UK*
Marzia Perluigi • *Department of Biochemical Sciences, Sapienza University of Rome, Rome, Italy*
Robert Perneczky • *Neuroepidemiology and Ageing Research Unit, Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London, UK; West London Cognitive Disorders Treatment and Research Unit, West London Mental Health Trust, London, UK; Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany*
V. Hugh Perry • *CNS Inflammation Group, Centre for Biological Sciences, Southampton General Hospital, University of Southampton, Southampton, UK*
Dorothea Pinotsi • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
Afsaneh Porzoor • *School of Applied Sciences and Health Innovations Research Institute, RMIT University, Melbourne, VIC, Australia*
Yang Qiu • *Informatics and Structure Biology, R&D China, GlaxoSmithKline, Shanghai, China*
Alberto Rábano • *Departamento de Neuropatología y Banco de Tejidos, Fundación CIEN, Madrid, Spain*
Alan Rembach • *The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia*

Blaine R. Roberts • *The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia*
Izaskun Rodal • *Departamento de Neuropatología y Banco de Tejidos, Fundación CIEN, Madrid, Spain*
Wera Roth • *German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
David C. Rubinsztein • *Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK*
Dan Rujescu • *Department of Psychiatry, Psychotherapy and Psychosomatics, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany*
Nathalie G. Saurat • *Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK; Gurdon Institute, University of Cambridge, Cambridge, UK*
Karen Schmitt • *Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; Department of Biochemistry, University of Cambridge, Cambridge, UK; Gurdon Institute, University of Cambridge, Cambridge, UK*
Jorge Sepulcre • *Division of Nuclear Medicine and Molecular Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; Athinioula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA*
Anne Streeter • *Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK*
Kinga Szigeti • *Department of Neurology, School of Medicine and Biomedical Sciences, University at Buffalo SUNY, Buffalo, NY, USA*
Jiangli Tan • *Department of Pathology, The University of Melbourne, Parkville, VIC, Australia*
Hiroshi Tanaka • *Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan*
Linh Tran • *David Geffen School of Medicine, University of California, Los Angeles, CA, USA*
Mathias Verduyckt • *Laboratory of Functional Biology, Yeast Biotechnology Group, KU Leuven, Heverlee, Belgium*
Hélène Vignaud • *Institut de Biochimie et Génétique Cellulaires, CNRS UMR 5095, Université Bordeaux 2, Bordeaux, France*
Miao Wang • *Diabetes and Obesity Research Center, Sanford-Burnham Medical Research Institute, Orlando, FL, USA*
Yue Wang • *Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China*
Jason P. Weick • *Department of Neurosciences, University of New Mexico, Albuquerque, NM, USA*
Joris Winderickx • *Laboratory of Functional Biology, Yeast Biotechnology Group, KU Leuven, Heverlee, Belgium*
Tian Xia • *Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China*
Justin Yerbury • *Illawarra Health and Medical Research Institute, School of Biological Sciences, University of Wollongong, Wollongong, NSW, Australia*
Andreas Zanzoni • *Laboratoire TAGC/INSERM UMR_S1090, Parc Scientifique de Luminy, Marseille cedex 9, France; Aix-Marseille Université, UMR_S1090 TAGC, Marseille, France*

Bin Zhang • *Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, USA; Institute of Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, NY, USA*

Jason Zhu • *Informatics and Structure Biology, R&D China, GlaxoSmithKline, Shanghai, China*

Jun Zhu • *Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, USA; Institute of Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, NY, USA*

Part I

Systems Biology of Multifactorial Diseases: Alzheimer's Disease

Chapter 1

Alzheimer's as a Systems-Level Disease Involving the Interplay of Multiple Cellular Networks

Juan I. Castrillo and Stephen G. Oliver

Abstract

Alzheimer's disease (AD), and many neurodegenerative disorders, are multifactorial in nature. They involve a combination of genomic, epigenomic, interactomic and environmental factors. Progress is being made, and these complex diseases are beginning to be understood as having their origin in altered states of biological networks at the cellular level. In the case of AD, genomic susceptibility and mechanisms leading to (or accompanying) the impairment of the central Amyloid Precursor Protein (APP) processing and tau networks are widely accepted as major contributors to the diseased state. The derangement of these networks may result in both the gain and loss of functions, increased generation of toxic species (e.g., toxic soluble oligomers and aggregates) and imbalances, whose effects can propagate to supra-cellular levels. Although well sustained by empirical data and widely accepted, this global perspective often overlooks the essential roles played by the main counteracting homeostatic networks (e.g., protein quality control/proteostasis, unfolded protein response, protein folding chaperone networks, disaggregases, ER-associated degradation/ubiquitin proteasome system, endolysosomal network, autophagy, and other stress-protective and clearance networks), whose relevance to AD is just beginning to be fully realized. In this chapter, an integrative perspective is presented. Alzheimer's disease is characterized to be a result of: (a) intrinsic genomic/epigenomic susceptibility and, (b) a continued dynamic interplay between the deranged networks and the central homeostatic networks of nerve cells. This interplay of networks will underlie both the onset and rate of progression of the disease in each individual. Integrative Systems Biology approaches are required to effect its elucidation. Comprehensive Systems Biology experiments at different 'omics levels in simple model organisms, engineered to recapitulate the basic features of AD may illuminate the onset and sequence of events underlying AD. Indeed, studies of models of AD in simple organisms, differentiated cells in culture and rodents are beginning to offer hope that the onset and progression of AD, if detected at an early stage, may be stopped, delayed, or even reversed, by activating or modulating networks involved in proteostasis and the clearance of toxic species. In practice, the incorporation of next-generation neuroimaging, high-throughput and computational approaches are opening the way towards early diagnosis well before irreversible cell death. Thus, the presence or co-occurrence of: (a) accumulation of toxic Aβ oligomers and tau species; (b) altered splicing and transcriptome patterns; (c) impaired redox, proteostatic, and metabolic networks together with, (d) compromised homeostatic capacities may constitute relevant '*AD hallmarks at the cellular level*' towards reliable and early diagnosis. From here, preventive lifestyle changes and tailored therapies may be investigated, such as *combined strategies* aimed at *both lowering the production of toxic species and potentiating homeostatic responses*, in order to prevent or delay the onset, and arrest, alleviate, or even reverse the progression of the disease.

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_1,

Key words Alzheimer's disease, Omics, Interactomes, Networks, APP, Amyloid-β (Aβ), Tau, Oligomer cascade, Amyloid cascade, Proteinopathy, Protein folding disease, Proteostasis, Homeostatic networks, Experimental systems biology, Next generation post-genomic techniques, Computational systems biology, Network biology, Disease models, Longitudinal studies, Combined multimodal neuroimaging and molecular biomarkers, Risk classification, Stratification, Standardization, Systems medicine, Public health, Active healthy lifestyle

1 Systems Biology of Multifactorial Diseases: Alzheimer's Disease

In 2003, just a couple of years after the publication of the first draft sequence [1, 2], the scientific community announced the successful completion of the Human Genome Project (Human genome project completion report; April 14, 2003) [3, 4]. Unanimously recognized as one of the greatest milestones in biology, which opened up the prospect of constructing a detailed catalog of human genetic variation (e.g., 1000 Genomes Project Consortium 2010 [5] and [6]), the fact is that we are still far from understanding the exquisite complexity of human biology, and the molecular basis and mechanisms underlying complex diseases, even at the basic cellular level [7–15].

Many complex diseases are being revealed as multifactorial in nature [7, 11, 13, 16–18], involving a combination of genomic, epigenomic, interactomic and environmental factors. While it is clear that the increased availability and efficiency of next generation-sequencing (NGS) technologies [19–22] are already delivering in terms of both the diagnosis and treatment of diseases with a basic genomic component (e.g., Mendelian and, as yet, uncharacterized diseases) [23–34], and will continue to be invaluable in the identification of genomic loci and their specific contribution [16, 33, 35, 36], the challenge of multifactorial diseases is to integrate all components involved and elucidate their interactions. This can only be achieved by the quantitative and holistic approach that Systems Biology offers.

Many complex phenotypes may be more directly related to "*alterations in the properties of systems or networks than to particular genome sequences*" (Vidal in ref. 7). The idea that multi-scale complex systems formed by interacting macromolecules, arranged in dynamic modular complexes and networks underlie some of the most fundamental aspects of life was proposed half a century ago (*see* ref. 37 and references therein). The new era of biology is not so much concerned with building blocks, or inventories of working parts, but, rather, with how those parts interact, and are arranged into functional modules, essential complexes and networks (e.g., DNA-protein; RNA-protein; protein-protein; protein-metabolite networks), in order to produce units of biological organization whose properties are much greater than the sum of their parts. This is what Systems Biology is about—what makes complex

networks and systems sustainable and viable, and how complex diseases can arise from altered networks states [38] that are the consequence of multifactorial perturbations, whose mechanisms and dynamics can perhaps best be studied initially by experiments using model organisms, with the conclusions later confirmed in humans [7, 13, 17, 18, 37, 39–47].

Many complex human diseases (including diabetes, neurodegenerative diseases and most cancers) are far too complex to have a single cause or to rely on genomic variations alone. What really defines these complex diseases is that they are: (1) multifactorial, i.e., with significant contributions from both genomic/epigenomic and environmental perturbations; (2) primarily the result of altered networks, affecting essential modules required for the correct functioning of basic pathways in the cell or organism; (3) fundamentally dynamic, with a fine balance between these impaired networks and homeostatic defense mechanisms [11, 18, 48–55].

We propose that multifactorial complex diseases should be contemplated essentially as shown in Fig. 1. The genome and epigenome underlie the essential networks, first homeostatic states and the basic 'genomic/epigenomic' susceptibility to dysregulation of an organism, which will be subjected to a particular sequence of environmental perturbations (mild or severe; transient or sustained in time) during its lifetime. Mild perturbations may be counteracted by intrinsic (stress) defense networks, such as the heat-shock response, protein homeostasis, inflammatory or immunological networks and others, that restore the 'healthy' state of the individual. However, severe and/or sustained perturbations can overcome these homeostatic defense networks, whose efficiency may well decrease through life [62–64]. This may lead to cascades of dysregulations through intertwined essential networks resulting in acute imbalances, pleiotropic effects and complex diseases. Eventually, the system's capacity to restore homeostasis may be overwhelmed, resulting in an irreversible catastrophic collapse [62, 65–67]. Until we understand complex diseases as altered states of human biological networks, in constant relation with the environment (e.g., external insults, perturbations, traumas, infections, and our own microbiome [61, 68]), with specific dynamics and interplay, our vision will be incomplete [13, 38]. Human individuality at the genomic and metabolomic levels [5, 6, 69] adds an additional layer of complexity, and presents us with an even more challenging picture. Where to start? How can we begin to address most complex human diseases?

Fortunately, for a majority of cases in which the disease first manifests itself mainly at the cellular level (early stage, likely asymptomatic), a good starting point will be to recapitulate the altered states of the systems and networks in simple disease models at the level of the eukaryotic cell. Together with this, since a majority of the essential homeostatic networks (e.g., DNA, RNA, and protein

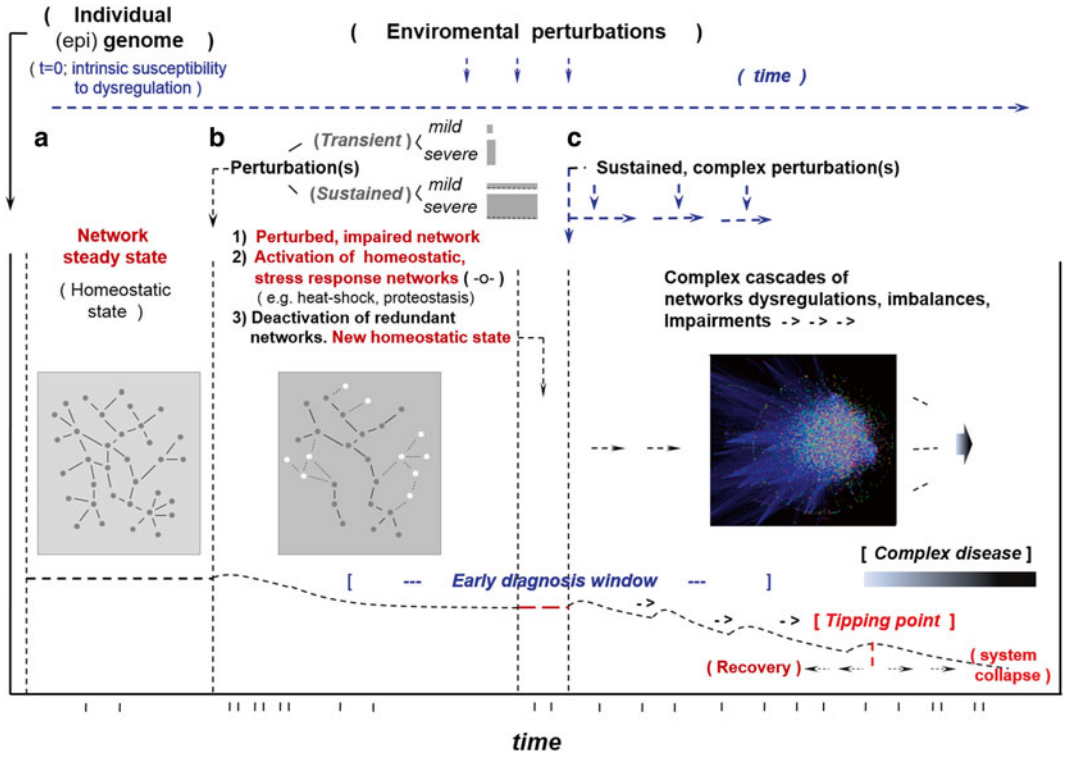


Fig. 1 Dynamics of interplay of biological networks leading to recovery of homeostasis or cascades of dysregulation, acute imbalances and complex diseases. Firstly, the genome and epigenome underlie the essential networks, homeostatic states and initial susceptibility to dysregulation of an individual (**a**) which will be subjected to a specific sequence of environmental perturbations (mild or severe; transient or sustained) during lifetime. Mild perturbations (**b**), result in deactivation of redundant networks (*grey nodes* and *edges*) and activation of homeostatic defense responses (e.g., ER stress, protein homeostasis (proteostasis), immunological and/or inflammatory networks; see new nodes and edges, --o--) until a new homeostatic state is restored. More importantly, severe, complex (e.g., multifactorial) and/or sustained perturbations (**c**) leading to acute imbalances overcoming homeostatic responses may result in cascades of dysregulations, which can propagate through intertwined networks resulting in acute impairments and diseases, with potential irreversible collapse of the whole 'system', at the cellular and/or physiological level. Periodic, longitudinal monitoring at different 'omic levels (e.g., transcriptome, proteome, metabolome and interactomes) towards characterization of the dynamics of impaired and homeostatic networks in molecular and Systems Biology comprehensive experiments in diseases model organisms and human trials (i.e., from yeast to human) [13, 56–60] have the potential to unveil the origin, early stages and dynamics of progression of multifactorial diseases (e.g., neurodegenerative diseases and cancers), well before the tipping point, towards early diagnosis (e.g., characterization of a panel of reliable biomarkers at different 'omic and physiological levels) and timely intervention. Adapted from ref. 13 with permission from Elsevier. Human interactome network picture visualized by Cytoscape 2.5. Human microbiome networks with direct interactions with the human interactome at the interface of health and disease [61] are omitted for clarity. Dataset created by Andrew Garrow at Unilever UK. Author: Keiono, reproduced under GNU Free Documentation License and Creative Commons (CC) licenses (http://en.wikipedia.org/wiki/File:Human_interactome.jpg)

quality control; proteostasis, autophagic mechanisms and clearance pathways) that are of greatest relevance to disease, are essentially conserved in all eukaryotes (e.g., from yeast to human) through millions of years of evolution [13, 17, 18, 41, 62–64, 70–74], well-designed studies with simple organisms may provide invaluable information at this basic cellular level. These are particularly accessible to Systems Biology studies [13, 17, 18, 75]. Confirmation of essentially conserved pathways and networks, and closer-to-human specific mechanisms underlying complex diseases will always require further studies in animal models and, ultimately, in human subjects. The greatest progress will come from integrating this systems-level information (some of which can only be obtained using model organisms, for either practical or ethical reasons) with data from well-designed longitudinal cohort studies.

1.1 Alzheimer's Disease: A Complex Multifactorial Disease

In 1906, a German neurologist, Dr Alois Alzheimer, first described the presence of distinctive pathologic abnormalities in the autopsied brain of a woman who was affected for years by memory problems, confusion, and language dysfunction. He reported the presence of a collection of dense deposits or plaques outside the neurons and bands of fibres or tangles within the brain cells. These senile plaques and neurofibrillary tangles (NFTs) have been recognized to be the two core pathological hallmarks of Alzheimer's disease (AD). Plaques are composed of amyloid beta (Aβ) protein and are called amyloid plaques, and the tangles consist of hyperphosphorylated tau protein. Associated with these changes are increased levels of inflammation, oxidative stress, and nerve cell death [76, 77]. Both senile plaques and neurofibrillary tangles are associated with the progressive loss of neurons and synapses, brain atrophy, and dilatation of the lateral ventricles due to loss of brain tissue, which (together) are the broad features of brain damage in dementia. Cellular dysfunction, tissue and brain changes underlying AD are likely to develop over a period of at least 20–30 years before the onset of symptoms, with the earliest signs appearing around the base of the brain in the fifth decade of life, and plaques and tangles later spreading up to the cortical regions, in a pattern distinct from normal aging and mild cognitive impairment (MCI) (Fig. 2) [78–81].

Alzheimer's disease causes a progressive dementia that currently is estimated to affect over 44 million individuals worldwide, its incidence is set to almost double by 2030 and more than triple by 2050. The global cost of dementia was estimated in 2010 at US $604 billion, and this is only set to rise [77, 82]. Doctors can offer no effective preventive or disease-modifying treatments, and this may be due to our inability to fully understand the underlying mechanisms and to detect the disease at the early asymptomatic stage, well before it has progressed to produce evident memory loss and functional decline ([47, 77]; Table 1 and references

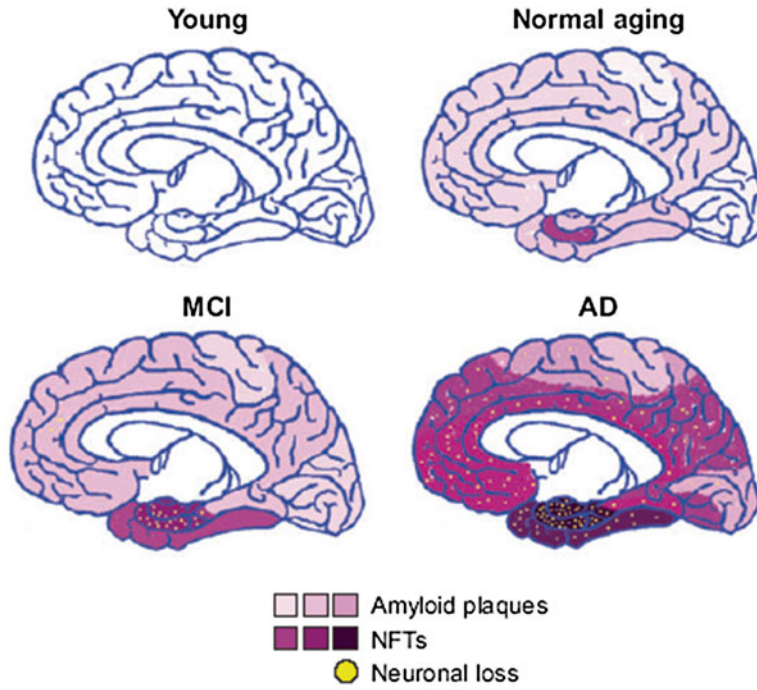


Fig. 2 Progression of neuropathology in aging and Alzheimer's disease. Neuroanatomical distribution of amyloid plaques, neurofibrillary tangles (NFTs), and neuronal loss during normal aging, mild cognitive impairment (MCI), and Alzheimer's disease (AD). In cognitively intact aging individuals, amyloid plaques can appear in the neocortex and hippocampus, whereas NFTs are localized predominantly to the entorhinal cortex. MCI is marked by the appearance of neuronal loss in layer 2 of the entorhinal cortex and the CA1 region of the hippocampus, and is often accompanied by an increase in the number and distribution of plaques and NFTs. Plaques and NFTs are generally more widespread in AD, although this is variable. However, the extent of neuronal and synaptic loss correlates with dementia. Republished with permission of "Annual Reviews", from "The aging brain", Yankner et al., *Annu Rev Pathol* 3 (2008); [79] with permission conveyed through Copyright Clearance Center, Inc

therein [13, 18, 83–299]). Criteria for AD diagnosis, validation and guidelines are, with the help of new knowledge and techniques, being developed and refined [99, 101, 102, 104, 105]; *see* also Table 1.

Although clear histological features and patterns of progression characterize AD [78, 79, 300] (Fig. 2), families and groups of individuals exhibit marked differences and heterogeneity, which are revealing distinct contributions of genomic/epigenomic and environmental factors in different cases. Thus, around <1 % of the AD cases are familial forms of autosomal dominant inheritance, which usually have an onset before age 65. This form of the disease

Table 1
Systems biology of Alzheimer's disease. Selected reference studies and advances (*)

Systems biology of Alzheimer's disease	References
I. Alzheimer's disease	
(1) Fundamentals. State-of-the-art and advances on standard guidelines for clinical diagnosis and stratification (e.g., Institute of Medicine USA, 2012 [99]; ([101, 102]; NIA International Working Group (IWG-2) criteria, 2014 [104]). Genetics, molecular, cellular and physiological mechanisms underlying pathological hallmarks. Protective and modifiable factors. Reference books, monographies, World Health Organization and Alzheimer's reports (e.g., Alzheimer's Association (2014). Alzheimer's disease facts and figures. http://www.alzheimersanddementia.com/article/S1552-5260(14)00062-4/pdf; World Alzheimer Report (2014). http://www.alz.co.uk/research/world-report-2014)	[76–78, 82–111]
(2) Systems Biology of multifactorial neurodegenerative diseases. Alzheimer's disease.	[13, 18, 47, 103, 106, 112–116]
II. Alzheimer's disease. Susceptibility. Main underlying mechanisms and networks	
Susceptibility. Strong genetic component. Genetics of AD. Genetic and genomic risk factors in early onset FAD (ADAD) and/or LOAD. Missing heritability. Need for new techniques/ NGS studies and global databases, curated repositories (early and late onset variants CNVs, rearrangement and others; e.g., AlzGene database; National Institute of Ageing (NIA) and Alzheimer's Disease Sequencing Project (ADSP), https://www.niagads.org/adsp/). Protective variants and essential studies on variants leading to homeostasis/proteostasis collapse. Towards rational risk classification / stratification.	[88, 93, 94, 107, 117–129]
Main underlying mechanisms and networks (1) Early imbalances and generation of toxic aβ oligomers, tau species, aggregates, tangles and cytotoxic compounds leading to, impaired networks and spreading of dysregulation.	[79, 81, 130, 131]
Alterations in APP processing and tau networks. Toxic oligomers and aggregates (intracellular, extracellular)	[85, 87, 132–141]
Amyloid cascade and oligomer cascade hypotheses. APP processing, toxic amyloid-β and tau oligomers and deposits and interplaying networks	[95, 115, 132, 134, 137, 139, 142–147]

(continued)

Table 1
(continued)

Systems biology of Alzheimer's disease	References
Oxidative stress/damage	[146, 148–155]
Calcium dysregulation (e.g., disruption of homeostatic regulation of neuronal Ca) and role of metals (e.g., Cu, Fe) and radical ions in APP and tau networks and protein aggregation.	[156–161]
Noncoding RNAs, miRNAs networks and splicing altered networks	[162–166]
Mitochondrial dysfunction	[167–170]
(2) Homeostatic, stress-protective networks susceptible to impairment themselves, particularly in Alzheimer's disease	
Protein folding and protein quality control homeostasis (proteostasis) networks. From basic ER stress, accumulation of aggregates to proteostasis at all cellular levels. ER folding and ER stress responses; Unfolded protein response (UPR); Endoplasmic reticulum-associated degradation (ERAD); heat shock and chaperone networks in cytosol, ER and mitochondria; Disaggregases and hexosamine pathway; REST protein pathway (*see* also Table 2)	[85, 135, 171–187]
Endolysosomal traffic network and autophagy	[181, 188–191]
(3) Mechanisms and networks at supra-cellular and physiological levels	
Extra-cellular propagation/spreading of the disease. Role of prion-like mechanisms.	[139, 192–194]
Clearance of toxic compounds in CNS and CSF.	[195–197]
Immunological responses—Inflammation—Glia	[198–204]
III. Disease models to unveil mechanisms and networks underlying Alzheimer's disease. From cellular models to human	
From single-celled cellular models to human	[187, 205–213]
Yeast as a model for complex diseases to unravel mechanisms of impairment and neurodegeneration	[13, 18, 154, 182, 214–218]
Yeast models of tau, amyloid-β aggregation and cytotoxicity in Alzheimer's disease	[215–217, 219–222]

Drosophila melanogaster (fruit fly)	[206, 223, 224]
Mammalian cell models, patient-derived neurons cell lines; animal models and human. Down's syndrome and AD longitudinal studies and trails	[153, 207–211, 213, 225–229]
First demonstration of a human model of AD. A 3D human neural cell culture model of AD recapitulating for the first time spontaneous formation of both Aβ plaques and neurofibrillary tangles in human cultured cells, showing treatment of the cultures with β- or γ-secretase inhibitors can dramatically reduce amyloid-β pathology and attenuate tauopathy	[230]
Experimental models for Neurodegenerative Diseases. EC-supported JPND action group. (January 2014) http://www.neurodegenerationresearch.eu/uploads/media/JPND_Exp_Models_Final_report_Jan_2014_-_DM.pdf	
IV. Next generation sequencing (NGS), genomics and molecular techniques, experimental systems biology and computational network biology approaches to unveil mechanisms and dynamic interplay of networks underlying Alzheimer's disease	
(a) Experimental Systems Biology. Omic levels. Interactomes. Networks involved in AD	
Next generation molecular and high-throughput methods for the study of disease susceptibility and networks at different 'omics levels (genome, transcriptome, proteome, metabolome levels).	[90, 107, 120, 122, 123, 126, 153, 155, 162, 163, 231–249]
Signaling pathways and complex networks involved in Alzheimer's disease	[130, 131, 162, 163]
(b) Computational Systems Biology. Network biology.	
Computational and integrative network biology approaches for the study of genomic susceptibility and network modules underlying AD, network dynamics and interplay in time-course experiments, selected groups, longitudinal cohort studies and trials	[246, 250–255]
V. Systems Biology of Alzheimer's disease in practice. New approaches towards early diagnostics and timely intervention, Preventive, promoting active lifestyle and therapeutic. From Systems Biology to Systems Medicine and Public Health	
Application of the National Institute on Aging-Alzheimer's Association AD criteria to ADNI	[99–103]
Clinical diagnostics. The neurological disease ontology	
Standardization of sampling, processing, bioinformatics and statistical protocols, and techniques validation is mandatory	[237, 256–258]

(continued)

Table 1
(continued)

Systems biology of Alzheimer's disease	References
From genomic susceptibility to periodic monitoring, early diagnostics and tailored timely intervention. From Systems Biology to Translational Systems Medicine	[56, 57, 59, 60, 84, 108, 259–264]
Research on combined, multimodal panels of biomarkers at physiological and molecular levels (e.g., neuroimaging techniques and fluid biomarkers) using multimodal techniques. From disease models to human. Longitudinal cohort studies. Towards noninvasive biomarkers and early diagnosis of AD. Alzheimer's Disease Neuroimaging Initiative (ADNI) and database (http://www.adni-info.org/); World Wide Alzheimer's Disease Neuroimaging Initiative (WW-ADNI; http://www.alz.org/research/funding/partnerships/WW-ADNI_overview.asp)	[47, 59, 80, 97, 106, 108, 127, 140, 152, 153, 162, 195, 226, 227, 231, 232, 240, 245, 259–261, 265–280]
Screening for natural compounds and new drugs. Rational screening strategies from disease models to human. For example: (a) Inhibitors or modulators of aberrant production of cytotoxic compounds and/or inflammation (see also books, monographies above (I) and references therein) (b) Compounds that rescue proteotoxicity and/or potentiate global homeostatic defense responses (redox homeostasis; disaggregase activities; UPS/ERAD, autophagy immunomodulation and others) (e.g., screening for proteasomal activators; therapeutic induction of autophagy to modulate neurodegenerative progression) (*see* also Table 2)	[138, 183, 213, 216, 219, 223, 281–290]
Repurposing of approved drugs. New therapeutics.	[116, 219, 291–296]
Case example:	
FDA approved lipoxygenase (LOX) inhibitors for treatment of asthma (e.g., zileuton) and with chemopreventive effects in cancers [294] together with new blood brain barrier (BBB) permeable curcumin-derivative LOX inhibitors lower Aβ accumulation through activation of UPR, proteostasis and autophagy, limit the accumulation of ubiquitinated aggregated proteins and improve memory, amyloid and tau pathology in AD mouse models [295, 296] Opportunities for drug repurposing (see also books, monographies above (I) and references therein)	

Studies on environmental conditions potentiating global homeostatic defense responses (redox homeostasis; proteostasis; disaggregase activities; UPS/ERAD, autophagy). Scientific evidence for minimizing chronic stress, enriched environment (EE), exercise and balanced diet, active healthy lifestyle (*see* also Table 2)	[184, 186, 207, 269, 297, 298]
Alzheimer's disease. Dementia and risk reduction: Analyses of protective and modifiable factors	[77, 299]

(*)(I) Alzheimer's disease. Basis diagnosis and guidelines, susceptibility, genetics, molecular, cellular and physiological studies and mechanisms underlying onset, progression and pathological hallmarks. (II) Main underlying mechanisms at different 'omic levels, interactomes and networks, dynamic interplay between impaired (e.g., APP processing, tau) and homeostatic networks (e.g., proteostasis). (III) Disease models to unveil mechanisms and networks underlying Alzheimer's disease. From cellular models to human. (IV) Next generation sequencing (NGS), genomics and molecular techniques, experimental systems biology and computational systems/network biology approaches to unveil mechanisms and dynamic interplay of networks underlying Alzheimer's disease. (V) Systems Biology of Alzheimer's disease in practice. New approaches towards early diagnostics and timely tailored intervention (preventive, promoting active lifestyle and therapeutic). From Systems Biology to Systems Medicine and Public Health. *ADAD* autosomal dominant Alzheimer's disease; *ADI* Alzheimer's disease international; *ADNI* Alzheimer's disease neuroimaging initiative; *ADSP* Alzheimer's disease sequencing project; *APP* amyloid precursor protein; *CNS* central nervous system; *CNVs* copy number variations; *CSF* cerebrospinal fluid; *EE* enriched environment; *ER stress* endoplasmic reticulum stress; *FAD* familial Alzheimer's disease; *JPND* EU Joint Programme Neurodegenerative Disease Research; *LOAD* late-onset alzheimer's disease; *NIA* US National Institute on Aging, *NGS* Next generation sequencing; *REST protein pathway* RE1-silencing transcription factor pathway; *UPR* unfolded protein response; *UPS/ERAD* ubiquitin proteasome system coupled to endoplasmic reticulum-associated degradation

is known as autosomal dominant Alzheimer's disease (ADAD) [301]. Most of autosomal dominant AD can be attributed to mutations in one of three genes: those encoding amyloid precursor protein (APP) and presenilins 1 and 2 [266, 302]. Most mutations in the APP and presenilin genes increase the production of the most toxic protein fragment $A\beta_{42}$, which is the main component of aggregates and senile plaques, or the ratio between $A\beta_{42}$ and the other forms e.g., $A\beta_{40}$ [303]. The latest studies on impairment in APP processing in ADAD have reported presenilin-1 mutations dramatically reduce trimming of long amyloid β-peptides (Aβ) by γ-secretase which leads to the increase of toxic $A\beta_{42}/A\beta_{40}$ ratio [304].

Most cases of Alzheimer's disease do not exhibit autosomal-dominant inheritance and are termed 'sporadic' or late-onset AD (LOAD), in which both genetic and environmental differences act as risk factors. The best known genetic risk factor in these cases is the inheritance of the ε4 allele of the apolipoprotein E (APOE) gene [305]. LOAD is the most common form of Alzheimer's disease, accounting for >90 % of cases, with overt symptoms usually occurring after age 65 [306].

Most relevant progress on mechanisms and pathways underlying AD has come from comprehensive studies on genetic factors and their contribution to the disease. This entails the study of not only gene sequences (with mutations occurring in both coding and non-coding regions), but also genomic alterations (including copy number variations (CNVs), structural rearrangements, aneuploidies and others) which can affect the regulation of pathways and networks underlying susceptibility to AD. All these are being subjected to exhaustive investigation. Most relevant discoveries on AD genomic susceptibility using molecular and high-throughput experimental systems biology techniques such as next generation sequencing (NGS) and network interactome approaches can be summarized in selected references ([88, 107, 117, 124, 128, 247]; AlzGene database (http://www.alzgene.org/); National Institute of Ageing (NIA) (http://www.nia.nih.gov/); Alzheimer's Disease Sequencing Project (ADSP) (https://www.niagads.org/adsp/); *see* also Table 1). The latest approaches using both experimental and computational systems biology techniques are revealing new genomic loci and pathways underlying complex diseases and are beginning to set new standards in terms of the evidence for causality [246].

Studies on genetic susceptibility to AD are already delivering key insights into the role of specific mechanisms and pathways such as impaired APP processing and tau networks. Thus, APP locus duplication has been shown to lead to autosomal dominant early-onset alzheimer disease (ADAD) [307]. Furthermore, as many as 50 % of people with Down's syndrome (trisomy 21) who live into their 60s may be affected by AD. This high risk has been linked to the presence of the extra copy of the APP gene in chromosome 21 (Alzheimer's Society UK; http://www.alzheimers.org.uk).

The important role of impaired APP processing in AD is supported by the existence of a protective mutation (A673T) in APP, present in a small group of Scandinavian people who do not get AD. This mutation reduces the effectiveness of APP as a substrate for β-secretase which, in turn, reduces the production of Aβ [121]. New variants and genomic loci are being studied, and those affecting homeostatic, stress-protective networks responsible for clearance of toxic compounds in age-related protein aggregation diseases (e.g., LOAD), compromising homeostatic responses or leading to proteostasis collapse should be included [129].

Landmark studies on families carrying known autosomal dominant mutations are providing relevant information on the early presymptomatic stages, onset and progression in ADAD. Thus, Fig. 3 shows the timeline of ADAD, obtained from cross-sectional studies, showing clear changes happening decades before symptom onset, in good agreement with clinicopathological evidence [80].

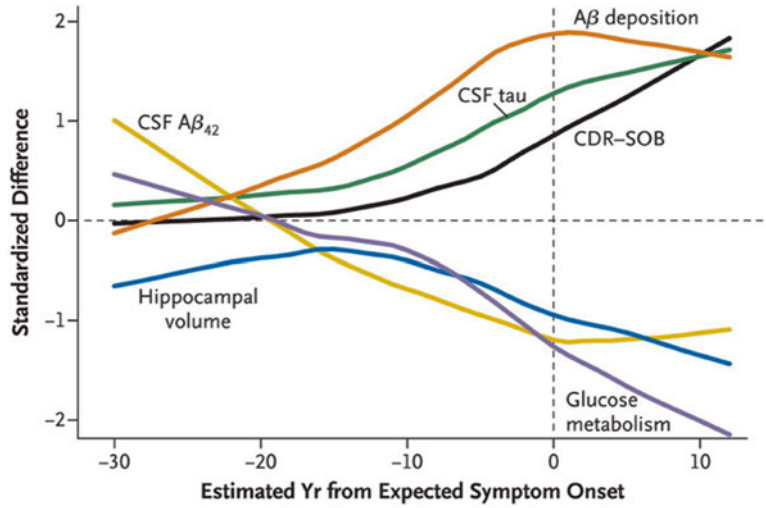


Fig. 3 Timeline in autosomal dominant Alzheimer's disease (ADAD). Cross-sectional studies in the Dominantly Inherited Alzheimer Network (DIAN) project. Comparison of clinical, cognitive, structural, metabolic and biochemical changes as a function of estimated years from expected symptom onset. The normalized differences between mutation carriers and noncarriers are shown versus estimated years from expected symptom onset and plotted with a fitted curve. The order of differences suggests decreasing $A\beta_{42}$ in the cerebrospinal fluid (CSF $A\beta_{42}$), followed by fibrillar Aβ deposition, then increased tau in the CSF (CSF tau), followed by hippocampal atrophy and hypometabolism, with cognitive and clinical changes (as measured by the Clinical Dementia Rating–Sum of Boxes [CDR-SOB]) occurring later. Mild dementia (CDR 1) occurred an average of 3.3 years before expected symptom onset. Reproduced from *The New England Journal of Medicine*, Bateman, Xiong, Benzinger et al. "Clinical and biomarker changes in dominantly inherited Alzheimer's disease", vol 367, 795–804. Copyright © (2012) [80] Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society

More recent longitudinal studies are confirming this, with a possible change of pattern (decrease in concentrations of CSF biomarkers of neuronal injury) after symptom onset, to be confirmed in further longitudinal within-person studies [265]. At this point, it is important to remark that the results and proposed model of biomarkers progression of dominantly inherited AD may not apply to patients with sporadic late-onset alzheimer's disease (LOAD) [80]. Comprehensive systems biology and physiology approaches will be required to translate these findings to sporadic disease [18, 47].

The reality is that there is a broad spectrum of AD patterns, with broad range in the age of onset and rates of progression [306, 308] different not only between LOAD and ADAD, or within LOAD, but also within Down's syndrome people, within ADAD individuals carrying the same mutation [309] and even in monozygotic twins [310–312]. These cases provide an excellent opportunity to investigate candidate epigenetic and environmental contributions underlying AD, a multifactorial disease [81] involving genomic, epigenomic, interactomics, networks dynamics and environmental factors. Any integrative perspective aimed at understanding the common underlying features, and global differences between AD cases will require continued scrutiny and validation using advanced techniques in both experimental and computational Systems Biology approaches (see next sections).

2 Susceptibility and Dynamic Interplay of Impaired and Homeostatic Networks Underlying the Onset and Progression of Alzheimer's Disease

Fundamental studies on neurodegeneration in AD are showing the existence of a high number of 'actors' relevant to brain cells' function, synaptic connectivity, development and plasticity being impaired in AD, at the molecular, cellular, supra-cellular, synaptic, glial, neuronal circuitry, and physiological levels. For instance: (a) Abnormal accumulation of Aβ and tau as oligomers, neuritic plaques and neurofibrillary tangles which may impair neuronal function, with tau aggregates destabilizing microtubules and axonal transport and thus compromising synaptic function; (b) Aβ aggregates inducing the proliferation and activation of astrocytes and microglia, leading to the production of neurotoxic cytokines and reactive oxygen species (ROS); (c) Aβ-induced endocytosis of synaptic NMDA and AMPA receptors, with increased calcium influx through calcium channels and impaired re-uptake of glutamate by astrocytes leading to synaptic dysfunction; (d) Aβ aggregates activating caspases through several pathways, including cell death receptors, calpain activation and mitochondrial damage, leading to neuronal apoptosis ([79] and references therein). Table 1 provides a fully referenced guide to the literature on mechanisms, pathways and networks that underlie AD.

Confronted with the high complexity of AD, there is a need to go 'back to basics' focusing on the core mechanisms, pathways, and networks that underlie the disease at its earliest cellular stages. At this level the principal actors are revealed to be: (1) Impaired amyloid precursor protein (APP) processing and tau networks. These may be impaired due to intrinsic genetic or epigenetic susceptibilities which may exert their effects at various stages of the individual's life, for example as a result of environmental perturbations or stresses. (2) The stress-response homeostatic networks (predominantly proteostasis) actively counteracting the earliest imbalances. These, often overlooked, essential homeostatic networks are the ones which, if overwhelmed by severe or sustained accumulative stress, may lead to the activation of senescence pathways and lead to apoptotic cell death (e.g., [313] and below). Importantly, these homeostatic networks may be progressively impaired through life [18, 62–64], thus contributing to AD being a disease of old age (LOAD).

Studies on the role the APP processing and tau networks in AD are yielding new results and more refined knowledge. As a consequence, the classical amyloid cascade hypothesis [144] is being refined with the incorporation of the new "oligomer cascade" hypothesis [134, 145]. This proposes that toxic Aβ and tau oligomers are the main initiating pathogenic agents in AD. The accumulation and spread of these soluble oligomers are likely to be responsible for earliest dysfunction in AD [134], followed by amyloid plaques and neurofibrillary tangles deposition [144]. The balance of current evidence substantiates the view of AD as essentially a 'proteinopathy', in which increased production of Aβ and tau oligomers and aggregates, and the failure of the homeostatic networks to clear these proteotoxic species underlies the earliest stages of the disease and its progression. In addition to this, new information on the role of tau, toxic tau oligomers, and neurofibrillary tangles in AD further illuminate our overall view of the cellular basis of AD [138, 141, 314, 315].

The relevant homeostatic networks include the essential proteostasis networks: e.g., protein-folding chaperone networks, endoplasmic reticulum (ER) stress and unfolded protein response (UPR), disaggregases, the ER-associated degradation/ubiquitin proteasome system (ERAD/UPS), the endolysosomal network, autophagy, and other stress-protective and clearance pathways (Fig. 4). These networks are part of the 'core protein machinery' that is conserved in all eukaryotes, from yeast to human [18, 53, 75, 180, 187, 316]. The proteostatic machinery is responsible for the continuous quality control of the proteome, and should prevent the accumulation of toxic misfolded proteins and aggregates; its underlying mechanisms are subject of intensive investigation ([173, 180, 181, 187, 316–318]; *see* also Table 1). The evolutionary conservation of these proteostatic networks opens the way to

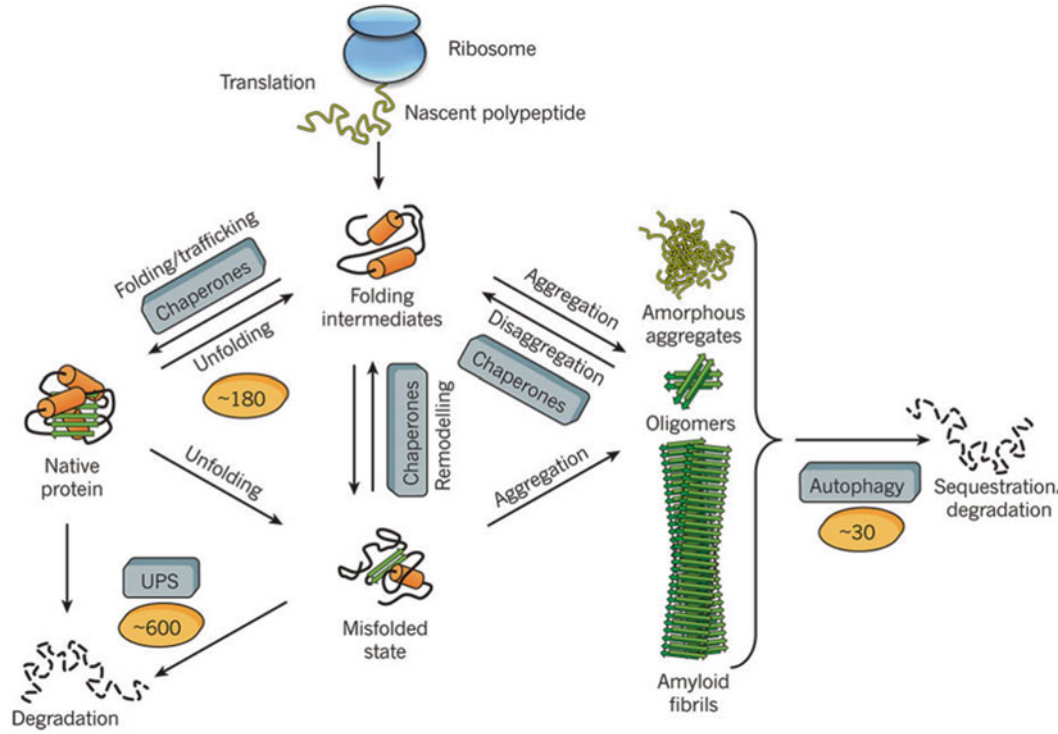


Fig. 4 Proteostasis network (PN). Protein fates. The proteostasis network (PN) integrates chaperone pathways in the cytosol, endoplasmic reticulum (ER), nucleus and mitochondria for the proper folding of newly synthesized proteins, for the remodeling of misfolded states and for disaggregation with the protein degradation mediated by the ubiquitin proteasome system-endoplasmic reticulum associated degradation networks (UPS/ERAD) and the autophagy system. The essential proteostasis network machinery is conserved in all eukaryotes, from yeast to human [173, 180, 316]. Approximately 180 different chaperone components and their regulators orchestrate these processes in mammalian cells, whereas the UPS/ERAD comprises ~600 and the autophagy system ~30 different components. The primary effort of the chaperone system is in preventing aggregation. Sub-cellular organization and organelles omitted for clarity. Crucial machinery for the disaggregation of aggregated proteins has been also detected in yeast and in metazoans, from *C. elegans* [171] to mammals (mouse and human) [172]. Reprinted by permission from Macmillan Publishers Ltd: *Nature* (Hartl et al., *Nature* 475, 324–332) [173], copyright (2011)

systems biology studies (first, in simple disease models such as yeast) aimed at dissecting the interplay of the genetic and environmental factors involved. Subsequent longitudinal studies in human subjects must be performed to validate the predictions made using these simple models ([13, 18] and references therein; see also below).

A considerable body of evidence supports the crucial role of homeostatic networks in determining the initial cascade of events that result in AD. Relevant findings include the following: (1) Tau accumulation activates the unfolded protein response (UPR) by impairing UPS/ERAD and this is reversible, which suggests tau-based therapeutics could significantly delay cell death and disease

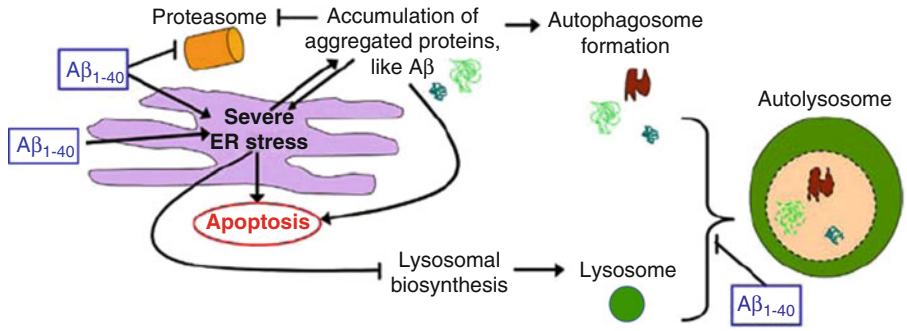


Fig. 5 Amyloid beta peptide $(A\beta)_{1-40}$ ER stress and deregulation of proteostasis in brain endothelial cells. ER stress caused by Aβ toxic species leads to the accumulation of misfolded proteins that can be targeted for degradation in the proteasome or in the lysosome by macroautophagy when become aggregated. However, oligomeric and large protein aggregates block the proteasome and further induce ER stress. Under conditions of severe ER stress, general protein translation is inhibited and impairs ER functioning, compromising the biogenesis of organelles, such as the lysosome. As a consequence, the degradation of autophagosomes' cargo decreases and protein aggregates accumulate. Therefore, ER stress and proteasome blockage are exacerbated, leading to cell death by apoptosis [319]; *see* also ref. 313. Reprinted adapted from *Biochimica et Biophysica Acta. Molecular Cell Research* vol. 1843, Fonseca et al. (2014) Loss of proteostasis induced by amyloid beta peptide in brain endothelial cells, pages 1150–1161. Copyright (2014), [319] with permission from Elsevier

progression [135]. (2) Amyloid beta peptide (Aβ) deregulates proteostasis in brain endothelial cells, leading to accumulation of ubiquitinated proteins, autophagy impairment and cell death [319] (Fig. 5). (3) Metabolic stress induces the phosphorylation of endogenous tau via activation of the UPR, and this is reversible. Upon intervention to restore homeostasis, the levels of UPR markers and tau phosphorylation can be reversed [320]. (4) There is accumulated evidence of the role of impairment of homeostatic and clearance pathways in neurodegenerative diseases other than AD [321, 322]. (5) Sustained protein folding stress, with accumulation of misfolded proteins, activates the UPR and causes a build-up in levels of the death cell receptor DR5. If the stress is relieved soon enough levels of the receptor decay back to normal and the cells stay alive (homeostasis being restored), otherwise induction of apoptotic cell death follows [313] (see also below).

Based on all this, an integrative perspective is presented in Fig. 6. Alzheimer's disease is shown to essentially result from: (a) intrinsic genomic/epigenomic susceptibilities and, (b) the dynamic interplay between impaired and central homeostatic networks. These homeostatic, stress-protective networks appear to be mobilized at the earliest stages of the disease. The interplay of these networks will underlie the time of onset and rate of progression of the disease in a given individual (Fig. 1).

This integrative perspective will need to be refined to include important signaling pathways and the effects of senescence [115, 130, 313], as well as being reconciled with other theories [115, 313].

Thus, focusing on LOAD only, Krstic and Knuesel have proposed this might be initiated by chronic inflammatory conditions causing dysregulation of the mechanisms that clear misfolded or damaged neuronal proteins which accumulate with age, as well as tau-associated impairments of axonal integrity and transport. Together these tip the balance towards events leading to the generation of the aggregation-prone toxic species and AD [335, 336]. It is likely that LOAD may have various origins in different individuals, but the initial sequence of events nevertheless converges on interplay between impaired and homeostatic networks (Fig. 6).

3 Potentiation of Homeostatic Networks at the Early Stages of AD May Delay the Onset, Arrest, or Even Reverse the Progression of the Disease

Studies on homeostatic networks are not only revealing their exquisite complexity, but also their key role in neurodegenerative diseases [316, 321]. Remarkably, studies in simple model organisms, mammalian models, and human cell lines are all beginning to reveal promising results that show that supposedly 'untreatable' proteinopathies such as AD may be reversible. They hold out the hope that the onset and progression of AD may be counteracted by early intervention by appropriate therapies or lifestyle changes. Such interventions might include the use of drugs to modulate proteostasis activities (these might be targeted at chaperones, disaggregases etc.) combined with changes in a individual's diet and exercise regime. These hopes are bolstered by the fact that main proteostasis networks are highly conserved through evolution. The characterization of the additional human-specific homeostatic networks, their role in counteracting, compensating AD pathology in AD and DS individuals, at the cellular, supra-cellular and physiological levels, and the signaling pathways responsible for their regulation, is being actively pursued as well. Relevant examples are summarized in Table 2 and selected case examples are given here:

1. Those providing evidence on the role of essential homeostatic mechanisms, towards candidate therapeutic strategies:
   (a) Tau accumulation activates the unfolded protein response (UPR) by impairing UPS/ERAD (increasing levels of ubiquitinated proteins). Depleting soluble tau levels in cells and brain could reverse UPR activation. The reversibility of

Fig. 6 (continued) *miRNAs* microRNAs networks; *RNA alt. splicing* RNA alternative splicing; *RNPs*, ribonucleoprotein complexes/networks; *PTMs*, post-translational modifications/pathways (e.g., proteolytic cleavage pathways in APP processing and tau networks). *UPR*, unfolded protein response; *UPS-ERAD*, ubiquitin proteasome system-endoplasmic reticulum associated degradation pathways/networks. *UDP-GlcNAc*, UDP-*N*-acetyl-glucosamine (hexosamine pathway) [183, 185]; *REST*, repressor element 1-silencing transcription factor [184]

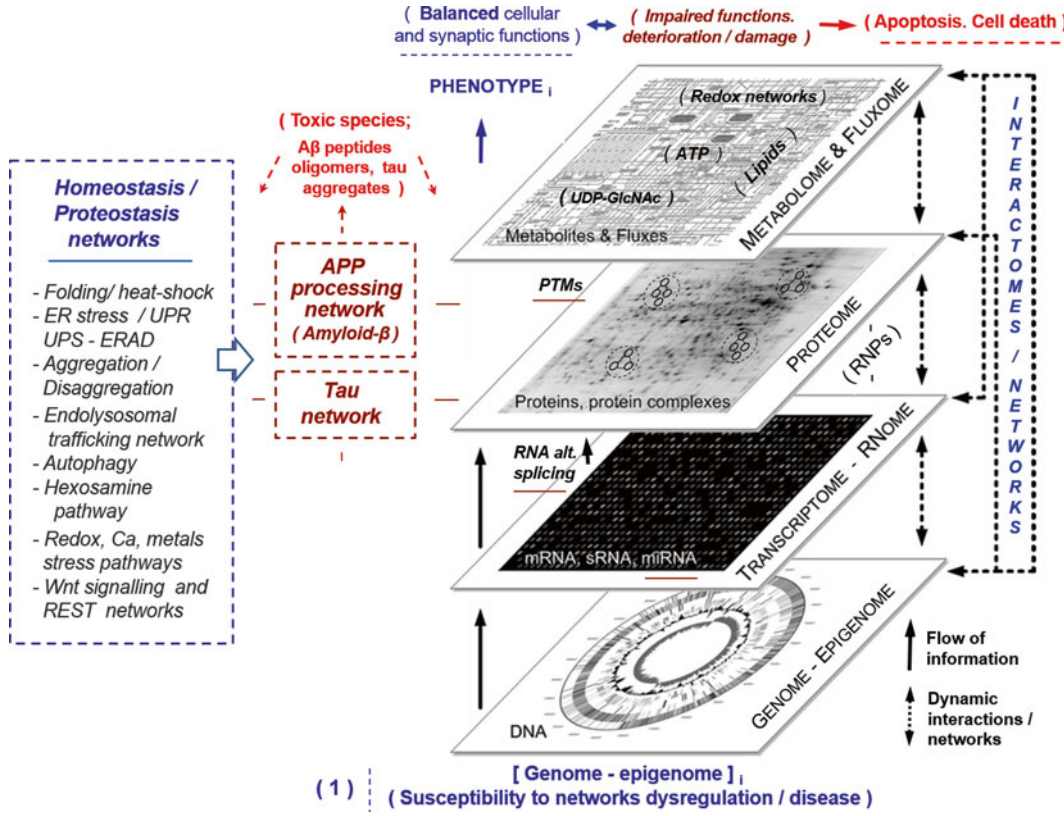


Fig. 6 Alzheimer's disease. An integrative perspective. Susceptibility and dynamic interplay of impaired and homeostatic networks. Main biological networks and 'omic levels relevant to brain cells' function, connectivity and development underlying AD: from genotype to phenotype, are shown. First, the genome and epigenome [1] (*bottom of the figure*) underlie the essential networks, homeostatic states and initial susceptibility to dysregulation/disease. Genetic and/or environmental perturbations/stressors may lead to progressive impairment of the central amyloid precursor protein (APP) processing and tau networks reported main responsible for the neurodegeneration features and characteristic hallmarks of AD [85–87, 89, 93, 106]. With independence of many interplaying mechanisms contributing to dysregulation and progressive impairment (e.g., leading to the accumulation of candidate cytotoxic species such as toxic soluble Aβ monomers, oligomers and/or aggregates; tau-microtubules interactions; tau-small nuclear RNAs aggregates affecting alternative splicing [165, 166], redox species and others;-omitted for clarity), the brain cells' function (e.g., neurons; glia, endothelial cells) is essentially displayed characterized by a fine balance, dynamic interplay between "central modular networks" (including the APP and tau networks and several others; [323, 324]) at different 'omic and interactomes levels [13, 18, 52, 53, 243] and the essential "homeostatic networks", stress-response/defense networks (e.g., redox homeostasis and proteostasis networks), activated to counteract, compensate or minimize cellular imbalances well before irrreversible damage and cell death [172, 189, 313, 325, 326]. The crucial role the homeostatic networks such as the protein quality control (proteostasis) networks including UPS/ERAD and disaggregase activities [171–173, 177–179, 183, 185, 186, 322, 327], endolysosomal trafficking network and autophagy [190, 328, 329] and the heat-shock/proteostasis signaling, Wnt signaling and REST networks [184, 330–332] is often overlooked or underestimated. Specific environmental perturbations/stressors (mild or severe, transient or sustained in time) may also impair homeostatic defenses networks during lifetime [62–64]. In all, intrinsic susceptibility and the fine balance between dynamic networks may result in a broad range of patterns, heterogeneity, in the age of onset and rate of progression of the disease, and diverse patient-specific phenotypes. Sub-cellular organization and organelles omitted for clarity. Additional signaling, impaired and defense homeostatic networks (e.g., glia, immunological and/or inflammatory responses; vascular and glymphatic clearance systems) and neuronal circuitry and networks occurring at the supra-cellular, tissue and/or physiological levels [86, 197, 259, 333] omitted for clarity. Adapted from ref. 334 with permission and ref. 13 with permission from Elsevier. *APP*, amyloid precursor protein; *mRNA*, messenger RNA; *sRNA*, small (noncoding) RNA;

Table 2
Latest studies on characterization of essential stress-protective homeostatic networks, pathways and mechanisms and their interplay in neurodegenerative diseases. Relevant case examples showing their crucial role which may arrest, delay or reverse toxic effects, from simple disease models to human. Support for new studies and molecular evidence of protective and modifiable factors [299]; World Alzheimer Report, 2014. http://www.alz.co.uk/research/world-report-2014) [77]

Description of study (disease model)	References
Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases (eukaryotes)	[322]
Molecular chaperones in protein folding and proteostasis (eukaryotes)	[173]
Biology of the heat shock response and protein chaperones (yeast)	[180]
A quantitative chaperone interaction network reveals the architecture of human cellular protein homeostasis pathways (human cells)	[316]
Essential role of dysfunctional protein homeostasis in neurodegenerative diseases (from single-celled models to human)	[187]
Discovery and characterization of mammalian amyloid disaggregation activities (mammalian tissues; mouse, human)	[172]
Defining human endoplasmic reticulum-associated degradation (ERAD) networks through an integrative mapping strategy (human)	[177]
The ubiquitin-proteasome system (UPS/ERAD) and the autophagy-lysosome system (eukaryotes, mammals, human)	[181]
Beta-amyloid accumulation inhibits the ubiquitin-proteasome system (UPS) and impairs multivesicular body (MVB) sorting (mouse primary neurons)	[337]
Amyloid beta peptide (Aβ) deregulates proteostasis in brain endothelial cells leading to accumulation of ubiquitinated proteins, autophagy impairment and cell death (rat brain endothelial cells)	[319]
Aggregation-mediated Abeta(1–42) toxicity can be reduced in *Caenorhabditis elegans* by decreased insulin signaling and downstream transcription factors heat shock factor 1 and DAF-16 (*C. elegans*, worm)	[171]
Starvation and inhibition of lysosomal function increase tau secretion (mouse primary cortical neurons)	[338]
Endocytosis of extracellular monomeric tau is sufficient to inititate tau pathology (CHO and human neuroblastoma cells)	[139]
Tau accumulation activates the unfolded protein response (UPR) by impairing UPS/ERAD. The reversibility of the process suggests tau-based therapeutics could significantly delay cell death and disease progression (mouse, human)	[135]
Tau promotes neurodegeneration through global epigenetic changes, heterochromatin loss and aberrant gene expression (*Drosophila*, mouse, human)	[339]
Progressive impairment of UPS modules during Alzheimer's disease progression revealed by computational network biology approaches (human)	[255]

(continued)

Table 2
(continued)

Description of study (disease model)	References
Proteasome dysfunction activates autophagy and anti-oxidative pathways (mouse)	[340]
Proteasome dysfunction causes mitochondria impairment leading to increase reactive oxygen species (ROS) production and cell death (Chinese hamster ovary cell lines, CHO)	[341]
Enhancing protein disaggregation restores reduced proteasome activity in aged cells (yeast)	[342]
NAD+ salvage pathway proteins suppress proteotoxicity in yeast models of neurodegeneration by promoting the clearance of misfolded/oligomerized proteins (yeast)	[182]
Reversing deleterious protein aggregation with re-engineered protein disaggregases (e.g., potentiated yeast Hsp104 disaggregase (yeast, *C. elegans*))	[343–345]
Sustained, unmitigated protein folding ER stress activates the UPR and build up levels of death cell receptor DR5. If stress is relieved soon enough levels of the receptor decay back to normal and the cells stay alive (effect counteracted/reversed; homeostasis restored), otherwise induction of apoptotic cell death (human cell lines, mouse)	[313]
Chronic mild stress accelerates the onset and progression of an AD phenotype (mouse)	[207]
Metabolic stress induces the phosphorylation of endogenous tau via activation of the UPR. This is reversible. Upon intervention to restore homeostasis the levels of UPR markers and tau phosphorylation are reversed (human cell lines and Syrian hamsters hypometabolic models)	[320]
Under stress, the spliced X-box binding protein 1 (Xbp1s) couples the UPR to the hexosamine biosynthetic pathway (mouse, rat, human heart tissue)	[186]
Hexosamine pathway metabolites enhance protein quality control, reduce aggregation and extend lifespan (*C. elegans,* worm)	[183]
Deficiency in LRP6-mediated Wnt signaling contributes to synaptic abnormalities and amyloid pathology in Alzheimer's disease (mouse). Restoring Wnt signaling can be explored as a viable strategy for AD therapy.	[323, 346]
Activation of Wnt signaling enhances cognitive function of adult mice and reverses cognitive deficits in an Alzheimer's disease model (mouse).	[332]
Repressor element 1-silencing transcription factor (REST), protecting neurons from oxidative stress and amyloid β-protein toxicity, is lost in mild cognitive impairment and Alzheimer's disease. REST as a new therapeutic target for neurodegenerative disorders (*C. elegans*, mouse, human)	[184, 347]
Environmental novelty activates β2-adrenergic signaling to prevent hippocampal impairment by Aβ oligomers (mouse)	[297]
Environmental enrichment strengthens corticocortical interactions and reduces amyloid-β oligomers in aged mice (mouse)	[298]

(continued)

Table 2
(continued)

Description of study (disease model)	References
8-hydroxyquinolines protect models of TDP-43 protein, α-synuclein, polyglutamine proteotoxicity and rescue Aβ toxicity (yeast)	[284, 348]
Yeast-to-human high-throughput small-molecule screening platform for compounds modulating toxic effects in synucleinopathies, validated in patient-derived neurons (yeast, human)	[213]
Therapeutic induction of autophagy as an strategy to modulate neurodegenerative disease progression (*Drosophila*)	[287]
A multifaceted drug exerts neuroprotective effects and reverses AD-like phenotype (mouse)	[349]
FDA approved lipoxygenase (LOX) inhibitors (e.g., zileuton) and new LOX inhibitors such as blood brain barrier (BBB) permeable curcumin-derivative compounds lower Aβ levels through activation of the UPR, proteostasis and autophagy, limit the accumulation of ubiquitinated aggregated proteins and improve memory, amyloid and tau pathology in AD mouse models	[295, 296]
Immunomodulatory treatment (Toll-like receptor 9 stimulation) with TLR9 agonists such as CpG oligodeoxynucleotides (ODNs) reduces both Aß and tau pathologies, and levels of toxic oligomers, and leads to cognitive rescue in the absence of inflammatory toxicity (mice)	[350]
Potential for primary prevention of Alzheimer's disease. Risk modifiable factors	[299]
World Alzheimer Report (2014). Dementia and Risk Reduction: An analysis of protective and modifiable factors. ADI http://www.alz.co.uk/research/world-report-2014	[77]

the process suggests tau-based therapeutics could significantly delay cell death and disease progression [135].

(b) Activation of the UPR, proteostasis and autophagy via FDA-approved lipoxygenase (LOX) inhibitors such as blood brain barrier permeable curcumin-derivative compounds lower Aβ levels, limit the accumulation of ubiquitinated protein aggregates proteins, and improve memory, as well as amyloid and tau pathology in AD mouse models [295, 296].

(c) Evidence of therapeutic induction of autophagy as a strategy to modulate neurodegenerative disease progression [287].

(d) Immunomodulatory treatments (e.g., Toll-like receptor 9 stimulation with TLR9 agonists) reduces both Aß and tau pathologies and levels of toxic oligomers, and leads to cognitive rescue in the absence of inflammatory toxicity [350].

(e) Sustained, unmitigated protein-misfolding ER stress activates the UPR and build up levels of death cell receptor DR5. If stress is relieved soon enough, levels of the receptor decay back to normal and the cells stay alive (effect counteracted/reversed; homeostasis restored), otherwise induction of apoptotic cell death follows [313].

(f) Metabolic stress induces the phosphorylation of endogenous tau via activation of the UPR. This is reversible—upon intervention to restore homeostasis, the levels of UPR markers and tau phosphorylation are reversed [320].

2. Evidence that the benefits of modifying risk factors such as poor diet or a sedentary lifestyle [77, 299] may be mediated by homeostatic networks and signaling pathways which might be investigated as candidate therapeutic targets as well:

(a) Environmental novelty activates β2-adrenergic signaling to prevent hippocampal impairment by Aβ oligomers [297].

(b) Environmental enrichment (EE) strengthens corticocortical interactions and reduces amyloid-β oligomers in aged mice [298].

(c) Modulation of protein homeostasis and lifespan depending on diet composition [351].

(d) Aggregation-mediated Abeta(1–42) toxicity can be reduced by decreased insulin signaling and downstream transcription factors heat shock factor 1 and DAF-16 [171].

(e) Deficiency in LRP6-mediated Wnt signaling contributes to synaptic abnormalities and amyloid pathology in AD. Restoring Wnt signaling can be explored as a viable strategy for AD therapy [323, 346].

(f) Activation of Wnt signaling enhances cognitive function of adult mice and reverses cognitive deficits in an AD model [332].

(g) The repressor element 1-silencing transcription factor (REST) pathway, which protects neurons from oxidative stress and amyloid β-protein toxicity, is lost in mild cognitive impairment and AD. Thus REST potentiation is a candidate strategy for neurodegenerative disorders [184, 347].

(h) In the brains of those with AD, both Wnt signaling and REST induction are suppressed, leading to neurodegeneration [184, 352].

(i) The hexosamine pathway, protein folding stress induces the UPR, with the spliced X-box binding protein 1 activating key enzymes of the hexosamine biosynthetic pathway (HBP) in studies on mouse, rat and human heart tissue [186]. HBP metabolites enhance protein quality control, reduce aggregation, and extend lifespan in *C. elegans* [183, 185].

As promising as these results may appear, they will need to be validated by new evidence, with early detection of impaired mechanisms and their dynamic interplay with homeostatic networks as one of the most formidable challenges ahead. Longitudinal studies applying new advances in the neuroimaging of small oligomers or aggregates, and of increasing deposition (indicative of the net

generation of toxic species, i.e., disease progression) [273, 276, 277, 353, 354] will be of paramount importance. Together with this, sensitive molecular and high-throughput techniques able to reliably detect earliest impairments at different 'omic levels, such as altered RNA splicing, redox imbalances, proteotoxic species (Aβ and tau species), altered RNAs, proteins and/or metabolic patterns, biomarkers for inflammation and homeostatic and clearance activities well before apoptotic cell death, would constitute the best-case scenario. Biomarkers for the presence of apoptosis (e.g., proteins or lipids membranes, detectable in biofluids) represent significant advances in revealing brain cells' lysis and disease progression. However, such methods may only be able to detect AD at a stage when counteracting strategies potentiating homeostatic mechanisms such as some of those presented in Table 2 may no longer reverse the disease. In this context, advances in the identification of patterns and candidate biomarkers of impaired and compensatory mechanisms and networks during the early decades of Down's syndrome (DS) and ADAD individuals (at the asymptomatic stage), able to counteract early acute impairments (e.g., altered RNA splicing, redox imbalances and systemic oxidative stress from the womb [165, 211, 355]), efficiently halting or delaying the onset of the disease until these patients' 40s or 50s will be of great interest.

At this point, it appears clear new studies are beginning to yield promising results and opening up still more avenues of research. However, in order to achieve significant and steady progress towards risk classification, early diagnosis, and mechanistically-based therapeutic and/or lifestyle interventions, a new strategy is needed. We submit that there is a need to advance our understanding of AD by taking an 'holistic' Systems Biology perspective (*see* Section 1 of this chapter). That is, AD as a multifactorial disease, primarily arising from altered networks affecting essential modules and pathways, and fundamentally 'dynamic' (Figs. 1 and 6). The greatest progress will come from the integration of knowledge from studies in several disease models and by the synergistic interaction of researchers from diverse disciplines. Their novel findings will need to be validated in humans in order to make the transition from Systems Biology to Translational Systems Medicine [57, 59, 60, 263, 264]. This transition can only be made within an enlightened framework of Public Health benefit policies, recommendations and incentives [77, 299].

4 Implementing a Systems Biology Approach to AD

The previous results and integrative perspective open the way to new hypothesis-driven comprehensive studies on the role of impaired and homeostatic networks and their dynamics in AD, but

how? If a Systems Biology approach [13, 18, 39, 46, 47] is to succeed, what is required? We suggest that the following will be needed:

1. Advanced molecular and high-throughput techniques for monitoring time-course experiments coupled to longitudinal studies that take into account the effects of specific environmental stresses (either transient or sustained), and the drug regimes to which individuals may be subject for other (apparently unrelated) conditions.
2. Computational and integrative network biology tools and approaches for the elucidation of genomic regions susceptibility and network modules underlying AD, and their dynamics in time-course experiments and longitudinal cohort studies. Raw data and integrative analysis approaches will need to be deposited in well-curated databases and data repositories, with essential metadata (e.g., conditions and techniques used) to guide in the identification of real comparable datasets, for solid analytical studies (i.e. computational systems biology).

Huge efforts are being made and advances steadily produced in both the experimental and computational systems biology areas. Selected examples are shown in Table 1, *see* also refs. 13, 18, Handbook of Systems Biology [114], and chapters and contributors in this Systems Biology of AD volume. Amongst most relevant experimental systems biology molecular and high-throughput techniques are, for instance: (a) New molecular tracers and neuroimaging approaches to study small oligomers, aggregates, and amyloid plaques and tangles deposition [276, 353, 354]; (b) Molecular super-fluorescence resolution microscopy/nanoscopy approaches able to unveil earliest molecular events, aggregation and dysfunction at the cellular level [356, 357] (http://www.nature.com/news/nobel-for-microscopy-that-reveals-inner-world-of-cells-1.16097); (c) new molecular biology techniques (e.g., CRISPR-Cas9 genome editing) opening the way to the construction of advanced disease models, from simple organisms (yeast) [358] to animal models (e.g., mouse) in vivo at any stage in the animal's life, or disease stage [359]; (d) Next generation high-throughput 'omics (genome/epigenome, transcriptome, proteome, metabolome and interactomes) techniques under controlled conditions including, among others: Next generation sequencing (NGS) (e.g., whole genome sequencing (WGS), exome-sequencing and others) to study not only genes but also non-coding regions, epigenetic patterns, CNVs and structural rearrangements, aneuploidies and mosaicism, to illuminate genomic/epigenomic risk susceptibility; Transcriptional (e.g., RNA-sequencing, impaired RNAs splicing and microRNAs expression methods) and proteomic techniques, to study genome-wide expression and altered patterns at the transcriptional and proteome/peptidome levels in

disease models under different conditions (e.g., top-down, http://www.the-scientist.com/?articles.view/articleNo/40248/title/Bird-s-Eye-Proteomics/, and bottom-up proteomics, http://www.the-scientist.com/?articles.view/articleNo/40051/title/Moving-Target/, single reaction monitoring (SRM/MRM), data independent acquisition DIA (SWATH-MS) next generation proteomics, fluorescence-activated cell sorting (FACS) and other targeted proteomics, metabolomics, lipidomics, and differential interactomes approaches [122, 126, 165, 166, 235, 241, 360–364], and many others (*see* Table 1, [13, 18, 114], this volume and references therein)).

Comprehensive integrative systems biology experiments studying transcriptome, proteome, metabolome patterns and interactions under defined conditions were first achieved in yeast, as a reference 'model eukaryote' [75] (*see* also ref. 18 and references therein). This has opened the way to studies in other organisms and finally to its implantation in human. Thus, multi-omics studies could be performed in longitudinal studies in human using integrative Personal 'Omics Profiling (iPOP), monitoring panels of biomarkers and patterns towards diagnosis and tailored personalized medicine. While still expensive, such approaches are progressively becoming more affordable [56, 60, 262, 365] and Table 1. These integrative 'omics approaches could be combined with excellent reference studies monitoring patterns of onset and progression in AD [80, 265].

Computational systems biology approaches are also being continuously developed and refined, for integrative data analyses and for the construction and analysis of networks and modules underlying complex diseases. For basic rules on integrative analyses of 'omics datasets under equivalent comparable conditions (i.e., without introduction of systematic error, noise or bias), and latest integrative tools on analyses of dynamic responses, affected modules and networks in time-course experiments and other approaches *see* refs. 13, 18, and Handbook of Systems Biology [114]. Relevant to neurodegenerative diseases and AD, the excellent work of groups such as those of Cruchaga, Califano, Bar-Joseph, Gitter, Ogishima, and Zhang, which combine large experimental datasets with advanced computational network biology approaches, are already delivering essential information on AD genomic susceptibility and new causal drivers of AD [122, 366]. More importantly, they also provide information on the dynamics of dysregulation of essential modules and networks, from the early stages, through progression (e.g., altered ERAD/UPS proteasomal modules/networks), to the late stages of AD (e.g., acute impaired gene expression with heterochromatin dysfunction; apoptopic, inflammatory and innate immunity responses) [251, 254, 255]. With progressive incorporation of better input datasets from carefully designed longitudinal experiments, including the earliest asymptomatic stages (e.g., since first decades of life in voluntary DS and ADAD individuals and people with antecedents,

family history of LOADs), these are expected to provide unprecedented insights on mechanisms and networks underlying AD. When combined with information on their dynamics across the span of human life, including the influence of environmental perturbations, they may provide an opportunity for tailored interventions. For more information on computational/networks biology and combined approaches *see* Table 1 and chapters in this volume.

At this point, it is important to note that these exciting new approaches do have their limitations.

4.1 The Need for Standardization of Techniques and Data Records

Newly developed techniques, protocols and approaches need to be tested, compared with previous ones and, once validated, lead to the establishment of new standards and guidelines. This is a continuous process, fundamental to scientific research, and of prime importance in diagnostics. For example, progress in early diagnosis of complex diseases will come with the establishment of new guidelines on clinical genome and exome sequencing [367], advances in validation and standardization of experimental techniques, and proper data integration of 'omics datasets (e.g., *see* Suppl. Methods in refs. 75, and 18, 60, 262, with guidelines for longitudinal studies and trials, e.g., US Food and Drug Administration (FDA) clinical trials guidances, with adherence to the principles of good clinical practice (GCP): http://www.fda.gov/regulatoryinformation/guidances/ucm122046.htm; http://clinicaltrials.gov/; European Medicine Agency (EMA) and European Clinical Trials Database (EudraCT): https://eudract.ema.europa.eu/, and [13, 18]). Together with this, the FDA recently issued draft guidance documents outlining its regulatory plans for laboratory-developed tests (LDTs) for molecular diagnostics; http://www.bio-itworld.com/2014/10/3/fda-issues-draft-guidance-laboratory-developed-tests.html.

More specifically, the main advances on standardization and guidelines in neurodegenerative diseases and AD are coming from global efforts on neuroimaging and biomarkers; for example, form the Alzheimer's disease neuroimaging initiative (ADNI) (http://www.adni-info.org/) and the Dominantly Inherited Alzheimer Network (DIAN) (http://dian-info.org/) (*see* also Table 1), together with new studies on sample quality and assessment of, for example, cerebrospinal fluid samples for biomarker investigations [237], and new guidelines for the standardization of preanalytic variables for blood-based biomarker studies in AD research, from the STandards for Alzheimer's Research in Blood biomarkers (STAR-B) and Blood-Based Biomarker Interest Group (BBBIG) working groups [368]. Still more initiatives are in progress.

4.2 From Single-Celled Models to Humans

By definition, comprehensive studies on multifactorial complex diseases such as AD need to address two main objectives:

1. The construction of reliable models which recapitulate altered mechanisms and features of the disease at the molecular and

cellular level, leading to the generation of the characteristic supra-cellular features, hallmarks of the disease (e.g., in AD, amyloid plaques and neurofibrillary tangles). This is the most important 'take-home message', the main requirement of good disease models: 'to be able to recapitulate molecular and physiological features of the disease', towards a better understanding of the real events in vivo [18].

2. To analyze the progression of the altered networks and phenotypes in well-designed experiments and models under controlled conditions, reproducing environmental in vivo perturbations (whether transient or sustained), whose effects are representative of those contributing to the disease. These experimental systems should be easy to implement, monitoring the impaired networks together with the activation of defense responses with their dynamics and interplay, and enable progress towards direct applications (e.g., earliest diagnosis and timely intervention). To do this effectively demands:
   (a) Careful selection of the disease model to be used for specific objective(s), with construction of new advanced models, including (and this is crucial) female and male disease models at the cellular, supra-cellular, animal model levels [369–371]. The greatest progress will come from the integration of knowledge from studies on several disease models, to be confirmed in humans.
   (b) The use of proper experimental design, minimizing confounding variables and bias, and putting in place a bioinformatic and statistical strategy from the outset. Early discussions with bioinformatic experts analyzing goals and expected results, number of experiments, conditions and replicates, costs, and alternative data analysis strategies, in order to get statistically significant results and solid conclusions, will be of clear benefit (*see* ref. 18 and references therein).

In the end, complex human diseases will be characterized by the dynamic interplay between impaired and counteracting homeostatic networks (Figs. 1 and 6). If the main objective is the study of essentially conserved networks common to all eukaryotes (e.g., main homeostatic networks, proteostasis and others) yeast can be a first (and excellent) model, with advanced techniques to unravel basic mechanisms, and time course experiments to study homeostatic networks counteracting proteotoxicity, monitoring dynamic responses to transient or sustained perturbations under controlled conditions. Such experiments are very difficult to implement or not affordable with mammalian and human cell lines, and supra-cellular animal models [18].

A selection of the most relevant studies with AD disease models is included in Table 1. The main disease models in studies showing

early intervention may arrest, delay or reverse toxic effects in AD, from single-celled organisms to human, are presented in Table 2.

The main challenge will always be the construction of models that recapitulate main mechanisms and disease features at the cellular, supra-cellular and tissue/brain regions levels (for example, in animal models e.g., rodents, see below). At the cellular level, good results could be expected from studies using human cell lines or patient-specific cell lines from induced pluripotent stem cells (iPSC) under controlled conditions [209, 210, 369, 372–375]. These are the cases, one would expect, in which human-specific mechanisms, e.g., RNA splicing patterns, protein isoforms, the stoichiometry of protein complexes (e.g., γ-secretase), interactions, interactomes, and interplay of networks (Fig. 6) may approach what happens in patients. However, it is important to note that neurons in culture may not reflect conditions of normal neurons in the brain and more studies and models will be needed. As general rule, since a single model cannot fully recapitulate all AD features at all levels, continued integration of knowledge from different models, from single-celled models, mammalian cell lines, patient-specific cell lines, to supra-cellular animal models and human longitudinal studies [43, 187, 209, 210, 373–378] will be required.

In the case of AD, most relevant landmark disease models recapitulating main AD features are:

1. Human cellular models. Human neurons derived from AD patients have been reported showing elevated levels of toxic amyloid-β species and phosphorylated tau [379–381].
2. Patient-specific cell lines from induced pluripotent stem cells (iPSC). Thus, Shi and coworkers (Livesey's group) showed for the first time both Aβ deposition and tau pathology, from iPSC-derived neuron cultures derived from Down's syndrome [209, 210].
3. Supra-cellular and animal models. The extreme difficulty of recapitulating all AD features in basic mouse models has been explained as likely to be a consequence of the different properties of the mouse Aβ peptide and intracellular mouse proteome compared to that of the human. The construction of a closer-to-human rat model, expressing mutant human APP and presenilin 1 genes only, has enabled the recapitulation of all AD human features, including plaques and tau pathology. These rats manifest age-dependent (in their 6–26 months) cerebral amyloidosis preceding tauopathy, gliosis, apoptotic loss of neurons in the cerebral cortex and hippocampus, and cognitive disturbance [136, 208].

Advanced mouse models showing both AD hallmarks are being obtained, thanks to the introduction of more carefully

designed, fine manipulation. Although distant from human and with their limitations, some of these models are providing important insights. Firstly, creation of a triple transgenic mouse that overexpressed pathogenic variants of APP, presenilin-1, and tau gave rise to progressive formation of amyloid plaques and NFTs, as well as synaptic dysfunction. This appeared at four months of age and correlated with hippocampal accumulation of intraneuronal Aβ, and preceded the appearance of plaques and tangles [382]; *see* also ref. 79 and references therein. More recently, triple mutants, APP/PS1 overexpressing sterol regulatory element-binding protein-2 (SREP-2), also exhibit combined Aβ accumulation and tau pathology, opening the way to examine the contribution of altered lipid metabolism (mitochondrial cholesterol loading and glutathione depletion), in the onset of AD-like pathological alterations in ADAD, with their interplay [229]. Together with this, Platt's group reported, just knock-in of human β-secretase (BACE1) in mouse cleaves murine APP and initiates amyloid pathogenesis. Thus, processing of murine APP in this mouse model resulted in the formation of toxic APP compounds that accumulated intra- and extraneuronally in hippocampus and cortex. Heightened levels of inflammation (gliosis) also appeared in several AD-related brain regions at 6 and 12 months of age [383]. The progressive incorporation of new advanced genome-editing techniques at any stage of the animal lifetime or disease stage [359] is expected to open the way to new next generation disease models and more refined, hypothesis-driven studies.

Finally, an important breakthrough has been the first demonstration of a supra-cellular human model of AD: In a relevant study, researchers led by Doo Yeon Kim and Rudolph Tanzi at Harvard described a three-dimensional human neural cell culture model of ADAD recapitulating spontaneous formation of both Aβ plaques and neurofibrillary tangles in human cultured cells [230]. This confirms and extends the results of Shi and coworkers [209]. The results strongly suggest that Aβ aggregation can directly lead to tau aggregation, and that inflammation may not be essential, or first causative to AD pathology. They also showed that treatment of the cultures with β- or γ-secretase inhibitors can dramatically reduce amyloid-β pathology and attenuate tauopathy. This represents a good candidate for a model with which to study how Aβ may drive tau pathology, and to potentially test new treatments targeting Aβ, tau or other targets [230] (but see some limitations of neurons in culture above, and http://www.alzforum.org/news/research-news/alzheimers-dish-av-stokes-tau-pathology-third-dimension).

At this point, with an integrative perspective and a new strategy, Systems Biology, with advanced molecular, experimental and computational approaches and disease models recapitulating AD features, all this potential opens the way to the formulation of new questions and hypothesis towards basic and translational (applicable

to patient) discoveries. Some challenging questions (the reader may think of better ones) are:

1. Whereas main hallmarks of AD have been recapitulated in animal models and human cell cultures in weeks/months (see above), these hallmarks take years, even decades to appear in ADAD, DS and LOAD individuals. What homeostatic mechanisms and networks are readily mobilized in human, effectively counteracting AD as a well-established proteinopathy? The fine balance, dynamic interplay between impaired and homeostatic networks, for intrinsic experimental reasons, will be difficult to study in animal models and cell lines. For a steady progress, this calls for time-course experiments in simple models and, ultimately in well-designed longitudinal studies with ADAD [301, 302], DS [153, 384–386] and high risk LOAD individuals.
2. Promising results showing early intervention may delay or reverse toxic effects in AD (Table 2) need to be validated, sustained with more evidence, in well-designed studies from simple disease models to humans.
3. Reported benefits of lifestyle modifiable factors (e.g., exercise, diet, active lifestyle) [77, 299] need to be confirmed, their underlying mechanisms unveiled (a few examples in Table 2), to support Public Health initiatives and incentives. Are signaling, activation and/or potentiation of homeostatic networks (e.g., proteostasis, Wnt signaling, REST protein and, hexosamine pathways) underlying any of their benefits?

5 Conclusions: Future perspectives

Alzheimer's disease is a complex multifactorial disease starting decades before the appearance of first cognitive symptoms [78–81, 265]. This evidence emphasizes the urgent need to study the earliest impairments and interplay of networks at the 'core' of the disease, at the cellular level (and beyond).

In this work, an integrative perspective has been presented. Alzheimer's disease was characterized to result from: (a) intrinsic genomic/epigenomic susceptibility and, (b) a continued fine balance, dynamic interplay, between impaired and central homeostatic networks, mobilized since the earliest stages of the disease. The combined, unique interplay of networks will underlie the specific onset and rate of progression of the disease of an individual, whose proper investigation requires new integrative Systems Biology approaches.

Systems Biology experimental and computational approaches in simple model organisms recapitulating basic AD features offer the potential to dissect and unveil basic mechanisms, short and

long-term effects of perturbations and dysregulations and their interplay with homeostatic networks at the fundamental cellular level, which may illuminate the onset and sequence of events underlying AD. This rationale may be reproduced in more complex model systems (e.g., mammalian cell lines, patient-specific cell lines, animal models) and ultimately in human longitudinal studies. The latest studies of commonly assumed 'untreatable' proteinopathies such as AD are beginning to reveal promising results (Table 2).

In practice, main efforts should aim at advances on risk classification, earlier diagnostics, and timely and tailored interventions. Next Generation Systems Biology experimental and computational approaches, individually and combined, are already delivering remarkable insights on basic AD genomic/epigenomic susceptibility (e.g., [122, 366] and Table 1) towards risk classification of individuals. With the progressive incorporation of experimental and computational systems biology approaches such as those presented in this volume, new rational strategies towards early diagnosis well before apoptotic cell death may begin to be envisaged. Thus, we expect that early detection and monitoring of the presence and/or co-occurrence of, among others: (a) continued imbalances and generation of toxic Aβ oligomers and tau species (e.g., aggregates, tangles); (b) altered pre-mRNA splicing and noncoding RNA networks; (c) biomarkers of impaired proteome and metabolic networks and redox imbalances, together with (d) compromised, progressively affected or declining homeostatic capacities, may constitute an initial strategy towards the achievement of reliable diagnostics at the earliest asymptomatic disease stages.

Together with this, candidate tailored interventions (preventive in lifestyle, and/or therapeutic) may be investigated. First, accumulative evidence of benefits of protective, modifiable factors (e.g., diet, exercise, enriched environment, active lifestyle) during lifetime [77, 299] should be confirmed mechanistically, to support Health Public initiatives and incentives, and expedite their implementation. Also, there is a need to avoid toxic perturbations, such as sustained stress, since (without time to recover) the homeostatic responses become overwhelmed such that damage accumulates and apoptosis or necrosis results (*see* Fig. 1 and examples and mechanisms in Table 2). In all, only multidisciplinary worldwide collaborations will allow us to progress from Systems Biology to Translational Systems Medicine and Public Health. As a research community, our responsibility lies in delivering excellent results and outputs to advance both basic knowledge and its translational impact for the benefit of society.

Acknowledgements

This work was supported by BBSRC grants BB/C505140/2 and BB/F00446X/1 as well as by a contract from the European Commission under the FP7 Collaborative Programme, UNICELLSYS (all to SGO). Current work on AD in the SGO laboratory is supported by the Wellcome Trust and MRC (grant code: 089703/Z/09/Z). JIC is the beneficiary of a senior post-doctoral aid program 2011 (mode A) of Bizkaia:Xede Foundation in collaboration with GENETADI Biotech Laboratories (Derio, Bizkaia, Spain).

References

1. Lander ES, Linton LM, Birren B, et al - International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921
2. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351
3. Human genome project completion report (2003) http://www.genome.gov/11006929
4. Weiss K (2012) What is the human genome? The Scientist. 17 Aug 2012. http://www.the-scientist.com/?articles.view/articleNo/32446/title/Opinion-What-Is-the-Human-Genome--/
5. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073, http://www.1000genomes.org/
6. Chen G, Wang C, Shi L et al (2013) Comprehensively identifying and characterizing the missing gene sequences in human reference genome with integrated analytic approaches. Hum Genet 132:899–911
7. Heard E, Tishkoff S, Todd JA et al (2010) Ten years of genetics and genomics: what have we achieved and where are we heading? Nat Rev Genet 11:723–733
8. Alberts B (2011) Lessons from genomics. Science 331:511
9. Hayden EC (2010) Human genome at ten: life is complicated. Nature 464:664–667
10. Marshall E (2011) Human genome 10th anniversary. Waiting for the revolution. Science 331:526–529
11. Vidal M, Cusick ME, Barabási AL (2011) Interactome networks and human disease. Cell 144:986–998
12. Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. Annu Rev Med 63:35–61
13. Castrillo JI, Pir P, Oliver SG (2013) Yeast Systems Biology: towards a systems understanding of regulation of eukaryotic networks in complex diseases and biotechnology. In: Walhout M, Vidal M, Dekker J (eds) Handbook of systems biology. Elsevier, New York, pp 343–365
14. Human Genome Organisation (2003) HUGO—a UN for the human genome. Nat Genetics 34:115–116 (http://www.hugo-international.org)
15. Collins, FS, Morgan M, Patrinos A (2003–) The human genome project: lessons from large-scale biology. Science 300: 286–290 (http://www.genome.gov/HGP10/)
16. Pang CP, Baum L, Lam DS (2000) Hunting for disease genes in multi-functional diseases. Clin Chem Lab Med 38:819–825
17. Castrillo JI, Oliver SG (2011) Yeast Systems Biology: the challenge of eukaryotic complexity. Methods Mol Biol 759:3–28
18. Castrillo JI, Oliver SG (2014) Yeast as a model for Systems Biology studies on complex diseases. In: Nowrousian M (ed) (Karl Essex, editor-in-chief) The Mycota XIII. A comprehensive treatise on experimental systems for basic and applied research. Chapter 1. Springer, New York, pp 1–28
19. Casals F, Idaghdour Y, Hussin J et al (2012) Next-generation sequencing approaches for genetic mapping of complex diseases. J Neuroimmunol 248:10–22
20. Ku CS, Cooper DN, Polychronakos C et al (2012) Exome sequencing: dual role as a discovery and diagnostic tool. Ann Neurol 71:5–14
21. Sheridan C (2014) Illumina claims $1,000 genome win. Nat Biotechnol 32:115
22. Wu Z, Hu Y, Melton PE (2014) Longitudinal data analysis for genetic studies in the whole-genome sequencing era. Genet Epidemiol 38(Suppl 1):S74–S80

23. Choi M, Scholl UI, Ji W et al (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106:19096–19101
24. Davies K (2010) Hugh Rienhoff's voyage round his daughter's DNA. Bio-IT World Sept–Oct 2010. http://www.bio-itworld.com/2010/issues/sept-oct/rienhoff.html
25. Mayer AN, Dimmock DP, Arca MJ et al (2011) A timely arrival for genomic medicine. Genet Med 13:195–196
26. Worthey EA, Mayer AN, Syverson GD et al (2011) Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 13:255–262
27. Auffray C, Caulfield T, Khoury MJ et al (2012) Looking back at genomic medicine in 2011. Genome Med 4:9
28. Calvo SE, Compton AG, Hershman SG et al (2012) Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Sci Transl Med 4:118ra10
29. Gilissen C, Hoischen A, Brunner HG et al (2012) Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20:490–497
30. Li MX, Gui HS, Kwan JS et al (2012) A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res 40:e53
31. Puffenberger EG, Jinks RN, Sougnez C (2012) Genetic mapping and exome sequencing identify variants associated with five novel diseases. PLoS One 7:e28936
32. Evans JP, Berg JS, Olshan AF et al (2013) We screen newborns, don't we?: realizing the promise of public health genomics. Genet Med 15:332–334
33. Maher B (2013) Father's genetic quest pays off. Nature 498:418–419
34. Young E (2013) We gained hope. The story of Lilly Grossman's genome. Phenomena National Geographic. http://phenomena.nationalgeographic.com/2013/03/11/we-gained-hope-the-story-of-lilly-grossmans-genome/
35. Goh A (2013) Hunting down culprit genes. Bio-IT World June 2013. http://www.bio-itworld.com/2013/6/7/hunting-down-culprit-genes.html
36. Goldstein DB, Allen A, Keebler J et al (2013) Sequencing studies in human genetics: design and interpretation. Nat Rev Genet 14:460–470
37. Vidal M (2009) A unifying view of 21st century systems biology. FEBS Lett 583:3891–3894
38. Friend SH (2010) The need for precompetitive integrative bionetwork disease model building. Clin Pharmacol Ther 87:536–539
39. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
40. Nurse P (2003) The great ideas of biology. Clin Med 3:560–568
41. Nurse P, Hayles J (2011) The cell in an era of systems biology. Cell 144:850–854
42. Noble D (2006) The music of life: biology beyond the genome. Oxford University Press, Oxford
43. Aitman TJ, Boone C, Churchill GA et al (2011) The future of model organisms in human disease research. Nat Rev Genet 12:575–582
44. Arkin AP, Schaffer DV (2011) Network news: innovations in 21st century systems biology. Cell 144:844–849
45. Kruger RP (2011) Systems biology. Cell 144(827):829
46. Walhout AJM, Aebersold R, Meyer T et al (2011) Systems biology: what's the next challenge? Cell 144:837–838, http://www.sciencedirect.com/science/article/pii/S0092867411002285
47. Hampel H, Lista S (2012) Alzheimer disease: from inherited to sporadic AD-crossing the biomarker bridge. Nat Rev Neurol 8:598–600
48. Barrenas F, Chavali S, Holme P et al (2009) Network properties of complex human disease genes identified through genome-wide association studies. PLoS One 4:e8090
49. Agarwal S, Deane CM, Porter MA et al (2010) Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLoS Comput Biol 6:e1000817
50. Ulitsky I, Krishnamurthy A, Karp RM et al (2010) DEGAS: de novo discovery of dysregulated pathways in human diseases. PLoS One 5:e13367
51. Kim YA, Wuchty S, Przytycka TM (2011) Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput Biol 7:e1001095
52. Chang X, Xu T, Li Y et al (2013) Dynamic modular architecture of protein-protein interaction networks beyond the dichotomy of 'date' and 'party' hubs. Sci Rep 3:1691
53. Gstaiger M, Aebersold R (2013) Genotype-phenotype relationships in light of a modular protein interaction landscape. Mol Biosyst 9:1064–1067
54. Ryan CJ, Krogan NJ, Cunningham P, Cagney G (2013) All or nothing: protein complexes flip essentiality between distantly related eukaryotes. Genome Biol Evol 5:1049–1059
55. Song J, Singh M (2013) From hub proteins to hub modules: the relationship between essentiality and centrality in the yeast interac-

tome at different scales of organization. PLoS Comput Biol 9:e1002910
56. Chen R, Mias GI, Li-Pook-Than J et al (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148:1293–1307
57. Hood L, Flores M (2012) A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. N Biotechnol 29:613–624
58. Gibbs WW (2014) Medicine gets up close and personal. Nature 506:144–145
59. Lausted C, Lee I, Zhou Y et al (2014) Systems approach to neurodegenerative disease biomarker discovery. Annu Rev Pharmacol Toxicol 54:457–481
60. Snyder M (2014) iPOP and its role in participatory medicine. Genome Med 6:6
61. Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. Nat Rev Genet 13:260–270
62. Holliday R (1995) Understanding ageing. Cambridge University Press, Cambridge
63. López-Otín C, Blasco MA, Partridge L et al (2013) The hallmarks of aging. Cell 153: 1194–1217
64. Waldera-Lupa DM, Kalfalah F, Florea A-M et al (2014) Proteome-wide analysis reveals an age-associated cellular phenotype of in situ aged human fibroblasts. Aging 6:856–878
65. Dai L, Vorselen D, Korolev KS et al (2012) Generic indicators for loss of resilience before a tipping point leading to population collapse. Science 336:1175–1177
66. Dai L, Korolev KS, Gore J (2013) Slower recovery in space before collapse of connected populations. Nature 496:355–358
67. Carpenter SR (2013) Complex systems: spatial signatures of resilience. Nature 496: 308–309
68. The Integrative HMP (iHMP) Research Network Consortium (2014) The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe 16:276–289, iHMP, http://hmp2.org
69. Suhre K, Shin SY, Petersen AK et al (2011) Human metabolic individuality in biomedical and pharmaceutical research. Nature 477:54–60
70. Darwin C, Wallace AR (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. J Proc Linn Soc Lond Zool 3:45–62
71. Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London
72. Dobzhansky T (1964) Biology, molecular and organismic. Am Zool 4:443–452
73. Breitenbach M, Jazwinski SM, Laun P (2012) Aging research in yeast, vol 57, Subcellular biochemistry. Springer, New York
74. Castrillo JI, Oliver SG (2011) Yeast systems biology. Methods and protocols, vol 759, Methods in molecular biology (MiMB Series. Editor-in-chief. Prof. John M. Walker). Humana Press. Springer, New York
75. Castrillo JI, Zeef LA, Hoyle DC et al (2007) Growth control of the eukaryote cell: a systems biology study in yeast. J Biol 6:4
76. Ballard C, Gauthier S, Corbett A et al (2011) Alzheimer's disease. Lancet 377:1019–1031
77. World Alzheimer Report (2014) Dementia and risk reduction: an analysis of protective and modifiable factors. ADI. http://www.alz.co.uk/research/world-report-2014
78. Braak H, Braak E, Bohl J, Bratzke H (1998) Evolution of Alzheimer's disease related cortical lesions. J Neural Transm Suppl 54:97–106
79. Yankner BA, Lu T, Loerch P (2008) The aging brain. Annu Rev Pathol 3:41–66
80. Bateman RJ, Xiong C, Benzinger TL et al. Dominantly Inherited Alzheimer Network (2012) Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 367:795–804
81. Huang Y, Mucke L (2012) Alzheimer mechanisms and therapeutic strategies. Cell 148:1204–1222
82. World Health Organization (2012) Dementia: a public health priority. World Health Organization and Alzheimer's Disease International, Geneva, http://www.who.int/mental_health/publications/dementia_report_2012/en/
83. Alzheimer's disease (2011) Nature 475(7555 Suppl):S1–S42. http://www.nature.com/nature/outlook/alzheimers/
84. Prince M, Bryce R, Ferri C (2011) World Alzheimer Report 2011. The benefits of early diagnosis and intervention. Alzheimer's disease International (ADI). http://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf
85. Selkoe DJ (2011) Alzheimer's disease. Cold Spring Harb Perspect Biol 3. pii: a004457. http://www.ncbi.nlm.nih.gov/pubmed/21576255
86. Selkoe DJ (2013) SnapShot: pathobiology of Alzheimer's disease. Cell 154:468–468.e1
87. Selkoe DJ, Mandelkow E, Holtzman D (2011) The biology of Alzheimer disease, Cold Spring Harbor perspectives in medicine. Cold Spring Harbor Laboratory Press, New York

88. Guerreiro R, Brás J, Hardy J (2013) SnapShot: genetics of Alzheimer's disease. Cell 155:968–968.e1
89. Mukhopadhyay R (2013) The quiet creep of Alzheimer's disease. ASBMB Today 12(4): 21–25, http://www.asbmb.org/asbmbtoday/asbmbtoday_article.aspx?id=33129
90. Hooper NM (2000) Alzheimer's disease: methods and protocols, vol 32, Methods in molecular medicine series. Editor-in-chief Prof. J. M. Walker. Humana Press, Totowa, NJ
91. Dawbarn D, Allen SJ (2007) Neurobiology of Alzheimer's disease. Oxford University Press, Oxford
92. Sisodia SS, Tanzi RE (2007) Alzheimer's disease: advances in genetics, molecular and cellular biology. Springer, New York
93. Budson AE, Kowall NW (2011) The handbook of Alzheimer's disease and other dementias. Willey-Blackwell, Hoboken, NJ
94. Eisenstein M (2011) Genetics: finding risk factors. Nature 475:S20–S22, http://www.nature.com/nature/journal/v475/n7355_supp/full/475S20a.html
95. Griffin WS (2011) Alzheimer's - looking beyond plaques. F1000 Med Rep 3:24
96. Roberson ER (2011) Alzheimer's disease and frontotemporal dementia. Methods and protocols, vol 670, Methods in molecular biology. Editor-in-chief Prof. J. M. Walker. Humana Press. Springer, New York
97. Hampel H, Carrillo MC (2012) Alzheimer's disease. Modernizing concept, biological diagnosis and therapy, vol 28, Advances in biological psychiatry (Series Editors. K. P. Ebmeir, W. F. Gattaz and W. P. Kaschka. Karger, Basel
98. Hyman BT, Phelps CH, Beach TG et al (2012) National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 8:1–13
99. Institute of Medicine USA (2012) Forum on neuroscience and nervous system disorders. Alzheimer's diagnostic guideline validation. Exploration of next steps. Workshop summary. National Academies Press, Washington, DC, http://www.nap.edu/catalog.php?record_id=13312
100. Lowe VJ, Peller PJ, Weigand SD et al (2013) Application of the National Institute on Aging-Alzheimer's Association AD criteria to ADNI. Neurology 80:2130–2137
101. Chertkow H, Feldman HH, Jacova C, Massoud F (2013) Definitions of dementia and predementia states in Alzheimer's disease and vascular cognitive impairment: consensus from the Canadian conference on diagnosis of dementia. Alzheimers Res Ther 5(Suppl 1):S2
102. Patterson C, Gauthier S (2013) Diagnosis and treatment of dementia: the fourth Canadian consensus conference. Alzheimers Res Ther 5(Suppl 1):S1, http://alzres.com/content/5/Suppl%201/S1
103. Jensen M, Cox AP, Chaudhry N et al (2013) The neurological disease ontology. J Biomed Semantics 4:42
104. Dubois B, Feldman HH, Jacova C et al (2014) Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 13:614–629
105. Cohen AD, Klunk WE (2014) Early detection of Alzheimer's disease using PiB and FDG PET. Neurobiol Dis. pii: S0969-9961(14)00110-7. doi:10.1016/j.nbd.2014.05.001
106. Perry G, Zhu X, Smith MA et al (2013) Alzheimer's disease: advances for a new century, vol 3, Advances in Alzheimer's disease. Ios Press, Amsterdam
107. Benitez BA, Jin SC, Guerreiro R et al (2014) Missense variant in TREML2 protects against Alzheimer's disease. Neurobiol Aging 35:1510.e19–1510.e26
108. Shen L, Thompson PM, Potkin SG et al - for the Alzheimer's Disease Neuroimaging Initiative (2014) Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. Brain Imaging Behav 8:183–207
109. Thies W, Bleiler L, Alzheimer's Association (2013) Alzheimer's disease facts and figures. Alzheimers Dement 9:208–245, http://download.journals.elsevierhealth.com/pdfs/journals/1552-5260/PIIS1552526013000769.pdf
110. Alzheimer's Association (2014) Alzheimer's disease facts and figures. Alzheimers Dement 10(2):e47–e92, http://www.alzheimersanddementia.com/article/S1552-5260(14)00062-4/pdf
111. World Alzheimer Reports (2009–2014) Alzheimer's disease International (ADI). http://www.alz.co.uk/research/world-report
112. Noorbakhsh F, Overall CM, Power C (2009) Deciphering complex mechanisms in neurodegenerative diseases: the advent of systems biology. Trends Neurosci 32:88–100
113. Juhász G, Földi I, Penke B (2011) Systems biology of Alzheimer's disease: how diverse molecular changes result in memory impairment in AD. Neurochem Int 58:739–750
114. Walhout AJM, Vidal M, Dekker J (eds) (2013) Handbook of systems biology. Wiley-Blackwell, New York
115. Hunter S, Arendt T, Brayne C (2013) The senescence hypothesis of disease progression

in Alzheimer disease: an integrated matrix of disease pathways for FAD and SAD. Mol Neurobiol 48:556–570
116. Mei H, Xia T, Feng G, Zhu J et al (2012) Opportunities in systems biology to discover mechanisms and repurpose drugs for CNS diseases. Drug Discov Today 171:1208–1216
117. St George-Hyslop PH, Petit A (2005) Molecular biology and genetics of Alzheimer's disease. C R Biol 328:119–130
118. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23
119. Hollingworth P, Williams J (2011) Genetic risk factors for dementia. In: Budson AE, Kowall NW (eds) The handbook of Alzheimer's disease and other dementias. Wiley-Blackwell, New York, pp 197–234
120. Jin SC, Pastor P, Cooper B et al. Ibero-American Alzheimer Disease Genetics Group Researchers, Cruchaga C (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res Ther 4:34
121. Jonsson T, Atwal JK, Steinberg S et al (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96–99
122. Cruchaga C, Haller G, Chakraverty S et al. NIA-LOAD/NCRAD Family Study Consortium (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset alzheimer's disease families. PLoS One 7:e31039
123. Hooli BV, Kovacs-Vajna ZM, Mullin K et al (2014) Rare autosomal copy number variations in early-onset familial Alzheimer's disease. Mol Psychiatry 19:676–681
124. Morgan K, Carrasquillo MM (2013) Genetic variants in Alzheimer's disease. Springer, New York
125. Ridge PG, Mukherjee S, Crane PK et al. Alzheimer's Disease Genetics Consortium. (2013) Alzheimer's disease: analyzing the missing heritability. PLoS One 8:e79771
126. Szigeti K, Lal D, Li Y et al. Texas Alzheimer Research and Care Consortium (2013) Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease. J Alzheimers Dis 33:517–523
127. Mak HK, Qian W, Ng KS et al (2014) Combination of MRI hippocampal volumetry and arterial spin labeling MR perfusion at 3-Tesla improves the efficacy in discriminating Alzheimer's disease from cognitively normal elderly adults. J Alzheimers Dis 41:749–758
128. Karch CM, Cruchaga C, Goate AM (2014) Alzheimer's disease genetics: from the bench to the clinic. Neuron 83:11–26
129. Koutras C, Braun JE (2014) J protein mutations and resulting proteostasis collapse. Front Cell Neurosci 8:191. doi:10.3389/fncel.2014.00191
130. Mizuno S, Iijima R, Ogishima S et al (2012) AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst Biol 6:52
131. Ogishima S, Mizuno S, Kikuchi M et al (2013) A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery. Clin Pharmacol Ther 93:399–401
132. O'Brien RJ, Wong PC (2011) Amyloid precursor protein processing and Alzheimer's disease. Annu Rev Neurosci 34:185–204
133. Cowan CM, Mudher A (2013) Are tau aggregates toxic or protective in tauopathies? Front Neurol 4:114
134. Hayden EY, Teplow DB (2013) Amyloid β-protein oligomers and Alzheimer's disease. Alzheimers Res Ther 5:60
135. Abisambra JF, Jinwal UK, Blair LJ et al (2013) Tau accumulation activates the unfolded protein response by impairing endoplasmic reticulum-associated degradation. J Neurosci 33:9498–9507
136. Tharp WG, Sarkar IN (2013) Origins of amyloid-β. BMC Genomics 14:290
137. Avila J, Simon D, Diaz-Hernandez M et al (2014) Sources of extracellular tau and its signaling. J Alzheimers Dis 40(Suppl 1):S7–S15
138. Medina M, Avila J (2014) New perspectives on the role of tau in Alzheimer's disease. Implications for therapy. Biochem Pharmacol 88:540–547
139. Michel CH, Kumar S, Pinotsi D et al (2014) Extracellular monomeric tau protein is sufficient to initiate the spread of tau protein pathology. J Biol Chem 289:956–967
140. Pinotsi D, Buell AK, Galvagnion C et al (2014) Direct observation of heterogeneous amyloid fibril growth kinetics via two-color super-resolution microscopy. Nano Lett 14: 339–345
141. Sahara N, Avila J (2014) "Tau oligomers," what we know and what we don't know. Front Neurol 5:1. doi:10.3389/fneur.2014.00001
142. Ittner LM, Götz J (2011) Amyloid-β and tau–a toxic pas de deux in Alzheimer's disease. Nat Rev Neurosci 12:65–72
143. Hunter S, Brayne C (2012) Relationships between the amyloid precursor protein and its various proteolytic fragments and neuronal systems. Alzheimers Res Ther 4:10
144. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress

and problems on the road to therapeutics. Science 297:353–356
145. Teplow DB (2013) On the subject of rigor in the study of amyloid β-protein assembly. Alzheimers Res Ther 5:39
146. Tan JL, Li QX, Ciccotosto GD et al (2013) Mild oxidative stress induces redistribution of BACE1 in non-apoptotic conditions and promotes the amyloidogenic processing of Alzheimer's disease amyloid precursor protein. PLoS One 8:e61246
147. Drachman DA (2014) The amyloid hypothesis, time to move on: amyloid is the downstream result, not cause, of Alzheimer's disease. Alzheimers Dement 10:372–380
148. Lin MT, Beal MF (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443:787–795
149. Patten DA, Germain M, Kelly MA et al (2010) Reactive oxygen species: stuck in the middle of neurodegeneration. J Alzheimers Dis 20(Suppl 2):S357–S367
150. Singh M, Dang TN, Arseneault M et al (2010) Role of by-products of lipid oxidation in Alzheimer's disease brain: a focus on acrolein. J Alzheimers Dis 21:741–756
151. Butterfield DA, Perluigi M, Reed T et al (2012) Redox proteomics in selected neurodegenerative disorders: from its infancy to future applications. Antioxid Redox Signal 17:1610–1655
152. Di Domenico F, Coccia R, Butterfield DA et al (2011) Circulating biomarkers of protein oxidation for Alzheimer disease: expectations within limits. Biochim Biophys Acta 1814: 1785–1795
153. Di Domenico F, Coccia R, Cocciolo A et al (2013) Impairment of proteostasis network in Down syndrome prior to the development of Alzheimer's disease neuropathology: redox proteomics analysis of human brain. Biochim Biophys Acta 1832:1249–1259
154. Pimentel C, Batista-Nascimento L, Rodrigues-Pousada C et al (2012) Oxidative stress in Alzheimer's and Parkinson's diseases: insights from the yeast *Saccharomyces cerevisiae*. Oxid Med Cell Longev 2012:132146
155. Swomley AM, Förster S, Keeney JT et al (2014) Abeta, oxidative stress in Alzheimer disease: evidence based on proteomics studies. Biochim Biophys Acta 1842:1248–1257
156. Kell DB (2010) Towards a unifying, systems biology understanding of large-scale cellular death and destruction caused by poorly liganded iron: Parkinson's, Huntington's, Alzheimer's, prions, bactericides, chemical toxicology and others as examples. Arch Toxicol 84:825–889
157. Copenhaver PF, Anekonda TS, Musashe D et al (2011) A translational continuum of model systems for evaluating treatment strategies in Alzheimer's disease: isradipine as a candidate drug. Dis Model Mech 4:634–648
158. Mukherjee A, Soto C (2011) Role of calcineurin in neurodegeneration produced by misfolded proteins and endoplasmic reticulum stress. Curr Opin Cell Biol 23:223–230
159. Roberts BR, Ryan TM, Bush AI et al (2012) The role of metallobiology and amyloid-beta peptides in Alzheimer's disease. J Neurochem 120(Suppl 1):149–166
160. Sarell CJ, Wilkinson SR, Viles JH (2010) Substoichiometric levels of Cu2+ ions accelerate the kinetics of fiber formation and promote cell toxicity of amyloid-{beta} from Alzheimer disease. J Biol Chem 285:41533–44150
161. Torres M, Encina G, Soto C et al (2011) Abnormal calcium homeostasis and protein folding stress at the ER: a common factor in familial and infectious prion disorders. Commun Integr Biol 4:258–261
162. Twine NA, Janitz K, Wilkins MR et al (2011) Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 6:e16266
163. Mills JD, Janitz M (2012) Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. Neurobiol Aging 33:1012.e11–1012.e24
164. Bai B, Hales CM, Chen PC et al (2013) U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. Proc Natl Acad Sci U S A 110: 16562–16567
165. Hales CM, Seyfried NT, Dammer EB et al (2014) U1 small nuclear ribonucleoproteins (snRNPs) aggregate in Alzheimer's disease due to autosomal dominant genetic mutations and trisomy 21. Mol Neurodegener 9:15
166. Hales CM, Dammer EB, Diner I et al (2014) Aggregates of small nuclear ribonucleic acids (snRNAs) in Alzheimer's disease. Brain Pathol 24:344–351
167. Eckert A, Schmitt K, Götz J (2011) Mitochondrial dysfunction - the beginning of the end in Alzheimer's disease? Separate and synergistic modes of tau and amyloid-beta toxicity. Alzheimers Res Ther 3:15
168. Eckert A, Nisbet R, Grimm A, Götz J (2014) March separate, strike together–role of phosphorylated tau in mitochondrial dysfunction in Alzheimer's disease. Biochim Biophys Acta 1842:1258–1266
169. Schon EA, Przedborski S (2011) Mitochondria: the next (neurode)generation. Neuron 70:1033–1053
170. Karbowski M, Neutzner A (2012) Neurodegeneration as a consequence of failed mitochondrial maintenance. Acta Neuropathol 123:157–171

171. Cohen E, Bieschke J, Perciavalle RM et al (2006) Opposing activities protect against age-onset proteotoxicity. Science 313:1604–1610
172. Murray AN, Solomon JP, Wang YJ et al (2010) Discovery and characterization of a mammalian amyloid disaggregation activity. Protein Sci 19:836–846
173. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332
174. Anderson JF, Siller E, Barral JM (2011) Disorders of protein biogenesis and stability. Protein Pept Lett 18:110–121
175. Araki K, Nagata K (2011) Protein folding and quality control in the ER. Cold Spring Harb Perspect Biol 3:a007526
176. Braakman I, Bulleid NJ (2011) Protein folding and modification in the mammalian endoplasmic reticulum. Annu Rev Biochem 80:71–99
177. Christianson JC, Olzmann JA, Shaler TA et al (2011) Defining human ERAD networks through an integrative mapping strategy. Nat Cell Biol 14:93–105
178. Smith MH, Ploegh HL, Weissman JS (2011) Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. Science 334:1086–1090
179. Walter P, Ron D (2011) The unfolded protein response: from stress pathway to homeostatic regulation. Science 334:1081–1086
180. Verghese J, Abrams J, Wang Y et al (2012) Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. Microbiol Mol Biol Rev 76:115–158
181. Lee MJ, Lee JH, Rubinsztein DC (2013) Tau degradation: the ubiquitin-proteasome system versus the autophagy-lysosome system. Prog Neurobiol 105:49–59
182. Ocampo A, Liu J, Barrientos A (2013) NAD+ salvage pathway proteins suppress proteotoxicity in yeast models of neurodegeneration by promoting the clearance of misfolded/oligomerized proteins. Hum Mol Genet 22:1699–1708
183. Denzel MS, Storm NJ, Gutschmidt A et al (2014) Hexosamine pathway metabolites enhance protein quality control and prolong life. Cell 156:1167–1178
184. Lu T, Aron L, Zullo J et al (2014) REST and stress resistance in ageing and Alzheimer's disease. Nature 507:448–454
185. Vincenz L, Hartl FU (2014) Sugarcoating ER stress. Cell 156:1125–1127
186. Wang ZV, Deng Y, Gao N et al (2014) Spliced X-Box binding protein 1 couples the unfolded protein response to hexosamine biosynthetic pathway. Cell 156:1179–1192
187. Narayan P, Ehsani S, Lindquist S (2014) Combating neurodegenerative disease with chemical probes and model systems. Nat Chem Biol 10:911–920
188. Rubinsztein DC, Codogno P, Levine B (2012) Autophagy modulation as a potential therapeutic target for diverse diseases. Nat Rev Drug Discov 11:709–730
189. Murrow L, Debnath J (2013) Autophagy as a stress-response and quality-control mechanism: implications for cell injury and human disease. Annu Rev Pathol 8:105–137
190. Nixon RA (2013) The role of autophagy in neurodegenerative disease. Nat Med 19:983–997
191. Puri C, Renna M, Bento CF et al (2013) Diverse autophagosome membrane sources coalesce in recycling endosomes. Cell 154:1285–1299
192. Prusiner SB (2012) Cell biology. A unifying role for prions in neurodegenerative diseases. Science 336:1511–1513
193. Stöhr J, Watts JC, Mensinger ZL et al (2012) Purified and synthetic Alzheimer's amyloid beta (Aβ) prions. Proc Natl Acad Sci U S A 109:11025–11030
194. Yong E (2012) Corrupted proteins spread disease. http://the-scientist.com/2012/06/18/corrupted-proteins-spread-disease/
195. Bateman RJ, Munsell LY, Morris JC et al (2006) Human amyloid-beta synthesis and clearance rates as measured in cerebrospinal fluid in vivo. Nat Med 12:856–861
196. Mawuenyega KG, Sigurdson W, Ovod V et al (2010) Decreased clearance of CNS beta-amyloid in Alzheimer's disease. Science 330:1774
197. Iliff JJ, Wang M, Liao Y et al (2012) A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. Sci Transl Med 4:147ra111
198. Griffin WS, Stanley LC, Ling C et al (1989) Brain interleukin 1 and S-100 immunoreactivity are elevated in Down syndrome and Alzheimer disease. Proc Natl Acad Sci U S A 86:7611–7615
199. Zotova E, Nicoll JA, Kalaria R et al (2010) Inflammation in Alzheimer's disease: relevance to pathogenesis and therapy. Alzheimers Res Ther 2:1
200. Masters SL, O'Neill LA (2011) Disease-associated amyloid and misfolded protein aggregates activate the inflammasome. Trends Mol Med 17:276–282
201. Nagele E, Han M, Demarshall C et al (2011) Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. PLoS One 6:e23112

202. Schnabel J (2011) Amyloid: little proteins, big clues. Nature 475:S12–S14, http://www.nature.com/nature/journal/v475/n7355_supp/full/475S12a.html
203. Alzheimer's Disease Anti-inflammatory Prevention Trial Research Group (2013) Results of a follow-up study to the randomized Alzheimer's Disease anti-inflammatory prevention trial (ADAPT). Alzheimers Dement 9:714–723
204. Gomez-Nicola D, Perry VH (2015) Microglial dynamics and role in the healthy and diseased brain: a paradigm of functional plasticity. Neuroscientist. 21:169–184
205. Bharadwaj P, Martins R, Macreadie I (2010) Yeast as a model for studying Alzheimer's disease. FEMS Yeast Res 10:961–969
206. Chen KF, Crowther DC (2014) Insights into amyloid disease from fly models. Essays Biochem 56:69–83
207. Cuadrado-Tejedor M, Ricobaraza A, Frechilla D et al (2012) Chronic mild stress accelerates the onset and progression of the Alzheimer's disease phenotype in Tg2576 mice. J Alzheimers Dis 28:567–578
208. Cohen RM, Rezai-Zadeh K, Weitz TM et al (2013) A transgenic Alzheimer rat with plaques, tau pathology, behavioral impairment, oligomeric aβ, and frank neuronal loss. J Neurosci 33:6245–6256
209. Shi Y, Kirwan P, Smith J et al (2012) A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Sci Transl Med 4:124ra29
210. Livesey FJ (2012) Stem cell models of Alzheimer's disease and related neurological disorders. Alzheimers Res Ther 4:44
211. Weick JP, Held DL, Bonadurer GF 3rd et al (2013) Deficits in human trisomy 21 iPSCs and neurons. Proc Natl Acad Sci U S A 110:9962–9967
212. Wisniewski KE, Dalton AJ, McLachlan C et al (1985) Alzheimer's disease in Down's syndrome: clinicopathologic studies. Neurology 35:957–961
213. Tardiff DF, Khurana V, Chung CY, Lindquist S (2014) From yeast to patient neurons and back again: a powerful new discovery platform. Mov Disord 29:1231–1240
214. Khurana V, Lindquist S (2010) Modelling neurodegeneration in *Saccharomyces cerevisiae*: why cook with baker's yeast? Nat Rev Neurosci 11:436–449
215. Pereira C, Bessa C, Soares J et al (2012) Contribution of yeast models to neurodegeneration research. J Biomed Biotechnol 2012: 941232
216. Porzoor A, Macreadie IG (2013) Application of yeast to study the tau and amyloid-β abnormalities of Alzheimer's disease. J Alzheimers Dis 35:217–225
217. Tenreiro S, Munder MC, Alberti S, Outeiro TF (2013) Harnessing the power of yeast to unravel the molecular basis of neurodegeneration. J Neurochem 127:438–452
218. Villar-Piqué A, Ventura S (2013) Protein aggregation propensity is a crucial determinant of intracellular inclusion formation and quality control degradation. Biochim Biophys Acta 1833:2714–2724
219. Bharadwaj PR, Verdile G, Barr RK et al (2012) Latrepirdine (dimebon) enhances autophagy and reduces intracellular GFP-Aβ42 levels in yeast. J Alzheimers Dis 32:949–967
220. De Vos A, Anandhakumar J, Van den Brande J et al (2011) Yeast as a model system to study tau biology. Int J Alzheimers Dis 2011:428970
221. Treusch S, Hamamichi S, Goodman JL et al (2011) Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. Science 334:1241–1245
222. D'Angelo F, Vignaud H, Di Martino J et al (2013) A yeast model for amyloid-β aggregation exemplifies the role of membrane trafficking and PICALM in cytotoxicity. Dis Model Mech 2013:206–216
223. Favrin G, Bean DM, Bilsland E et al (2013) Identification of novel modifiers of Aβ toxicity by transcriptomic analysis in the fruit fly. Sci Rep 3:3512
224. Hermansson E, Schultz S, Crowther D et al (2014) The chaperone domain BRICHOS prevents amyloid β-peptide CNS toxicity in Drosophila melanogaster. Dis Model Mech 7:659–665
225. Tomiyama T, Matsuyama S, Iso H et al (2010) A mouse model of amyloid beta oligomers: their contribution to synaptic alteration, abnormal tau phosphorylation, glial activation, and neuronal loss in vivo. J Neurosci 30:4845–4856
226. DaRocha-Souto B, Scotton TC, Coma M et al (2011) Brain oligomeric β-amyloid but not total amyloid plaque burden correlates with neuronal loss and astrocyte inflammatory response in amyloid precursor protein/tau transgenic mice. J Neuropathol Exp Neurol 70:360–376
227. Giménez-Llort L, Rivera-Hernández G, Marin-Argany M et al (2013) Early intervention in the 3xTg-AD mice with an amyloid β-antibody fragment ameliorates first hallmarks of Alzheimer disease. MAbs 5:665–677
228. Cairney CJ, Sanguinetti G, Ranghini E et al (2009) A systems biology approach to Down syndrome: identification of Notch/Wnt dysregulation in a model of stem cells aging. Biochim Biophys Acta 1792:353–363

229. Barbero-Camps E, Fernández A, Martínez L et al (2013) APP/PS1 mice overexpressing SREBP-2 exhibit combined Aβ accumulation and tau pathology underlying Alzheimer's disease. Hum Mol Genet 22:3460–3476
230. Choi SH, Kim YH, Hebisch M et al (2014) A three-dimensional human neural cell culture model of Alzheimer's disease. Nature. doi:10.1038/nature1380
231. Mills JD, Kawahara Y, Janitz M (2013) Strand-specific RNA-Seq provides greater resolution of transcriptome profiling. Curr Genomics 14:173–181
232. Mills JD, Nalpathamkalam T, Jacobs HI et al (2013) RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. Neurosci Lett 536:90–95
233. Gilleron J, Querbes W, Zeigerer A et al (2013) Image-based analysis of lipid nanoparticle-mediated siRNA delivery, intracellular trafficking and endosomal escape. Nat Biotechnol 31:638–646
234. Lothian A, Hare DJ, Grimm R et al (2013) Metalloproteomics: principles, challenges and applications to neurodegeneration. Front Aging Neurosci 5:35
235. Gillet LC, Navarro P, Tate S et al (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 11:O111.016717
236. Lambert JP, Ivosev G, Couzens AL et al (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. Nat Methods 10:1239–1245
237. Greco V, Pieragostino D, Piras C et al (2014) Direct analytical sample quality assessment for biomarker investigation: qualifying cerebrospinal fluid samples. Proteomics 14:1954–1962
238. Sultana R, Butterfield DA (2013) Oxidative modification of brain proteins in Alzheimer's disease: perspective on future studies based on results of redox proteomics studies. J Alzheimers Dis 33(Suppl 1):S243–S251
239. Cheng H, Wang M, Li JL et al (2013) Specific changes of sulfatide levels in individuals with pre-clinical Alzheimer's disease: an early event in disease pathogenesis. J Neurochem 127: 733–738
240. Griffiths WJ, Koal T, Wang Y et al (2010) Targeted metabolomics for biomarker discovery. Angew Chem Int Ed Engl 49:5426–5445
241. Kaddurah-Daouk R, Zhu H, Sharma S et al (2013) Alterations in metabolic pathways and networks in Alzheimer's disease. Transl Psychiatry 3:e244
242. Nibbe RK, Chowdhury SA, Koyutürk M et al (2011) Protein-protein interaction networks and subnetworks in the biology of disease. Wiley Interdiscip Rev Syst Biol Med 3:357–367
243. Carvunis A-R, Roth FP, Calderwood MA et al (2013) Interactome networks. In: Walhout M, Vidal M, Dekker J (eds) Handbook of systems biology. Elsevier, New York, pp 45–63
244. Li X, Wang X, Snyder M (2013) Systematic investigation of protein-small molecule interactions. IUBMB Life 65:2–8
245. Harari O, Cruchaga C, Kauwe JS et al. - Alzheimer's Disease Neuroimaging Initiative (2014) Phosphorylated tau-Aβ(42) ratio as a continuous trait for biomarker discovery for early-stage Alzheimer's Disease in multiplex immunoassay panels of cerebrospinal fluid. Biol Psychiatry 75:723–731.
246. Novarino G, Fenstermaker AG, Zaki MS et al (2014) Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. Science 343:506–511
247. Tanzi RE (2014) Decoding Alzheimer's in the age of genome-wide analyses. Mol Neurodegener 8(Suppl 1):O1
248. Guerreiro R, Brás J, Hardy J, Singleton A (2014) Next generation sequencing techniques in neurological diseases: redefining clinical and molecular associations. Hum Mol Genet 23(R1):R47–R53
249. Gu L, Li C, Aach J et al (2014) Multiplex single-molecule interaction profiling of DNA-barcoded proteins. Nature. doi:10.1038/nature13761
250. Soler-López M, Zanzoni A, Lluís R et al (2011) Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. Genome Res 21:364–376
251. Bar-Joseph Z, Gitter A, Simon I (2012) Studying and modelling dynamic biological processes using time-series gene expression data. Nat Rev Genet 13:552–564
252. Gitter A, Bar-Joseph Z (2013) Identifying proteins controlling key disease signalling pathways. Bioinformatics 29:i227–i236
253. Gitter A, Carmi M, Barkai N, Bar-Joseph Z (2013) Linking the signaling cascades and dynamic regulatory networks controlling stress responses. Genome Res 23:365–376
254. Zhang B, Gaiteri C, Bodea LG et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease. Cell 153:707–720
255. Kikuchi M, Ogishima S, Miyamoto T et al (2013) Identification of unstable network modules reveals disease modules associated with the progression of Alzheimer's disease. PLoS One 8:e76162

256. Gika H, Theodoridis G (2011) Sample preparation prior to the LC-MS-based metabolomics/metabonomics of blood-derived samples. Bioanalysis 3:1647–1661
257. Simpson RJJ, Greening DWW (2011) Serum/plasma proteomics: methods and protocols, vol 728, Methods in molecular biology (MiMB) series. Springer, New York
258. Tammen H, Hess R (2011) Collection and handling of blood specimens for peptidomics. Methods Mol Biol 728:151–159
259. Perrin RJ, Fagan AM, Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. Nature 461:916–922
260. Wang K, Lee I, Carlson G et al (2010) Systems biology and the discovery of diagnostic biomarkers. Dis Markers 28:199–207
261. Abu-Asab MS, Chaouchi M, Alesci S et al (2011) Biomarkers in the age of omics: time for a systems biology approach. OMICS 15:105–112
262. Li-Pook-Than J, Snyder M (2013) iPOP goes the world: integrated personalized Omics profiling and the road toward improved health care. Chem Biol 20:660–666
263. Hood L, Tian Q (2012) Systems approaches to biology and disease enable translational systems medicine. Genomics Proteomics Bioinformatics 10:181–185
264. Hood L, Auffray C (2013) Participatory medicine: a driving force for revolutionizing healthcare. Genome Med 5:110
265. Fagan AM, Xiong C, Jasielec MS et al. Dominantly Inherited Alzheimer Network (2014). Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. Sci Transl Med. 6:226ra30.
266. Acosta-Baena N, Sepulveda-Falla D, Lopera-Gómez CM et al (2011) Pre-dementia clinical stages in presenilin 1 E280A familial early-onset alzheimer's disease: a retrospective cohort study. Lancet Neurol 10:213–220
267. Breitner JC, Baker LD, Montine TJ et al (2011) Extended results of the Alzheimer's disease anti-inflammatory prevention trial. Alzheimers Dement 7:402–411
268. Buckholtz NS (2011) Perspective: in search of biomarkers. Nature 475:S8, http://www.nature.com/nature/journal/v475/n7355_supp/full/475S8a.html
269. Galvin JE (2011) Dementia screening, biomarkers and protein misfolding: implications for public health and diagnosis. Prion 5:16–21
270. Taguchi A, Politi K, Pitteri SJ et al (2011) Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. Cancer Cell 20:289–299
271. Mayeux R, Schupf N (2011) Blood-based biomarkers for Alzheimer's disease: plasma Abeta40 and Abeta42, and genetic variants. Neurobiol Aging 32(Suppl 1):S10–S19
272. Perluigi M, Butterfield DA (2011) The identification of protein biomarkers for oxidative stress in Down syndrome. Expert Rev Proteomics 8:427–429
273. Rabinovici GD, Rosen HJ, Alkalay A et al (2011) Amyloid vs FDG-PET in the differential diagnosis of AD and FTLD. Neurology 77:2034–2042
274. Liu Y, Hüttenhain R, Collins B, Aebersold R (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. Expert Rev Mol Diagn 13:811–825
275. Williams R (2011) Biomarkers: warning signs. Nature 475:S5–S7, http://www.nature.com/nature/journal/v475/n7355_supp/full/475S5a.html
276. Masdeu JC, Kreisl WC, Berman KF (2012) The neurobiology of Alzheimer disease defined by neuroimaging. Curr Opin Neurol 25:410–420
277. Sepulcre J, Sabuncu MR, Becker A et al (2013) In vivo characterization of the early states of the amyloid-beta network. Brain 136(Pt 7):2239–2252
278. Guo LH, Alexopoulos P, Wagenpfeil S et al (2013) Alzheimer's Disease Neuroimaging Initiative. Plasma proteomics for the identification of Alzheimer disease. Alzheimer Dis Assoc Disord 27:337–342
279. Perneczky R, Alexopoulos P, Kurz A (2014) Soluble amyloid precursor proteins and secretases as Alzheimer's disease biomarkers. Trends Mol Med 20:8–15
280. Mitsis EM, Riggio S, Kostakoglu L et al (2014) Tauopathy PET and amyloid PET in the diagnosis of chronic traumatic encephalopathies: studies of a retired NFL player and of a man with FTD and a severe head injury. Transl Psychiatry 4:e441
281. Yerbury JJ, Wilson MR (2010) Extracellular chaperones modulate the effects of Alzheimer's patient cerebrospinal fluid on Abeta(1-42) toxicity and uptake. Cell Stress Chaperones 15:115–121
282. Park SK, Pegan SD, Mesecar AD et al (2011) Development and validation of a yeast high-throughput screen for inhibitors of $A\beta_{42}$ oligomerization. Dis Model Mech 4:822–831
283. López LC, Dos-Reis S, Espargaró A et al (2012) Discovery of novel inhibitors of amyloid β-peptide 1-42 aggregation. J Med Chem 55:9521–9530
284. Tardiff DF, Tucci ML, Caldwell KA et al (2012) Different 8-hydroxyquinolines protect models of TDP-43 protein, α-synuclein, and polyglutamine proteotoxicity through distinct mechanisms. J Biol Chem 287: 4107–4120

285. Tardiff DF, Lindquist S (2013) Phenotypic screens for compounds that target the cellular pathologies underlying Parkinson's Disease. Drug Discov Today Technol 10:e121–e128
286. Holmes C, Boche D, Wilkinson D et al (2008) Long-term effects of Abeta42 immunisation in Alzheimer's disease: follow-up of a randomised, placebo-controlled phase I trial. Lancet 372:216–223
287. Hochfeld WE, Lee S, Rubinsztein DC (2013) Therapeutic induction of autophagy to modulate neurodegenerative disease progression. Acta Pharmacol Sin 34:600–604
288. Zhao W, Bonem M, McWhite C et al (2014) Sensitive detection of proteasomal activation using the Deg-On mammalian synthetic gene circuit. Nat Commun 5:3612
289. Lista S, Faltraco F, Prvulovic D, Hampel H (2013) Blood and plasma-based proteomic biomarker research in Alzheimer's disease. Prog Neurobiol 101–102:1–17
290. Sun Q, Hampel H, Blennow K et al (2014) Increased plasma TACE activity in subjects with mild cognitive impairment and patients with Alzheimer's disease. J Alzheimers Dis 41:877–886
291. Corbett A, Pickett J, Burns A et al (2012) Drug repositioning for Alzheimer's disease. Nat Rev Drug Discov 11:833–846
292. Appleby BS, Cummings JL (2013) Discovering new treatments for Alzheimer's disease by repurposing approved medications. Curr Top Med Chem 13:2306–2327
293. Iqbal K, Liu F, Gong CX (2014) Alzheimer disease therapeutics: focus on the disease and not just plaques and tangles. Biochem Pharmacol 88:631–639
294. Li N, Sood S, Wang S et al (2005) Overexpression of 5-lipoxygenase and cyclooxygenase 2 in hamster and human oral cancer and chemopreventive effects of zileuton and celecoxib. Clin Cancer Res 11:2089–2096
295. Valera E, Dargusch R, Maher PA, Schubert D (2013) Modulation of 5-lipoxygenase in proteotoxicity and Alzheimer's disease. J Neurosci 33:10512–10525
296. Chu J, Li J-G, Praticò D (2013) Zileuton improves memory deficits, amyloid and tau pathology in a mouse model of Alzheimer's disease with plaques and tangles. PLoS One 8:e70991
297. Li S, Jin M, Zhang D, Yang T et al (2013) Environmental novelty activates β2-adrenergic signaling to prevent the impairment of hippocampal LTP by Aβ oligomers. Neuron 77:929–941
298. Mainardi M, Di Garbo A, Caleo M et al (2014) Environmental enrichment strengthens corticocortical interactions and reduces amyloid-β oligomers in aged mice. Front Aging Neurosci 6:1
299. Norton S, Matthews FE, Barnes DE et al (2014) Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 13:788–794
300. Wood H (2014) Alzheimer disease: functional connectivity changes show similar trajectories in autosomal dominant and sporadic Alzheimer disease. Nat Rev Neurol 10:483
301. Blennow K, de Leon MJ, Zetterberg H (2006) Alzheimer's disease. Lancet 368: 387–403
302. Waring SC, Rosenberg RN (2008) Genome-wide association studies in Alzheimer disease. Arch Neurol 65:329–334
303. Selkoe DJ (1999) Translating cell biology into therapeutic advances in Alzheimer's disease. Nature 399(6738 Suppl):A23–A31
304. Fernandez, MA, Klutkowski JA, Freret T, Wolfe MS (2014) Alzheimer presenilin-1 mutations dramatically reduce trimming of long amyloid β-peptides (Aβ) by γ-secretase to increase 42-to-40-residue Aβ. J Biol Chem. jbc.M114.581165. doi:10.1074/jbc.M114.581165
305. Strittmatter WJ, Saunders AM, Schmechel D et al (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90:1977–1981
306. Alzheimer's Society UK (2007) What is Alzheimer's disease? Alzheimers.org.uk. August 2007. http://www.alzheimers.org.uk/site/scripts/documents_info.php?documentID=100
307. Rovelet-Lecrux A, Hannequin D, Raux G et al (2006) APP locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–26
308. Sherva R, Tripodis Y, Bennett DA et al (2014) Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimers Dement 10:45–52
309. Ryman DC, Acosta-Baena N, Aisen PS et al (2014) Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. Neurology 83:253–260
310. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174
311. Brickell KL, Leverenz JB, Steinbart EJ et al (2007) Clinicopathological concordance and discordance in three monozygotic twin pairs with familial Alzheimer's disease. J Neurol Neurosurg Psychiatry 78:1050–1055

312. Ketelaar ME, Hofstra EM, Hayden MR (2012) What monozygotic twins discordant for phenotype illustrate about mechanisms influencing genetic forms of neurodegeneration. Clin Genet 81:325–333
313. Lu M, Lawrence DA, Marsters S (2014) Cell death. Opposing unfolded-protein-response signals converge on death receptor 5 to control apoptosis. Science 345:98–101
314. Hernández F, Avila J (2007) Tauopathies. Cell Mol Life Sci 64:2219–2233
315. Zhang Z, Song M, Liu X et al (2014) Cleavage of tau by asparagine endopeptidase mediates the neurofibrillary pathology in Alzheimer's disease. Nat Med. doi:10.1038/nm.3700
316. Taipale M, Tucker G, Peng J et al (2014) A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. Cell 158:434–448
317. Kaganovich D, Kopito R, Frydman J (2008) Misfolded proteins partition between two distinct quality control compartments. Nature 454:1088–1095
318. Bagola K, Sommer T (2008) Protein quality control: on IPODs and other JUNQ. Curr Biol 18:R1019–R1021
319. Fonseca AC, Oliveira CR, Pereira CF et al (2014) Loss of proteostasis induced by amyloid beta peptide in brain endothelial cells. Biochim Biophys Acta 1843:1150–1161
320. Van der Harg JM, Nölle A, Zwart R et al (2014) The unfolded protein response mediates reversible tau phosphorylation induced by metabolic stress. Cell Death Dis 5:e1393
321. Takalo M, Salminen A, Soininen H et al (2013) Protein aggregation and degradation mechanisms in neurodegenerative diseases. Am J Neurodegener Dis 2:1–14
322. Hetz C, Mollereau B (2014) Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases. Nat Rev Neurosci 15: 233–249
323. Inestrosa NC, Varela-Nallar L (2014) Wnt signaling in the nervous system and in Alzheimer's disease. J Mol Cell Biol 6:64–74
324. Killick R, Ribe EM, Al-Shawi R et al (2014) Clusterin regulates β-amyloid toxicity via Dickkopf-1-driven induction of the wnt-PCP-JNK pathway. Mol Psychiatry 19:88–98
325. Roth DM, Balch WE (2011) Modeling general proteostasis: proteome balance in health and disease. Curr Opin Cell Biol 23:126–134
326. Niforou K, Cheimonidou C, Trougakos IP (2014) Molecular chaperones and proteostasis regulation during redox imbalance. Redox Biol 2:323–332
327. Christianson JC, Ye Y (2014) Cleaning up in the endoplasmic reticulum: ubiquitin in charge. Nat Struct Mol Biol 21:325–335
328. Seaman MN (2012) The retromer complex - endosomal protein recycling and beyond. J Cell Sci 125:4693–4702
329. Mecozzi VJ, Berman DE, Simoes S et al (2014) Pharmacological chaperones stabilize retromer to limit APP processing. Nat Chem Biol 10:443–449
330. Anckar J, Sistonen L (2011) Regulation of HSF1 function in the heat stress response: implications in aging and disease. Annu Rev Biochem 80:1089–1115
331. Vihervaara A, Sistonen L (2014) HSF1 at a glance. J Cell Sci 127:261–266
332. Vargas JY, Fuenzalida M, Inestrosa NC (2014) In vivo activation of Wnt signaling pathway enhances cognitive function of adult mice and reverses cognitive deficits in an Alzheimer's disease model. J Neurosci 34: 2191–2202
333. Cronk JC, Kipnis J (2013) Microglia - the brain's busy bees. F1000 Prime Rep 5:53
334. Kohlstedt M, Becker J, Wittmann C (2010) Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism. Appl Microbiol Biotechnol 88:1065–1075
335. Krstic D, Knuesel I (2013) Deciphering the mechanism underlying late-onset alzheimer disease. Nat Rev Neurol 9:25–34
336. Wyss-Coray T (2006) Inflammation in Alzheimer disease: driving force, bystander or beneficial response? Nat Med 12:1005–1015
337. Almeida CG, Takahashi RH, Gouras GK (2006) Beta-amyloid accumulation impairs multivesicular body sorting by inhibiting the ubiquitin-proteasome system. J Neurosci 26: 4277–4288
338. Mohamed NV, Plouffe V, Rémillard-Labrosse G et al (2014) Starvation and inhibition of lysosomal function increased tau secretion by primary cortical neurons. Sci Rep 4:5715
339. Frost B, Hemberg M, Lewis J, Feany MB (2014) Tau promotes neurodegeneration through global chromatin relaxation. Nat Neurosci 17:357–366
340. Kageyama S, Sou YS, Uemura T et al (2014) Proteasome dysfunction activates autophagy and the Keap1-Nrf2 Pathway. J Biol Chem. pii: jbc.M114.580357
341. Maharjan S, Oku M, Tsuda M et al (2014) Mitochondrial impairment triggers cytosolic oxidative stress and cell death following proteasome inhibition. Sci Rep 4:5896
342. Andersson V, Hanzén S, Liu B et al (2013) Enhancing protein disaggregation restores proteasome activity in aged cells. Aging (Albany NY) 5:802–812
343. Jackrel ME, Shorter J (2014) Reversing deleterious protein aggregation with

re-engineered protein disaggregases. Cell Cycle 13:1379–1383
344. Jackrel ME, Shorter J (2014b) Potentiated Hsp104 variants suppress toxicity of diverse neurodegenerative disease-linked proteins. Dis Model Mech. pii: dmm.016113.
345. Jackrel ME, DeSantis ME, Martinez BA et al (2014) Potentiated Hsp104 variants antagonize diverse proteotoxic misfolding events. Cell 156:170–182
346. Liu CC, Tsai CW, Deak F et al (2014) Deficiency in LRP6-mediated Wnt signaling contributes to synaptic abnormalities and amyloid pathology in Alzheimer's disease. Neuron 84:63–77
347. Hattori N (2014) REST as a new therapeutic target for neurodegenerative disorders. Mov Disord 29:869
348. Matlack KE, Tardiff DF, Narayan P et al (2014) Clioquinol promotes the degradation of metal-dependent amyloid-β (Aβ) oligomers to restore endocytosis and ameliorate Aβ toxicity. Proc Natl Acad Sci U S A 111:4013–4018
349. Cuadrado-Tejedor M, Ricobaraza AL, Torrijo R et al (2013) Phenylbutyrate is a multifaceted drug that exerts neuroprotective effects and reverses the Alzheimer's disease-like phenotype of a commonly used mouse model. Curr Pharm Des 19:5076–5084
350. Scholtzova H, Chianchiano P, Pan J et al (2014) Amyloid ß and tau Alzheimer's disease related pathology is reduced by toll-like receptor 9 stimulation. Acta Neuropathol Commun 2:101
351. Sun X, Wheeler CT, Yolitz J et al (2014) Mitochondrial ATP synthase subunit interacts with TOR signaling to modulate protein homeostasis and lifespan in Drosophila. Cell Rep 8:1781–1792
352. Tsai L-H, Madabhushi R (2014) Alzheimer's disease: a protective factor for the ageing brain. Nature 507:439–440
353. Leuzy A, Zimmer ER, Heurling K et al (2014) Use of amyloid PET across the spectrum of Alzheimer's disease: clinical utility and associated ethical issues. Amyloid 21:143–148
354. Zimmer ER, Leuzy A, Bhat V et al (2014) In vivo tracking of tau pathology using positron emission tomography (PET) molecular imaging in small animals. Transl Neurodegener 3:6
355. Garlet TR, Parisotto EB, de Medeiros GDS et al (2013) Systemic oxidative stress in children and teenagers with Down syndrome. Life Sci 93:558–563
356. Gustafsson MG (2008) Super-resolution light microscopy goes live. Nat Methods 5:385–387
357. Johnson RD, Steel DG, Gafni A (2014) Structural evolution and membrane interactions of Alzheimer's amyloid-beta peptide oligomers: new knowledge from single-molecule fluorescence studies. Protein Sci 23:869–883
358. Dicarlo JE, Norville JE, Mali P et al (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343
359. Platt RJ, Chen S, Zhou Y et al (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell 159:440–455
360. Cerciello F, Choi M, Nicastri A et al (2013) Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. Clin Proteomics 10:16
361. Han SH, Kim JS, Lee Y et al (2014) Both targeted mass spectrometry and flow sorting analysis methods detected the decreased serum apolipoprotein e level in Alzheimer's disease patients. Mol Cell Proteomics 13:407–419
362. Hye A, Riddoch-Contreras J, Baird AL et al (2014) Plasma proteins predict conversion to dementia from prodromal disease. Alzheimers Dement. pii: S1552-5260(14)02454-6
363. Roemmelt AT, Steuer AE, Poetzsch M, Kraemer T (2014) LC QTOF with SWATH acquisition: systematic studies on its use for screenings in clinical and forensic toxicology and comparison with IDA and targeted MRM approaches. Anal Chem. doi:10.1021/ac503144p
364. Mapstone M, Cheema AK, Fiandaca MS et al (2014) Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med 20:415–418
365. Cabezas-Wallscheid N, Klimmeck D, Hansson J et al (2014) Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell 15:507–522
366. Chen JC, Alvarez MJ, Talos F et al (2014) Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell 159:402–414
367. Biesecker LG, Green RC (2014) Diagnostic clinical genome and exome sequencing. N Engl J Med 370:2418–2425
368. O'Bryant SE, Gupta V, Henriksen K et al. STAR-B and BBBIG working groups (2014) Guidelines for the standardization of preanalytic variables for blood-based biomarker studies in Alzheimer's disease research. Alzheimers Dement. pii: S1552-5260(14)02765-4
369. Pollitzer E (2013) Biology: cell sex matters. Nature 500:23–24
370. Jazin E, Cahill L (2010) Sex differences in molecular neuroscience: from fruit flies to humans. Nat Rev Neurosci 11:9–17
371. Deng X, Berletch JB, Nguyen DK, Disteche CM (2014) X chromosome regulation:

diverse patterns in development, tissues and disease. Nat Rev Genet 15:367–378
372. Das J, Vo TV, Wei X et al (2013) Cross-species protein interactome mapping reveals species-specific wiring of stress response pathways. Sci Signal 6:ra38
373. Ebert AD, Liang P, Wu JC (2012) Induced pluripotent stem cells as a disease modeling and drug screening platform. J Cardiovasc Pharmacol 60:408–416
374. Dianat N, Steichen C, Vallier L et al (2013) Human pluripotent stem cells for modelling human liver diseases and cell therapy. Curr Gene Ther 13:120–132
375. Musunuru K (2013) Genome editing of human pluripotent stem cells to generate human cellular disease models. Dis Model Mech 6:896–904
376. Chang MC, Grieder FB (2008) The baffling multitude of disease models for the study of human disease–how can the scientist navigate the huge amount of data and receive guidance? Dis Model Mech 1:99–102
377. Marx V (2013) Model organisms: beyond the inner circle. Nat Methods 10:471–473
378. Murphy MP (2013) Special issue: animal models of disease. Biochim Biophys Acta 1832:1361
379. Yagi T, Ito D, Okada Y, Akamatsu W et al (2011) Modeling familial Alzheimer's disease with induced pluripotent stem cells. Hum Mol Genet 20:4530–4539
380. Israel MA, Yuan SH, Bardy C et al (2012) Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature 482:216–220
381. Choi SH, Tanzi RE (2012) iPSCs to the rescue in Alzheimer's research. Cell Stem Cell 10:235–236
382. Oddo S, Vasilevko V, Caccamo A et al (2006) Reduction of soluble Aβ and tau, but not soluble Aβ alone, ameliorates cognitive decline in transgenic mice with plaques and tangles. J Biol Chem 281:39413–39423
383. Plucińska K, Crouch B, Koss D et al (2014) Knock-in of human BACE1 cleaves murine APP and reiterates Alzheimer-like phenotypes. J Neurosci 34:10710–10728
384. Zis P, Dickinson M, Shende S et al (2012) Oxidative stress and memory decline in adults with Down syndrome: longitudinal study. J Alzheimers Dis 31:277–283
385. Hussain NK, Diering GH, Sole J et al (2014) Sorting Nexin 27 regulates basal and activity-dependent trafficking of AMPARs receptors. Proc Natl Acad Sci U S A 111:11840–11845
386. Wang X, Huang T, Zhao T et al (2014) Sorting Nexin 27 regulates Aβ production through modulating γ-secretase activity. Cell Rep. doi:http://dx.doi.org/10.1016/j.celrep.2014.09.037

Chapter 2

Application of Systems Theory in Longitudinal Studies on the Origin and Progression of Alzheimer's Disease

Simone Lista, Zaven S. Khachaturian, Dan Rujescu, Francesco Garaci, Bruno Dubois, and Harald Hampel

Abstract

This chapter questions the prevailing "implicit" assumption that molecular mechanisms and the biological phenotype of dominantly inherited early-onset alzheimer's disease (EOAD) could serve as a linear model to study the pathogenesis of sporadic late-onset alzheimer's disease (LOAD). Now there is growing evidence to suggest that such reductionism may not be warranted; these suppositions are not adequate to explain the molecular complexities of LOAD. For example, the failure of some recent amyloid-centric clinical trials, which were largely based on the extrapolations from EOAD biological phenotypes to the molecular mechanisms in the pathogenesis of LOAD, might be due to such false assumptions. The distinct difference in the biology of LOAD and EOAD is underscored by the presence of EOAD cases without evidence of familial clustering or Mendelian transmission and, conversely, the discovery and frequent reports of such clustering and transmission patterns in LOAD cases. The primary thesis of this chapter is that a radically different way of thinking is required for comprehensive explanations regarding the distinct complexities in the molecular pathogenesis of inherited and sporadic forms of Alzheimer's disease (AD). We propose using longitudinal analytical methods and the paradigm of systems biology (using transcriptomics, proteomics, metabolomics, and lipidomics) to provide us a more comprehensive insight into the lifelong origin and progression of different molecular mechanisms and neurodegeneration. Such studies should aim to clarify the role of specific pathophysiological and signaling pathways such as neuroinflammation, altered lipid metabolism, apoptosis, oxidative stress, tau hyperphosphorylation, protein misfolding, tangle formation, and amyloidogenic cascade leading to overproduction and reduced clearance of aggregating amyloid-beta (Aβ) species. A more complete understanding of the distinct difference in molecular mechanisms, signaling pathways, as well as comparability of the various forms of AD is of paramount importance. The development of knowledge and technologies for early detection and characterization of the disease across all stages will improve the predictions regarding the course of the disease, prognosis, and response to treatment. No doubt such advances will have a significant impact on the clinical management of both EOAD and LOAD patients. The approach propped here, combining longitudinal studies with the systems biology paradigm, will create a more effective and comprehensive framework for development of prevention therapies in AD.

Key words Early-onset alzheimer's disease, Late-onset alzheimer's disease, Sporadic Alzheimer's disease, Longitudinal studies, Systems biology, Transcriptomics, Proteomics, Metabolomics, Lipidomics, Biological markers

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_2,

1 Introduction

The prevailing conceptual model of Alzheimer's disease (AD) is that of a primary degenerative brain disease, a protein misfolding disease, characterized by the hallmark of (a) senile plaques with primary constituent of amyloid-beta (Aβ) peptide, and (b) neurofibrillary tangles, consisting of hyperphosphorylated tau protein (p-tau) [1]. The biological genotypes as well as the clinical-behavioral phenotypes of Alzheimer's syndrome are highly heterogeneous in both forms: early-onset familial AD (EOAD or fAD) and late-onset or sporadic AD (LOAD or sAD) [2].

The fAD form of the syndrome is extremely rare; only 1–5 % of all cases are fAD type with mutations in one of the three specific genes: amyloid precursor protein (*APP*, located at chromosome region 21q21.2), presenilin 1 (*PSEN1*, located at 14q24.3), and presenilin 2 (*PSEN2*, located at 1q42.13) [3]. These mutations are inherited by autosomal dominant transmission; therefore, the clinical entity is referred to as autosomal dominant AD (ADAD). The term fAD is often used to denote the early-onset AD (EOAD), with an average age of onset at ca. 50 years, although there are rare case of earlier onset where symptoms are manifested as early as the age of 35 or less. Although the current state of knowledge regarding the mutations in these three genes offers substantial information on plausible molecular mechanism underlying the fAD or Mendelian form of AD, this knowledge does not appear to be adequate to explain all cases or other forms of the syndrome. Thus, the search for a comprehensive elucidation of the complete cascade of molecular processes leading to AD syndrome will require the discovery of not only other genes but also the role of epigenetic factors or gene-gene interactions [2]. In this regard, pathogenic mutations in progranulin (*GRN*) gene, which are traditionally known to cause the autosomal-dominant form of frontotemporal dementia (FTD) [4, 5], can be linked to familial early-onset forms of dementia [6–10]. Thus, it is likely that the variants in *GRN* may also play a role as risk factors in AD.

The majority of AD cases (>95 %), which do not show Mendelian inheritance, have a typical onset age of 65 years or older and belong to the late-onset AD (LOAD) or sAD form [3]. The best established and perhaps the most important genetic susceptibility for LOAD is the presence of one or two copies of the *ε4* allele in the apolipoprotein E gene (*APOE*, chromosome 19q13.2) [11]. The employment of high-throughput DNA sequencing methods, which allow concurrent genotyping of large number of subjects, has enabled the execution of unbiased genome-wide association studies (GWAS). These comprehensive examinations of genomes allow wide range of explorations that are not restricted to genetic variants of presumed pathophysiological significance.

Accordingly, the whole genome has been inspected for association with disease risk (for a summary of the gathered evidence, see the AlzGene database at http://www.alzgene.org) [12]. Recent large-scale GWAS have detected no less than ten novel loci related to an increased risk of developing LOAD [13–17]. These genes may interact with crucial events of the AD pathogenesis, including the amyloidogenic cascade, tau hyperphosphorylation, apoptosis, oxidative processes, cell membrane and endocytic pathways, cholesterol/lipid metabolism, apoptosis, and immune-inflammatory mechanisms [18]. Notably, Jones and colleagues have sought to investigate the functional role of genetic variants not quite reaching genome-wide significance in AD and concluded that especially pathways related to immune system response and lipid metabolism seem to be significantly overrepresented [19].

The recognition for the need of large-scale meta-analyses of GWAS to enhance the search for additional genetic risk factors has led to launch the International Genomics of Alzheimer's Project (I-GAP). This collaborative effort links the resources of four consortia conducting research on the genetics of AD, namely the Alzheimer's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium, the European Alzheimer's Disease initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) consortium. Following the meta-analysis of 74,046 subjects, this international collaboration has allowed the identification of 11 novel susceptibility loci for AD [20]. Some of these novel genes highlight the importance of pathways previously assumed to be associated with an increased risk for AD, such as immune response and inflammation, cell migration, lipid transport, and endocytosis. Intriguingly, the presence of novel pathways underpinning AD, e.g., hippocampal synaptic function, cytoskeletal function, axonal transport, and microglial and myeloid cell function, has also been proposed [20]. These novel areas deserve further investigation since they might provide new insights in terms of AD pharmacological research therapy.

Remarkably, novel significant data have provided new incentives in the explication of the genetics mechanisms underpinning the development of LOAD. In this regard, a rare genetic variant—rs75932628—which leads to a substitution of arginine by histidine at the amino acidic residue 47 (R47H) in the *TREM2* gene (encoding for the triggering receptor expressed on myeloid cells 2) is significantly associated with increased susceptibility to LOAD [21, 22]. The amplified risk of LOAD associated with the rs75932628 variant might be caused by a dysregulation of the inflammatory processes in the central nervous system (CNS) [23]. In all likelihood, further AD susceptibility variants are expected to be revealed in imminent GWAS based on larger sample sizes and/or higher resolution genetic maps, as successfully demonstrated in

the large-scale GWAS meta-analysis conducted by Lambert and colleagues [20].

The rare form of dominantly inherited EOAD, with almost complete mutation penetrance and defined age of disease onset, has been traditionally considered as a model to study the very early disease mechanisms that are also supposed to underlie the common sporadic LOAD. However, this commonly accepted dichotomy between EOAD versus LOAD is turning out to be too simplistic a model because of growing evidence, for cases of EOAD without any familial clustering, and, conversely, there are findings reporting familial clustering—transmission patterns in LOAD cases [2]. The emerging knowledge regarding complexity of this multi-genetic disease with a prolonged time course mandated longitudinal analytical methods to determine the interactions among multiple systems: a time-series analysis of the sequences and patterns of changes in various neural networks and signaling pathways. Thus, there is the need for a "systems theory" approach to explain the origin and time course of the functional failure underlying the pathogenesis of all forms of AD. Therefore, we propose that systems biology and longitudinal investigations will provide a more comprehensive characterization of the complex molecular pathogenesis of inherited and sporadic forms of AD. The working hypothesis is that all forms of AD evolve through the convergence of failures in several "systems," networks, signaling pathways, or pathophysiological processes such as neuroinflammation, altered lipid metabolism, apoptosis, oxidative stress, tau hyperphosphorylation, tangle formation, and amyloidogenic cascade including the generation of different Aβ species. This knowledge is of key significance for the understanding of altered disease mechanisms and signaling pathways as well as for the comparability of the purely genetic and the sporadic forms of AD.

2 The Need for Longitudinal Studies

Longitudinal studies in AD involve following EOAD or LOAD patients over time, thereby measuring one or more outcome variables (e.g., neuropsychological performances) and different imaging or biological markers at least at two different points in time, and often more. In general, the primary focus of a longitudinal study is to elucidate if and how much an explanatory variable (or changes in this variable) may cause changes in the outcome variables [24]. The advantages of longitudinal studies over other study designs, such as cross-sectional studies, are well documented. Longitudinal data involve repeated measures of the same subjects over time, while cross-sectional data involve measures at one time point only. Thus, cross-sectional research can only measure the

prevalence of a factor of interest at a certain point in time, while longitudinal research measures prevalence at several points in time and can provide information on causation, prognosis, stability, deviation, and change [25]. While cross-sectional data only allow investigation of differences between individuals, a longitudinal study can examine change within and between individuals as well as differences in variations between them (i.e., interaction effects) [26]. It should be noted, however, that prospective longitudinal studies are very expensive to undertake, particularly in terms of personnel costs. In this context, the future implementation of consortia and collaborative research networks is expected to increase to provide large, representative EOAD and LOAD study samples with high-quality data that can be used to generate longitudinal evidence on the natural history and to inform how molecular and imaging markers can predict the disease course, response to treatment, and clinical decision making [27].

In utilizing neurochemical, imaging, and neuropsychological biomarkers to track changes in longitudinal studies of both inherited and sporadic forms of AD, the physician needs to be aware of the characteristics required of clinically useful markers for this purpose, the sources of their measurement variance, and the principles to minimize this [28]. In longitudinal analyses of AD, biomarker estimates need to be interpreted in the context of the normal variance both within an individual and between individuals [29]. Thus, the main determinants of the normal variance of a biomarker within an individual consist of the within-subject biological variance and analytical variance. To the physician, whose role is to diagnose AD at the asymptomatic stages and monitor disease or response to treatment as early as possible, it is the degree of standardization, e.g., total within-subject and between-subject variance of any given biomarker, that is of greatest relevance in the context of interpreting biomarker levels released by a laboratory and relating this to the clinical state of an individual patient [30]. A longitudinal study investigating the performance of repeated measures of multimodal markers in tracking cognitive changes in EOAD and LOAD relies on all repeated measures being collected and analyzed under optimal conditions to minimize measurement errors and all repeated measures to be consistently recorded at all time points [31].

In conducting longitudinal studies in the setting of EOAD and LOAD, an important issue lies in the population that will be included in the study and how to sample that population; indeed, it is important that any longitudinal project is able to extend the implications of the findings beyond the participants who actually participated in the study. To achieve this goal, studies should be designed not only to recruit sufficient participants to meet the sample size and power requirements of the study but also to

adequately represent the target population. The timing of collecting biomarker data is especially important, as it affects the effectiveness of data-gathering. Timing needs to be considered at all stages of a longitudinal study at recruitment, during follow-up visits, and possibly after a participant has exited the research protocol [28].

The choice of suitable study populations in which to conduct well-designed longitudinal AD prevention trials is a fundamental issue for communities, researchers, sponsors, and stakeholders. Such studies are feasible only among at-risk subjects presenting with evolving mild cognitive impairment (MCI) or prodromal at-risk individuals meeting characteristic biomarker patterns, where high AD incidence rates make prevention trials feasible. In such settings, preliminary feasibility studies with a longitudinal design are considered essential to inform the design of future phase III efficacy and safety trials. Researchers typically frame their assessment of feasibility within an "epidemiological paradigm," focusing on a limited number of key outcome parameters to guide decision making, like biochemical or imaging markers [28].

Until now, researchers have mainly related cross-sectional biomarker data to longitudinal clinical or neuroimaging markers or time before the expected onset of disease to construct pseudo-longitudinal AD studies. However, the composition of pseudo-longitudinal cohorts changes over time as most severe cases may decease or exit the study for various reasons of attrition. This may introduce a bias in the clinical trajectories estimated from cross-sectional data. If, for example, patients who exit the sample are more likely to have poorer clinical outcomes than those who stay, the clinical trajectories based on pseudo-longitudinal cohorts will be biased towards a delayed onset of amyloid deposition and neurodegeneration compared with what actually would be. In the future, the use of true longitudinal follow-up data is expected to reduce the confounding effects that may originate from variation in collection modes and procedures across datasets and/or different participating centers.

Recent evidences derived from longitudinal sAD data collected by Villemagne and colleagues from the Australian Imaging Biomarkers and Lifestyle (AIBL) research group [32] reconcile well with cross-sectional results obtained in EOAD mutation carriers enrolled in the Dominantly Inherited Alzheimer Network (DIAN) study [33], thus suggesting that sporadic and inherited forms of AD may share common pathogenic pathways. However, the results from cross-sectional studies and true longitudinal data source seem to suggest significant differences in terms of other biomarkers. Moreover, in light of the biological complexity of AD, it is reasonable that not only longitudinal studies but also a systems biology approach would be required to unravel the different sAD endophenotypes and to explicate the differences and similarities between EOAD and LOAD.

3 Systems Biology of Alzheimer's Disease

Traditional biological strategies to the analysis of the pathogenesis of neurodegenerative diseases, including AD, have commonly been focused on a few important genes and their associated products, resulting in limited knowledge of the complex etiopathogenesis. Differently, systems biology is an emerging integrative interdisciplinary approach applying advances in multimodal high-throughput method development that permits the investigation of networks of biological pathways where high amounts of structurally/functionally dissimilar molecules are simultaneously detected over time in cells, groups of cells, tissues, organs, or whole organisms [34]. According to Noorbakhsh and colleagues, systems biology is dependent on comprehensive enumeration and quantification of biological processes, followed by data exploration and integration, enabling the formulation of hypotheses that can be verified at a system level (for instance, cell, tissue, or organism) [34].

Systems biology-based strategies have become practicable only in recent times, owing to the development of various high-throughput methods applied to the "omics" sciences including genomics, transcriptomics, proteomics, and metabolomics/lipidomics. Technological platforms employed in the "omics" disciplines, in conjunction with accurate and dedicated statistical/computational tools, allow the characterization of various biomolecules including DNA sequences, transcripts, proteins, metabolites, and lipids [35].

3.1 Transcriptomics

In addition to genetic variants, other sources of variability are associated with transcriptional, translational, and posttranslational modifications. Therefore, after inspecting the collective genotypes of a subject, it is reasonably appropriate to move towards the area of transcriptomics. Transcriptomics designates the examination of the genome-wide gene expression products. The transcriptome represents the whole set of RNA transcripts in a distinct cell type or tissue at a specific developmental stage and/or under certain physiological conditions, including messenger RNA (mRNA), transfer RNA, ribosomal RNA, and other noncoding RNAs. While the genome is stable in a given individual, the transcriptome is subject to variations along the cell cycle and the life cycle of the organism and along all various kinds of tissues; moreover, it can be affected by epigenetic modifications [36]. Transcriptomics frequently utilizes high-throughput methods including hybridization- or sequence-based approaches. More recently, the development of novel high-throughput DNA sequencing technologies has provided a novel system for mapping/quantifying transcriptomes referred to as RNA sequencing (RNA-Seq) that has been predicted to revolutionize the way in which eukaryotic transcriptomes will be scrutinized [37].

Transcriptomic studies in AD compare mRNA expression levels between AD patients and healthy controls. Significant differences in mRNA expression levels indicate that genes are expressed to a diverse extent. This, in turn, may result in dissimilar levels of proteins. An influential study by Liang and colleagues [38] has reported gene expression profiles from postmortem examination of six anatomically and functionally separate brain regions in AD patients. Notably, the authors have disclosed significant regional differential expression in AD brains compared with healthy control brains including expression changes of genes previously implicated in AD pathogenesis and especially concerning plaques and tangle formation [38]. In another study, consistent patterns of alteration in the gene expression profile in the neocortex of AD patients versus healthy subjects have been described. The exploration of the AD transcriptome unveiled synaptic dysfunction, disrupted neurotransmission, and generation of neuroinflammation [39].

Given the high structural and histological heterogeneity of the brain, the selection of the anatomical cerebral area to inspect might significantly affect the final results. Furthermore, human brain tissues for analysis can only be attained postmortem and sampled only once. In light of this, transcriptomic research efforts have been directed towards blood that is easily accessible and can be sampled repeatedly, thus allowing for longitudinal assessment of gene dysregulation at different disease stages. In this regard, Maes and colleagues have described the gene expression profiling of blood mononuclear cells of mild sAD patients versus cognitively healthy individuals [40]. Twenty-eight percent of the upregulated genes and 16 % of the downregulated genes in AD blood mononuclear cells have been demonstrated to have analogous expression patterns in AD brain [41–43], whereas only 4 % of affected genes differ in terms of expression between blood and brain [42]. More specifically, AD blood mononuclear cells show a major decrease in the expression of genes related to cytoskeletal preservation, cellular trafficking and stress response, redox homeostasis, transcription, and DNA repair. Therefore, these data emphasize the systemic nature of gene dysregulation in sAD. Fehlbaum-Beurdeley and colleagues [44] have reported a blood RNA signature that can properly discriminate AD patients from non-demented control subjects with a sensitivity of 100 % and specificity of 96 %. This signature includes genes that are part of pathways related to macrophages and lymphocytes in AD patients: transforming growth factor signaling, oxidative stress, innate immunity/inflammation, and cholesterol homeostasis [44]. The implementation of a transcriptome study of leukocytes from MCI subjects, AD patients, as well as healthy controls by using oligonucleotide microarrays has led to disclosure of eight genes significantly associated with purine metabolism and ATP-binding cassette (ABC) transporters [45]. Interestingly, Booji and colleagues have established and validated a

blood gene expression signature that is able to categorize AD patients and cognitively healthy individuals with high accuracy (87 %) [46]. More recently, Fehlbaum-Beurdeley and colleagues [47] have demonstrated that AclarusDx™, a blood-based transcriptomic test, can be of valuable support in differentiating AD patients from healthy control individuals. This noninvasive test, in conjunction with standard assessments, can provide physicians with unbiased data and indications to facilitate the diagnosis of AD. AclarusDx™ is currently being assessed in many prospective cohorts that will be of benefit in improving its clinical utility [47]. Finally, Han and colleagues [48], after systematically exploring data on AD blood transcriptome, have observed a distinctive perturbation of cellular functional units, including upregulation of environmental responses (immune responses, survival/death signaling, and cellular recycling pathways) and downregulation of central metabolism (energy metabolism and translation/splicing mechanisms). This peculiar perturbation has been found to be characteristic of AD when compared to blood transcriptomes from other neurological diseases. More importantly, a comparable degree of perturbation has been reported in both AD patients and MCI subjects [48].

The easiness with which fresh blood mononuclear cells can be noninvasively and repeatedly collected should (a) allow scrupulous longitudinal evaluation of gene dysregulation in the different stages of MCI and AD, (b) simplify assessment of disease-modifying interventions, and (c) be of support during the differential diagnosis of dementia in elderly people.

In brief, transcriptomic studies have identified a large number of genes and putative pathways. The growing list of mechanisms/pathways substantiates the heterogeneous nature of AD and needs further standardization, replication, and validation.

3.2 Proteomics

Among the developing platforms needed to perform research in neurodegenerative diseases, with AD being investigated most extensively, proteome analysis (proteomics) has gathered high consideration. The word "proteome" has been originally employed to designate the cluster of all proteins expressed by a specific genome [49]. Given that its composition varies from tissue to tissue and from cell to cell, the proteome refers to a set of proteins in a certain time and space, which highlights its dynamic nature [50]. Proteins are characterized by alternative splicing mechanisms, conformational changes, and posttranslational modifications. These features account for the high variability of the proteomic information. In light of this, the number of proteins is much greater than that of their corresponding genes. Therefore, proteins are assumed to have a high potential as dynamic biomarkers in the diagnosis, prediction, and progress monitoring of a given disease [51].

Proteomics has been used on a global scale in neuroscience to scrutinize and decipher the molecular "bar code" of the brain. Such a system-based strategy is called neuroproteomics, the large-scale profiling and functional annotation of brain proteins [52]. The description of the CNS proteome under both normal and pathological conditions is an important initiative sponsored by the Human Proteome Organization (HUPO) (available at http://www.hupo.org/), the largest international consortium supporting proteomic research and study of human tissues. Interestingly, the target of the HUPO Brain Proteome Project (HUPO BPP), an initiative launched by HUPO, is the elucidation of the CNS proteome in both ageing and neurodegenerative diseases [53].

Progresses in analytical instrumentation, especially in mass spectrometry and improvement of ionization techniques, have enhanced proteomic examination of tissues. The interpretation of mass spectra may allow portions of the spectrum to be labeled with a protein/peptide identity. However, when a protein/peptide has been recognized, what confers it the status of a candidate marker is its consistent variation in some features such as abundance between two states (for instance, presence or absence of a pathophysiology) [54]. Different approaches have been employed to inspect proteome alterations in AD during the last 15 years. Proteomics of AD has been investigated at different stages of the disease using a plethora of high-throughput systems. These have been exploited in different types of clinical biofluids, especially cerebrospinal fluid (CSF) [55–70] and blood (i.e., plasma/serum) [71–80]. Moreover, critical overviews of the literature in the area of AD proteomics have been produced in the last decade [81–87].

Given the dynamic progresses in AD proteomic research and the continuous introduction of new technology platforms, proteins encompass the majority of feasible candidate biomarkers for AD diagnosis and are of value for measuring statistical differences between AD patients and healthy controls. Furthermore, several of these protein molecules show potential to increase the diagnostic accuracy of the currently recognized CSF markers of AD, i.e., $A\beta_{1-42}$, total tau (t-tau), and tau phosphorylated at threonine 181 (p-tau_{181}) [88].

3.3 Metabolomics and Lipidomics

The most recent of the "omics" sciences—metabolomics—offers powerful approaches that allow the examination of perturbations in metabolic pathways/networks [80, 89]. Metabolomics accumulates quantitative data on a large amount of metabolites in order to delineate a complete depiction of metabolism and to characterize metabolic fluctuations associated with pathological conditions [90]. Metabolomics implicates the inspection of small molecules—metabolites—detected in cells, tissues, organs, or biological fluids. To this end, several techniques have been developed to separate and measure the components of the metabolome. The selection of

the appropriate platform will enable the analysis of these molecules with regard to their individual features [89].

The technologies employed by most studies are (a) mass spectrometry, utilized in combination with gas/liquid chromatography or capillary electrophoresis in order to attain the initial separation of metabolites, and (b) nuclear magnetic resonance (NMR) spectroscopy. It should be noted that each methodological approach shows advantages as well as drawbacks since the classes of metabolites detected by each system are different. At present, no specific analytical methodology is able to capture the whole metabolomic information from a single sample. Thus, various approaches need to be combined, especially in the area of biomarker discovery [91]. Notably, owing to progress in instrumentation, it has recently become possible to simultaneously measure thousands of metabolites using extremely low sample volumes [92]. In addition, advances in technology coupled with recent bioinformatic tools and software have allowed the comprehensive inspection of cellular metabolites without bias. However, several molecules whose importance has been highlighted have not been integrated in metabolite repositories; therefore, the depiction of cellular metabolism is still incomplete [93].

Several pathologies have been shown to result in the disruption of metabolic pathways. Therefore, they can produce long-term metabolic alterations that can be reported in terms of metabolic signatures. Metabolomic profiling can be executed quite easily in peripheral tissues as well as in biofluids including CSF or plasma/serum, thus making this approach suitable for clinical applications [94]. Metabolomic signatures have been documented for some pathological conditions, including AD [95]. These signatures are represented by numerous dysregulated metabolites. Their levels fluctuate in the disease state or after drug exposure. The analysis of these signatures might offer significant information concerning disease pathophysiology [96]. In addition, several projects aiming at disclosing serum-derived metabolic markers in AD are currently available, such as HUSERMET (available at http://www.husermet.org/) [97, 98] and PredictAD (available at http://www.predictad.eu/).

Metabolomic profiling has been utilized to evaluate cross-sectional modifications in brain samples [99], CSF [100–102], and plasma/serum specimens [103–105] from patients with a different extent of AD clinical severity. Nevertheless, attempts of matching the obtained data have been limited since metabolomic platforms and biofluids employed were dissimilar and the studies diverged in the range of the detected metabolites. Very recently, nontargeted mass spectrometry-based metabolomic profiling has been employed to search for global modifications in metabolites and several recognized metabolic pathways in both CSF and plasma specimens from the same subjects in the context of AD progression [94].

Notably, the altered pathways revealed by CSF and plasma analysis of MCI individuals and AD patients have been primarily associated with mitochondrial activity and energy metabolism as well as biosynthesis/trafficking/metabolism of lipids, amino acids, neurotransmitters, and hormones [94].

The evidence that *APOE*—encoding for apolipoprotein E, a lipid transport/chaperone protein—is the most important gene for sporadic LOAD points to the importance of lipid dynamics as an area that needs to be explored in AD research. Lipidomics is a specific branch of metabolomics focused on detecting and quantifying a wide range of polar and nonpolar lipid metabolites in cells and biofluids to attain a complete representation of human lipid biochemical pathways [106]. The comprehensive analysis of lipid metabolic pathways in AD might provide novel concepts on the modulation of lipid homeostasis, which may, therefore, explicate the pathogenesis of the disease [107]. Intriguingly, lipid molecular profiling, combined with biophysical modeling of membrane systems to examine lipid membranes [108] or lipoproteins [109], provides the opportunity to associate the molecular pathway alterations with cell- and tissue-level physiology and structure. This may not only suggest new insights into disease pathogenesis, but also offer new opportunities from a diagnostic and therapeutic viewpoint.

4 Molecular Networks in Alzheimer's Disease

The main goal of high-content analyses centered on AD pathogenesis is the system-wide investigation of qualitative/quantitative alterations in transcripts, proteins, metabolites, and lipids, followed by the formulation of pathways linking biomolecules and pathophysiological, disease-related mechanisms. These descriptive analyses represent several systems biology-based investigations focused on AD [34, 110]. Notably, according to Kitano [110], a second category of systems biology analyses, which exhibits higher complexity, has been more recently developed. These are molecular networks, also called "molecular modules," that are created within a cell. Understanding the arrangement and connectivity of molecular networks allows the prediction of their dynamical behavior, including the depiction of their physiological or pathological status [34, 110].

In this regard, an influential study by Miller and colleagues employed weighted gene co-expression network analysis (WGCNA) [111–113] on microarray data from the CA1 region of the hippocampus to identify co-expression modules associated with AD [114]. This strategy arranges results from gene expression studies into a framework to inspect AD pathophysiology from a systems viewpoint. WGCNA explicates the higher degree links among

genes via their co-expression interactions, defining groups (i.e., modules) of biologically connected genes and enabling an unbiased depiction of transcriptome organization, which is relevant to the pathology. Moreover, within these clusters of highly co-expressed genes, WGCNA is able to detect the most strictly related genes, often termed "hub genes" [115]. This approach has allowed to identify a number of AD-associated co-expression modules participating in synaptic transmission, mitochondrial activities, metabolic functions, immune response mechanisms, extracellular transport, and myelination [114]. The study of the module comprising a substantial number of mitochondrial genes (i.e., the "mitochondrial module") has led to discovery of some hub genes playing a part in ion transport: voltage-dependent anion channel 1 and 3 (*VDAC1*, *VDAC3*) and ATP synthase, H^+ transporting, mitochondrial F_0 complex, subunit B1 (*ATP5F1*) [114]. Other hub genes belonging to "synaptic modules," such as synaptojanin 1 (*SYNJ1*), syntaxin-binding protein 1 (*STXBP1*), and synaptosomal associated protein, 91 kDa (*SNAP91*), have clearly been implicated in fusion of synaptic vesicles and endocytosis. Therefore, they are expected to participate in synaptic transmission [114]. Notably, the implementation of a comparative network analysis of transcriptome organization between AD and physiological aging has disclosed a significant overlap between mitochondrial and synaptic modules. In light of this, a certain degree of similarity between the two conditions has been shown. However, local network examination of well-established AD genes revealed altered connectivity for *PSEN1* between AD and aging. Thus, the role of *PSEN1* in CA1 region in AD may diverge from that in normal aging. The modified activity of *PSEN1* in the AD network is in line with its acknowledged function in the pathology [114].

Protein-protein interaction (PPI) network models are of help in detecting important proteins and biochemical pathways in a specific pathological condition and offer a context for the examination of complex disorders such as AD. Network models have also been employed in integrating data from other sources, such as gene expression data [116, 117]. This combined strategy helps discover crucial cellular pathways or complexes where up- or down-regulated gene products are grouped, thus finding genes potentially related to the disease.

In a study by Hallock and Thomas, a "core" PPI network for AD has been created by reviewing the primary literature as well as web resources associated with AD [118]. Then, the employment of data from the Human Protein Reference Database (HPRD) [119] has led to assembly of an "expanded" network, enriched with supplementary proteins interacting with those of the core PPI network. After inspecting the structure and protein content of both networks to detect central proteins, protein interactions, and molecular pathways implicated in AD, existing gene expression

studies have been mapped to the core network. This combined model has allowed the identification of two cellular pathways exhibiting clusters of differentially modulated genes: (a) the mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) pathway, which is significantly involved in synaptic plasticity mechanisms and is altered in AD, and (b) clathrin-mediated receptor endocytosis, involved in the internalization of APP that can result in grown intracellular levels of Aβ. The first and the second pathways have been shown to be downregulated and upregulated in AD, respectively [118]. Interestingly, some genes recognized by recent GWAS as being associated with AD play also a role in receptor-mediated endocytosis and/or clathrin interactions [17, 20]. These include CD2-associated protein (*CD2AP*), bridging integrator 1 (*BIN1*), phosphatidylinositol-binding clathrin assembly protein (*PICALM*) [17], and sortilin-related receptor, L(DLR class) 1 (*SORL1*) [20]. All these findings highlight the prominent function of clathrin-mediated endocytosis in AD.

In summary, the analyses of complex molecular interactions, such as transcriptional modules, gene-interaction networks, PPI networks, and signaling networks, are of significant relevance because they help comprehensively unveil the previously unknown complex molecular network properties of AD as well as detect key genes, proteins, and cellular pathways involved in AD mechanisms. This, in turn, supports the establishment and selection of gene/protein targets for treating the disease.

5 Conclusions

The high degree of heterogeneity in the biological as well as behavioral-clinical genotypes of AD syndrome is reflected in the extensive variations in the neuropathological lesions, age of onset, and pattern and types of behavioral-clinical manifestations. These well-validated observations regarding the "Alzheimer's syndrome" speak for the complexity of the multigenic nature of this disorder.

Thus, this extreme complexity of AD pathogenesis mandates the necessity for integrating knowledge-information from several distinct but parallel sources (systems or networks) using both clinical and laboratory observations such as high-throughput molecular profiling strategies (transcriptomics, proteomics, metabolomics, and lipidomics), multiple brain imaging, neurophysiological measures, and psychometric cognitive assessments. From a basic research standpoint, the development of the "omics" sciences has provided the strategies for the identification of novel molecular biomarkers from biofluids, cells, and tissues. High-throughput methods have the ability to collect large amounts of data with reference to a specific phenotype or disease status in an unbiased way.

The implementation of a systems biology strategy to investigate AD will require the amalgamation of heterogeneous data. The combination of these data includes the development of tools not only for storing and mining the data, but also for modeling of the data in the context of disease pathophysiology [120]. Sophisticated methods are required to scrutinize molecules related to AD and their interactions in the spatial and temporal setting. Since data attained at different levels may carry complementary information on the pathophysiology of AD, their integration is likely to enhance the diagnosis and the interpretation of the pathophysiology [120] and the elucidation of the commonalities and dissimilarities between inherited and sporadic forms of AD. From a clinical standpoint, the implementation of consortia and collaborative research networks is expected to provide large, representative EOAD and LOAD study samples that can be used to generate longitudinal characterization of the natural history and to inform how novel molecular and imaging biomarkers can predict the disease course, response to treatment, and clinical decision making.

Acknowledgements

HH and BD would like to thank for the support of the Fondation Pour La Recherche Sur Alzheimer (FRA), Paris, France.

References

1. Blennow K, De Leon MJ, Zetterberg H (2006) Alzheimer's disease. Lancet 368: 387–403
2. Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. Neuron 68:270–281
3. Bertram L, Tanzi RE (2012) The genetics of Alzheimer's disease. Prog Mol Biol Transl Sci 107:79–100
4. Van Deerlin VM, Wood EM, Moore P et al (2007) Clinical, genetic, and pathologic characteristics of patients with frontotemporal dementia and progranulin mutations. Arch Neurol 64:1148–1153
5. Huey ED, Grafman J, Wassermann EM et al (2006) Characteristics of frontotemporal dementia patients with a Progranulin mutation. Ann Neurol 60:374–380
6. Perry DC, Lehmann M, Yokoyama JS et al (2013) Progranulin mutations as risk factors for Alzheimer disease. JAMA Neurol 70:774–778
7. Jin SC, Pastor P, Cooper B et al (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res Ther 4:34
8. Kelley BJ, Haidar W, Boeve BF et al (2010) Alzheimer disease-like phenotype associated with the c.154delA mutation in progranulin. Arch Neurol 67:171–177
9. Finch N, Baker M, Crook R et al (2009) Plasma progranulin levels predict progranulin mutation status in frontotemporal dementia patients and asymptomatic family members. Brain 132:583–591
10. Brouwers N, Nuytemans K, van der Zee J et al (2007) Alzheimer and Parkinson diagnoses in progranulin null mutation carriers in an extended founder family. Arch Neurol 64: 1436–1446
11. Strittmatter WJ, Saunders AM, Schmechel D et al (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90:1977–1981
12. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23

13. Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41:1088–1093
14. Lambert JC, Heath S, Even G et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094–1099
15. Seshadri S, Fitzpatrick AL, Ikram MA et al (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303:1832–1840
16. Hollingworth P, Harold D, Sims R et al (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43:429–435
17. Naj AC, Jun G, Beecham GW et al (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 43:436–441
18. Zetzsche T, Rujescu D, Hardy J et al (2010) Advances and perspectives from genetic research: development of biological markers in Alzheimer's disease. Expert Rev Mol Diagn 10:667–690
19. Jones L, Holmans PA, Hamshere ML et al (2010) Genetic evidence implicates the immune system and cholesterol metabolism in the etiology of Alzheimer's disease. PLoS One 5:e13950
20. Lambert JC, Ibrahim-Verbaas CA, Harold D et al (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45: 1452–1458
21. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116
22. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127
23. Neumann H, Daly MJ (2013) Variant TREM2 as risk factor for Alzheimer's disease. N Engl J Med 368:182–184
24. Chinn S (1989) Longitudinal studies: design and analysis. Rev Epidemiol Sante Publique 37:431–441
25. Gibbons RD (2008) Design and analysis of longitudinal studies. Psychiatr Ann 38: 758–761
26. Liu C, Cripe TP, Kim MO (2010) Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. Mol Ther 18:1724–1730
27. Hampel H, Lista S (2012) Alzheimer disease: from inherited to sporadic AD-crossing the biomarker bridge. Nat Rev Neurol 8:598–600
28. Hampel H, Lista S, Khachaturian ZS (2012) Development of biomarkers to chart all Alzheimer's disease stages: the royal road to cutting the therapeutic Gordian Knot. Alzheimers Dement 8:312–336
29. Hampel H, Lista S (2013) Use of biomarkers and imaging to assess pathophysiology, mechanisms of action and target engagement. J Nutr Health Aging 17:54–63
30. Rosén C, Hansson O, Blennow K et al (2013) Fluid biomarkers in Alzheimer's disease - current concepts. Mol Neurodegener 8:20
31. Carrillo MC, Blennow K, Soares H et al (2013) Global standardization measurement of cerebral spinal fluid for Alzheimer's disease: an update from the Alzheimer's Association Global Biomarkers Consortium. Alzheimers Dement 9:137–140
32. Villemagne VL, Burnham S, Bourgeat P et al (2013) Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 12:357–367
33. Bateman RJ, Xiong C, Benzinger TL et al (2012) Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 367:795–804
34. Noorbakhsh F, Overall CM, Power C (2009) Deciphering complex mechanisms in neurodegenerative diseases: the advent of systems biology. Trends Neurosci 32:88–100
35. Tyers M, Mann M (2003) From genomics to proteomics. Nature 422:193–197
36. Dong Z, Chen Y (2013) Transcriptomics: advances and approaches. Sci China Life Sci 56:960–967
37. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63
38. Liang WS, Dunckley T, Beach TG et al (2008) Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol Genomics 33:240–256
39. Tan MG, Chua WT, Esiri MM et al (2010) Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. J Neurosci Res 88:1157–1169
40. Maes OC, Xu S, Yu B et al (2007) Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. Neurobiol Aging 28:1795–1809
41. Colangelo V, Schurr J, Ball MJ et al (2002) Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and upregulation of apoptotic and pro-inflammatory signaling. J Neurosci Res 70: 462–473
42. Blalock EM, Geddes JW, Chen KC et al (2004) Incipient Alzheimer's disease: microarray

correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci U S A 101:2173–2178
43. Yao PJ, Zhu M, Pyun EI et al (2003) Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer's disease. Neurobiol Dis 12:97–109
44. Fehlbaum-Beurdeley P, Jarrige-Le Prado AC, Pallares D et al (2010) Toward an Alzheimer's disease diagnosis via high-resolution blood gene expression. Alzheimers Dement 6:25–38
45. Chen KD, Chang PT, Ping YH et al (2011) Gene expression profiling of peripheral blood leukocytes identifies and validates ABCB1 as a novel biomarker for Alzheimer's disease. Neurobiol Dis 43:698–705
46. Booij BB, Lindahl T, Wetterberg P et al (2011) A gene expression pattern in blood for the early detection of Alzheimer's disease. J Alzheimers Dis 23:109–119
47. Fehlbaum-Beurdeley P, Sol O, Désiré L et al (2012) Validation of AclarusDx™, a blood-based transcriptomic signature for the diagnosis of Alzheimer's disease. J Alzheimers Dis 32:169–181
48. Han G, Wang J, Zeng F et al (2013) Characteristic transformation of blood transcriptome in Alzheimer's disease. J Alzheimers Dis 35:373–386
49. Wasinger VC, Cordwell SJ, Cerpa-Poljak A et al (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. Electrophoresis 16:1090–1094
50. Tambor V, Fucíková A, Lenco J et al (2010) Application of proteomics in biomarker discovery: a primer for the clinician. Physiol Res 59:471–497
51. Thambisetty M, Lovestone S (2010) Blood-based biomarkers of Alzheimer's disease: challenging but feasible. Biomark Med 4:65–79
52. Becker M, Schindler J, Nothwang HG (2006) Neuroproteomics – the tasks lying ahead. Electrophoresis 27:2819–2829
53. Hamacher M, Meyer HE (2005) HUPO Brain Proteome Project: aims and needs in proteomics. Exp Rev Proteomics 2:1–3
54. Rifai N, Gillette MA, Carr SA et al (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24:971–983
55. Portelius E, Gustavsson MK, Zetterberg H et al (2012) Evaluation of the performance of novel Aβ isoforms as theragnostic markers in Alzheimer's disease: from the cell to the patient. Neurodegener Dis 10:138–140
56. Portelius E, Price E, Brinkmalm G et al (2011) A novel pathway for amyloid precursor protein processing. Neurobiol Aging 32: 1090–1098
57. Perrin RJ, Craig-Schapiro R, Malone JP et al (2011) Identification and validation of novel cerebrospinal fluid biomarkers for staging early Alzheimer's disease. PLoS One 6:e16032
58. Craig-Schapiro R, Kuhn M, Xiong C et al (2011) Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimer's disease diagnosis and prognosis. PLoS One 6:e18850
59. Craig-Schapiro R, Perrin RJ, Roe CM et al (2010) YKL-40: a novel prognostic fluid biomarker for preclinical Alzheimer's disease. Biol Psychiatry 68:903–912
60. Portelius E, Dean RA, Gustavsson MK et al (2010) A novel Abeta isoform pattern in CSF reflects gamma-secretase inhibition in Alzheimer disease. Alzheimers Res Ther 2:7
61. Albertini V, Bruno A, Paterlini A et al (2010) Optimization protocol for amyloid-β peptides detection in human cerebrospinal fluid using SELDI TOF MS. Proteomics Clin Appl 4:352–357
62. Portelius E, Brinkmalm G, Tran AJ et al (2009) Identification of novel APP/Abeta isoforms in human cerebrospinal fluid. Neurodegener Dis 6:87–94
63. Simonsen AH, McGuire J, Podust VN et al (2008) Identification of a novel panel of cerebrospinal fluid biomarkers for Alzheimer's disease. Neurobiol Aging 29:961–968
64. Simonsen AH, McGuire J, Hansson O et al (2007) Novel panel of cerebrospinal fluid biomarkers for the prediction of progression to Alzheimer dementia in patients with mild cognitive impairment. Arch Neurol 64:366–370
65. Simonsen AH, McGuire J, Podust VN et al (2007) A novel panel of cerebrospinal fluid biomarkers for the differential diagnosis of Alzheimer's disease versus normal aging and frontotemporal dementia. Dement Geriatr Cogn Disord 24:434–440
66. Finehout EJ, Franck Z, Choe LH et al (2007) Cerebrospinal fluid proteomic biomarkers for Alzheimer's disease. Ann Neurol 61:120–129
67. Portelius E, Tran AJ, Andreasson U et al (2007) Characterization of amyloid beta peptides in cerebrospinal fluid by an automated immunoprecipitation procedure followed by mass spectrometry. J Proteome Res 6:4433–4439
68. Portelius E, Zetterberg H, Andreasson U et al (2006) An Alzheimer's disease-specific beta amyloid fragment signature in cerebrospinal fluid. Neurosci Lett 409:215–219
69. Portelius E, Westman-Brinkmalm A, Zetterberg H et al (2006) Determination of beta-amyloid peptide signatures in cerebrospinal

fluid using immunoprecipitation–mass spectrometry. J Proteome Res 5:1010–1016
70. Carrette O, Demalte I, Scherl A et al (2003) A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. Proteomics 3:1486–1494
71. Guo LH, Alexopoulos P, Wagenpfeil S (2013) Plasma proteomics for the identification of Alzheimer disease. Alzheimer Dis Assoc Disord. doi:10.1097/WAD.0b013e31827b60d2
72. Doecke JD, Laws SM, Faux NG et al (2012) Blood-based protein biomarkers for diagnosis of Alzheimer disease. Arch Neurol 69: 1318–1325
73. Hu WT, Holtzman DM, Fagan AM et al (2012) Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. Neurology 79:897–905
74. Soares HD, Potter WZ, Pickering E et al (2012) Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease. Arch Neurol 69:1310–1317
75. Johnstone D, Milward EA, Berretta R et al (2012) Multivariate protein signatures of preclinical Alzheimer's disease in the Alzheimer's disease neuroimaging initiative (ADNI) plasma proteome dataset. PLoS One 7:e34341
76. O'Bryant SE, Xiao G, Barber R et al (2011) A blood-based algorithm for the detection of Alzheimer's disease. Dement Geriatr Cogn Disord 32:55–62
77. O'Bryant SE, Xiao G, Barber R et al (2011) A blood-based screening tool for Alzheimer's disease that spans serum and plasma: findings from TARC and ADNI. PLoS One 6:e28092
78. O'Bryant SE, Xiao G, Barber R et al (2010) A serum protein-based algorithm for the detection of Alzheimer disease. Arch Neurol 67: 1077–1081
79. Ray S, Britschgi M, Herbert C et al (2007) Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. Nat Med 13:1359–1362
80. Hye A, Lynham S, Thambisetty M et al (2006) Proteome-based plasma biomarkers for Alzheimer's disease. Brain 129: 3042–3050
81. Henriksen K, O'Bryant SE, Hampel H et al (2013) The future of blood-based biomarkers for Alzheimer's disease. Alzheimers Dement. doi:10.1016/j.jalz.2013.01.013
82. Gupta VB, Sundaram R, Martins RN (2013) Multiplex biomarkers in blood. Alzheimers Res Ther 5:31
83. Lista S, Faltraco F, Prvulovic D et al (2013) Blood and plasma-based proteomic biomarker research in Alzheimer's disease. Prog Neurobiol 101–102:1–17
84. Blennow K, Hampel H, Weiner M et al (2010) Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol 6: 131–144
85. Portelius E, Zetterberg H, Gobom J et al (2008) Targeted proteomics in Alzheimer's disease: focus on amyloid-beta. Expert Rev Proteomics 5:225–237
86. Davidsson P, Sjogren M (2005) The use of proteomics in biomarker discovery in neurodegenerative diseases. Dis Markers 21:81–92
87. Irizarry MC (2004) Biomarkers of Alzheimer disease in plasma. NeuroRx 1:226–234
88. Fagan AM, Perrin RJ (2012) Upcoming candidate cerebrospinal fluid biomarkers of Alzheimer's disease. Biomark Med 6: 455–476
89. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM (2008) Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol 48:653–683
90. Kristal BS, Kaddurah-Daouk R et al (2007) Metabolomics: concept and potential neuroscience application. In: Lajtha A, Gibson G, Dienel G (eds) Handbook of neurochemistry and molecular neurobiology. Brain energetics. Integration of molecular and cellular processes, 3rd edn. Springer, New York, pp 889–912
91. Sun J, Beger RD, Schnackenberg LK (2013) Metabolomics as a tool for personalizing medicine: 2012 update. Personal Med 10: 149–161
92. Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787
93. Baker M (2011) Metabolomics: from small molecules to big ideas. Nat Methods 8: 117–121
94. Trushina E, Dutta T, Persson XM et al (2013) Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's disease using metabolomics. PLoS One 8:e63644
95. Han X, MHoltzman D, McKeel DW Jr (2002) Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: potential role in disease pathogenesis. J Neurochem 82:809–818
96. Kaddurah-Daouk R, Krishnan KR (2009) Metabolomics: a global biochemical approach to the study of central nervous system diseases. Neuropsychopharmacology 34:173–186
97. Dunn WB, Broadhurst D, Begley P et al (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc 6: 1060–1083

98. Cottingham K (2008) HUSERMET researchers look to the metabolome for answers. J Proteome Res 7:4213
99. Han X (2010) Multi-dimensional mass spectrometry-based shotgun lipidomics and the altered lipids at the mild cognitive impairment stage of Alzheimer's disease. Biochim Biophys Acta 1801:774–783
100. Kaddurah-Daouk R, Zhu H, Sharma S et al (2013) Alterations in metabolic pathways and networks in Alzheimer's disease. Transl Psychiatry 3:e244
101. Czech C, Berndt P, Busch K et al (2012) Metabolite profiling of Alzheimer's disease cerebrospinal fluid. PLoS One 7:e31501
102. Kaddurah-Daouk R, Rozen S, Matson W et al (2011) Metabolomic changes in autopsy-confirmed Alzheimer's disease. Alzheimers Dement 7:309–317
103. Sato Y, Suzuki I, Nakamura T et al (2012) Identification of a new plasma biomarker of Alzheimer's disease using metabolomics technology. J Lipid Res 53:567–576
104. Orešič M, Hyötyläinen T, Herukka SK et al (2011) Metabolome in progression to Alzheimer's disease. Transl Psychiatry 1:e57
105. Han X, Rozen S, Boyle SH et al (2011) Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. PLoS One 6:e21643
106. Wenk MR (2010) Lipidomics: new tools and applications. Cell 143:888–895
107. Astarita G, Piomelli D (2011) Towards a whole-body systems [multi-organ] lipidomics in Alzheimer's disease. Prostaglandins Leukot Essent Fatty Acids 85:197–203
108. Niemelä PS, Ollila S, Hyvönen MT et al (2007) Assessing the nature of lipid raft membranes. PLoS Comput Biol 3:e34
109. Yetukuri L, Söderlund S, Koivuniemi A et al (2010) Composition and lipid spatial distribution of HDL particles in subjects with low and high HDL-cholesterol. J Lipid Res 51:2341–2351
110. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
111. Horvath S, Dong J (2008) Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4:e1000117
112. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559
113. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1:54
114. Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J Neurosci 28:1410–1420
115. Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 107:12698–12703
116. Kann MG (2007) Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform 8:333–346
117. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–115
118. Hallock P, Thomas MA (2012) Integrating the Alzheimer's disease proteome and transcriptome: a comprehensive network model of a complex disease. OMICS 16:37–49
119. Prasad TSK, Goel R, Kandasamy K et al (2009) Human Protein Reference Database-2009 update. Nucleic Acids Res 37:D767–D772
120. Orešič M, Lötjönen J, Soininen H (2010) Systems medicine and the integration of bioinformatic tools for the diagnosis of Alzheimer's disease. Genome Med 2:83

Part II

Alzheimer's Disease: Main Underlying Pathways and Networks

Chapter 3

The APP Proteolytic System and Its Interactions with Dynamic Networks in Alzheimer's Disease

Sally Hunter, Steven Martin, and Carol Brayne

Abstract

Diseases of aging are often complex and multifactorial, involving many genetic and life course modifiers. Systems biology is becoming an essential tool to investigate disease initiation and disease progression. Alzheimer's disease (AD) can be used as a case study to investigate the application of systems biology to complex disease. Here we describe approaches to capturing biological data, representing data in terms of networks and interpreting their meaning in relation to the human population. We highlight issues that remain to be addressed both in terms of modeling disease progression and in relating findings to the current understanding of human disease.

Key words Alzheimer's disease, Amyloid precursor protein, Amyloid-beta-protein, Presenilin, Network modeling, Systems biology

1 Introduction

Diseases of aging, such as cancer and neurodegeneration, are complex and multifactorial, involving many genetic and life course modifiers. As more evidence becomes available, many links between different diseases of aging are becoming apparent [1], such as the roles of cell cycle proteins in cancer and neurodegeneration [2] or the contributions of Alzheimer's disease (AD) related and cardiovascular related genes in both normal aging and neurodegeneration [3]. Systems biology, a field that aims to integrate data from diverse biological areas, is becoming an essential tool to investigate processes relating to initiation and progression in complex disease. AD is the most common form of dementia associated with aging and is increasingly being accepted as a complex multifactorial neurodegenerative syndrome. AD can be used as a case study to investigate the application of systems biology to complex molecular disease pathways and relate these to brain behavior and ultimately treatment strategies.

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_3,

2 Overview of Alzheimer's Disease (AD)

AD is characterized clinically by memory loss, cognitive impairments and dementia [4, 5]. These symptoms lead to impairments in activities of daily living with the result that individuals with AD require an increasing degree of support and care as the disease progresses. Neuropathologically, the hallmarks of AD include intracellular neurofibrillary tangles (NFT) composed of paired helical filaments of the microtubule associated protein tau, extracellular senile plaques containing aggregated amyloid-beta-protein (Aβ) and neuritic plaques and dystrophic neurites that are tau reactive and are also often associated with aggregated Aβ [6, 7].

The importance of the amyloid precursor protein (APP) proteolytic system to dementia initiation and progression in AD is highlighted by both neuropathological and genetic evidence. Various mutations within APP and the γ-secretase associated Presenilin (PS) genes, PS1 and PS2, are associated with early onset familial Alzheimer's disease (FAD) [8]. The genetic data is further linked to disease progression by the deposition of the Aβ, a proteolytic fragment of APP, in neuritic and senile plaques. Additionally, the deposition of Aβ in the brain vasculature as congophilic amyloid angiopathy (CAA) is common in AD and may have independent effects on cognitive function [9, 10]. For late onset AD, accounting for >95 % of cases, the genetic contributions to disease are estimated to be between 48 and 79 % [11, 12] and include contributions from genes such as ApoE [13], CLU and PICALM [14] and CR1 [15] amongst others (reviewed in [16, 17]). Lifestyle modifiers that may contribute to dementia risk include education [18], exercise [19] and diet [20].

The relationship between neuropathology and cognitive status is not straight forward [21]. While considered as neuropathological hallmarks of AD, clinicopathological population studies show that the relationships between various neuropathologies, age and dementia status are complex [22] and that very few "pure" AD cases exist [23]. Population studies of the aging brain commonly find the neuropathological hallmarks of AD in cognitively normal individuals, albeit generally at lower severities, and demented individuals may show little neuropathology [21, 22, 24]. This raises questions around how these neuropathologies, and the neurochemistry associated with them, contribute to disease initiation and progression and how AD is defined both clinically and neuropathologically. If the aim is to devise treatment strategies, where some medication may alleviate or prevent the clinical manifestation of dementia, then the relationships between the human genome, (the complete set of genetic material in a cell), the transcriptome, (the entire collection of gene transcripts both destined to be expressed as proteins and as regulatory elements), the proteome,

(the complete set of expressed proteins in a specific cell type), the interactome, (the complete set of molecular interactions in a cell), the functional brain connectome, (the complete set of neural and synaptic connections in the human), and the whole body within its ever-changing environment must be elucidated. Computational models can be a tool to investigate these relationships and how they change due to disease.

3 Basic Background for Biomolecular Networks

Molecular pathways are dynamic functional systems involving multiple players often with complex regulatory systems involving both direct and indirect feedback loops. Flow of biological information through these pathways can be represented as computational networks based on molecular communication theories [25]. Within a cell as a whole, the probability that an interaction or biological reaction will occur between specific molecules and not others depends on many factors including, compartmentation, relative affinity, concentration, half-life, protein modifications, the presence of co-factors, and the formation of biologically active protein complexes.

3.1 Compartmentation

A cell is divided into compartments and forms organized structures that allow cellular processes to occur in a controlled way. Organelles, such as the nucleus, endoplasmic reticulum and mitochondria, isolate specific cellular processes within semi-permeable membranes that concentrate components of a particular cellular process and increase the chance that they will combine. Compartmentation also isolates reactions that would otherwise be deleterious for the whole cell, such as lysosomal reactions involved in the breakdown of proteins tagged for destruction. Within organelles, specific compartments can be defined by further interactions between factors, such as relatively rigid cholesterol-rich lipid raft areas within a more fluid phospholipid membrane. In order to maintain cellular compartments, the cell must express all the various components in the correct place and at the appropriate time and this involves the complex process of cellular trafficking.

3.2 Relative Affinity

The relative affinity of one protein for another contributes to the probability that they will react and this affinity depends on shape and charge distribution which ultimately depend on the amino acid sequence and protein folding. Protein shape and charge distribution are altered by the protein modifications described below and by many other factors including pH, metal ion binding and interactions with other cellular molecules.

3.3 Concentration

The concentration of the active form of a protein depends on many factors including gene expression, protein synthesis, protein modification, trafficking and storage mechanisms and protein degradation amongst others. Concentration is usually tightly regulated and over- or under- expression of active proteins can be disruptive to normal cellular processes.

3.4 Half Life

The rate at which a protein is synthesized and degraded is its turnover and this is characterized by its half-life, i.e. the time it takes for half the amount of a particular protein to be degraded. The length of time a protein is active and available can contribute to the likelihood that it will be involved in a cellular reaction. The concentration of a protein with a short half-life is more easily manipulated by the cell.

3.5 Protein Modifications

After translation, proteins are often processed and/or modified before achieving an active form and more than 200 different types of modification are known [26]. Modifications can be permanent or transient. Permanent modifications include proteolytic processing, where an immature protein, such as immature PS, requires cleavage to attain its active form [27, 28]. Transient and reversible enzymatic modifications are fundamental to the regulation cellular processes and include (1) glycosylation, the addition of sugar groups, (2) phosphorylation and dephosphorylation, the addition and removal of phosphate groups and (3) acetylation and deacetylation, the addition or removal of acetyl groups. Phosphorylation and dephosphorylation in particular form a major mechanism by which cells can switch processes on or off or change the flow through a biochemical pathway. Additionally, proteins may be modified non-enzymatically by metabolites, e.g. the modification of various lysine residues by the glycolytic metabolite 1,3-bisphosphoglycerate [29].

3.6 Co-factors

Co factors are molecules or ions that are required for biological functions or reactions to occur. For many proteins, metal ions are central to their mechanism of action. For example, the *N*-methyl D-Aspartate (NMDA) glutamate receptor allows calcium ions into a neuron when both electrical and neurotransmitter signals are received. The Ca^{2+} channel is normally blocked by Mg^{2+}. This block is removed briefly when a previous electrical signal changes the electrical potential of the membrane surrounding the NMDA glutamate receptor. If glutamate binds at this time, the calcium channel opens to allow Ca^{2+} ions into the cell. With no change in electrical potential, glutamate binding cannot open the channel. In effect, Mg^{2+} contributes mechanistically to the way the NMDA receptor senses coincidence in electric and neurotransmitter signals and this process contributes to one mechanism of synaptic plasticity. Other examples of co-factors include small molecules such as

vitamins which are often involved in enzyme reactions as part of the chemical process.

3.7 Protein Complexes

The formation of tightly associated proteins within large complexes is often required for biological activity. An example of this is the endopeptidase γ secretase complex, discussed later, where at least four different proteins are required to form an active enzyme [30]. These include one of the presenilins, either PS1 (UniProt P49768) or PS2 (UniProt P49810), which forms the catalytic core and the proteins Pen-2 (UniProt Q9NZ42), nicastrin (UniProt Q92542) and APH-1 (UniProt Q96BI3) that may contribute to the activation of the protein complex and regulate how the complex interacts with its various substrates [31].

3.8 Environmental Factors

In addition to processes regulated by the cell via gene and protein expression, features such as temperature, pH or redox state associated with the cellular environment may also affect the likelihood of a reaction, for example pH modulates Aβ aggregation [32, 33] and oxidative stress may increase Aβ production and also be increased by Aβ [34].

3.9 Describing Protein Interactions

The properties of affinity and concentration for active forms of a protein in relation to its biological outcomes can be illustrated by dose response curves (Fig. 1). Further, interactions such as enzyme reactions can be described by various kinetic constants such as the affinity constant $K_{(a)}$, the catalytic efficiency $K_{(cat)}$, maximal reaction velocity $V_{(max)}$ and K_m, an inverse measure of affinity defined as the amount of substrate at half V_{max}. These values are calculated from experimental data using equations such as the Michaelis-Menten equation [35] and associated variations. The basic biochemical properties should be captured in any mechanistic model of a molecular pathway. Some pathways will be more complex than others but most will feature these properties in regulatory mechanisms. It must be remembered that molecules and signaling pathways in different cell types may be associated with different functions and these may also vary between species making a generally applicable model of any one molecular pathway impossible.

4 Networks and Their Analysis as Tools to Investigate Complexity in Molecular Pathways

One approach to teasing apart the complexity of molecular pathways is to model molecular interactions as networks to describe and characterize the complex relationships and components within and between pathways. A molecular system can be represented as a graph in the form of a collection of nodes (objects) and edges (relationships). The functional relevance of nodes and edges can be

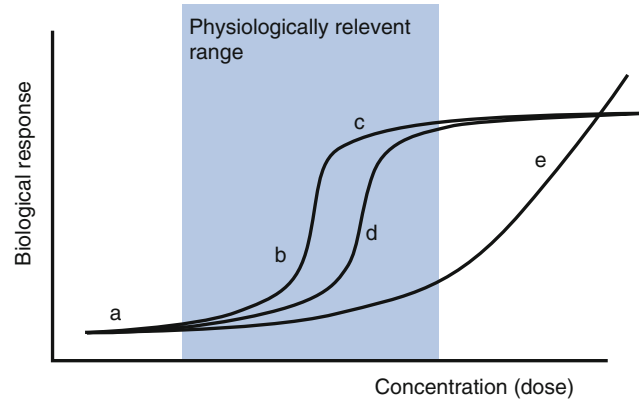


Fig. 1 A generalized dose response curve. Where the concentration of an active protein is very low, the probability that it will interact with its target is also very low and any associated biological outcome will be minimal (**a**). As concentration increases towards a physiologically relevant range, the high affinity biological outcome will also increase (**b**). At a certain point the system is maximally active and any further increase in protein concentration will not increase the high affinity biological outcome as other features of the system may be rate limiting and the biological outcome reaches a steady state (**c**). At increasing concentrations of the active protein, other pathways may become more relevant as the chances of lower affinity reactions increase (**d**); other features of the lower affinity systems may be rate limiting for the relevant biological outcomes which will reach a steady state. At very high concentrations, there are increased chances of aberrant or inappropriate reactions/interactions between the active protein and other pathways with which it would not normally associate (**e**), and these may not be rate limited

described by assigning various attributes derived from the molecular system in question.

Nodes can be used to represent molecules and annotations can represent the various factors such as concentration, affinity and compartmentation. Edges can be either directed, specifying a source (starting point) and a target (endpoint), or non-directed. Directed edges are suitable for representing flow while non-directed edges are used to represent mutual interactions. Mixed graphs contain both directed and undirected edges and have various sets of relations.

A network of molecular relationships can be built in several ways. One way is to iteratively search literature databases using keywords relevant to the system being investigated [36]. An iterative procedure can be used to develop the search strategy, with input from clinician advisors, neuropathologists, information specialists etc. A search of PubMed (28 August 2013) for the keywords systems biology AND Alzheimer disease retrieved 183 results and the increasing number of references over time indicates that the application of systems biology to AD research is of increasing importance.

Not all of these references will be relevant and manual curation will be required. A search of PubMed (28 August 2013) using the MeSH terms ("Systems Biology"[Mesh]) AND "Alzheimer Disease"[Mesh] retrieved 24 results, with some relevant references missing. A comprehensive search of several bibliographic databases as well as hand searches of key journals would also need to be undertaken to ensure all literature would be identified. All titles and abstracts should be screened by two independent reviewers and a third reviewer would resolve any disagreements about inclusion. This underlines the importance of a reliable and repeatable search strategy.

Once a collection of papers has been generated, there are various ways to filter these results to obtain only those papers of interest, involving either automated text search, human search of abstracts or both. Using this approach, networks can be built based on the information available, analyzed and then used to generate questions for further experimentation.

It must be remembered that any defined literature search, whilst being reproducible, may not retrieve all the papers of interest and a manual search of paper references may be required until no more useful references are found. Specific molecules in older literature may not be named in a standard way and in one network construction study [37], two APP interacting proteins were excluded as they could not be identified with certainty due to inconsistent naming. Additionally, only information that is published is available, leading to an unquantifiable bias in network construction due to missing information and this has important consequences for the analysis and interpretation of any resultant molecular network.

Molecular interactions can also be extracted from databases such as those listed in Table 1. While each database may be slightly different, there are now systematic ways to query such databases and extract relevant information in standard formats [38]. However, these databases are built from the existing literature and will therefore share the unquantifiable bias due to missing information. Automated methods of text searching are often used in database construction as they can be fast and repeatable. However, automation can lead to errors of misclassification and manual curation is used in most databases to minimize this. Manual curation can also lead to errors which must be repaired when found.

Most molecular databases are built using data from a variety of sources and are annotated with the experimental system from which the data were derived; this generally includes the species, whether in-vivo or in vitro and the exact method used, such as co-immunoprecipitation, various gene [39, 40] and protein [41, 42] expression systems or co-migration in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), all methods have strengths and weaknesses.

Table 1
Examples of molecular pathway and interaction databases

Database	Database description	Reference/link
MINT	Experimentally verified protein interactions; uses automated literature mining and expert curation	http://mint.bio.uniroma2.it/mint/Welcome.do
IntAct	Molecular interactions derived from literature mining or from direct user submissions; expert curation	http://www.ebi.ac.uk/intact/?conversationContext=1
DIP	Experimentally verified protein interactions; data from a variety of sources including automated literature mining and expert curation	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
KEGG	A collection of databases covering various areas including ontology, genomics and molecular networks	http://www.genome.jp/kegg/
HPRD	Database of human specific protein interactions, expert curation, no automation	http://www.hprd.org
BioGRID	Protein interactions from a number of species models, automatic literature mining and expert curation	http://thebiogrid.org/
STRING	Known and predicted direct and indirect protein interactions; uses automated literature mining and expert curation	http://string-db.org/

The studies listed in Table 2 have approached map construction in different ways, using different combinations of protein-protein interaction (PPI) databases, with different literature searching protocols and different inclusion or exclusion criteria. The networks generated in these studies do not always correspond and different studies highlight different pathways or biological processes, e.g. Fe^{2+} [43], apoptosis [44], or cardiovascular disease/diabetes [3]. Each study has different starting points, inclusion/exclusion criteria and network construction methods, so this lack of agreement is to be expected. It is difficult to assess the degree to which the various starting points, criteria and network construction methods bias results towards an outcome. For example, the study by Soler-Lopez et al. [45] may not represent the interactions of full length APP in the membrane adequately, as many of the extracellular matrix (ECM) proteins that might be expected to interact with APP are excluded due to difficulties involved in expressing them in the experimental microarray used. This may shift the focus of their network more towards intracellular interactions. Given the importance of the various interactions of APP with components of the ECM (*see* Fig. 2), any network excluding such proteins and proteoglycans could be seriously confounded and any findings would have to be interpreted carefully. Additionally, the correspondence between gene expression as mRNA and viable functional proteins within a cell is not absolute, varying from 9 to 87 % depending on which genes are investigated [46].

Table 2
Protein–Protein interaction (PPI) network studies (adapted from [36])

Reference	Selection criteria	Exclusion criteria	Main focus
[37]	Evidence of direct PPI from literature searches	Non-protein molecules and metals, poorly characterized proteins, specific peptides are included as parent genes	Direct PPI involving APP and associated fragments by domain with reference to APP770 isoform; molecular networks with reference to biological processes
[45]	Genes in close proximity with 12 "seed" genes previously associated with AD	Proteins without open reading frames, highly glycosylated proteins, transcription factors, extracellular proteins, proteins with several transmembrane regions	Identification of genes in AD with reference to direct PPI and biological processes
[3]	Co-expressed genes that differ between controls and AD	Probe-sets not mapping to any gene or mapping to hypothetic proteins are removed	Variations in transcriptomes of AD similar to cardiovascular disease and diabetes. Cis regulatory elements identified in several diseases known to co-occur with AD
[51]	Genes with variable expression between human and mouse datasets	Outlier removal, removed datasets with low interspecies expression and connectivity; top 5,000H and 3,000 M connected genes included, rest removed to reduce noise	Mouse and human networks are similar—expression levels more preserved than connectivity, species differences in gene co-expression in astroglia and microglia but not neurons, human specific role of PSEN1 in myelination and evidence of species differences in glial cells linked to neuroinflammation

Molecular networks built from PPI databases or literature searches do not explicitly take into account differences between cell types arising through the processes of differentiation during development which can lead to different susceptibilities of different cell types to neuropathology, such as the well-recognized difference in susceptibility to tau reactive NFT pathology of Ca4, Ca3, Ca2 and Ca1 neurons in the hippocampus as reflected in Braak Staging [47]; a widely accepted semi-quantitative measure of NFT pathology. Different cellular systems may have very different functions depending on cell type: an example of this is the way many cell cycle proteins, involved in regulating cell proliferation, are involved in synaptic plasticity in non-proliferative neurons [2]. The differences between cell types potentially undermine many of the current network approaches, especially where different experimental systems have been used to generate interaction data. Ideally there should be a database for each cell type, and for the brain this would need to include different neuron types as not all neurons necessarily share similar signaling and interaction pathways.

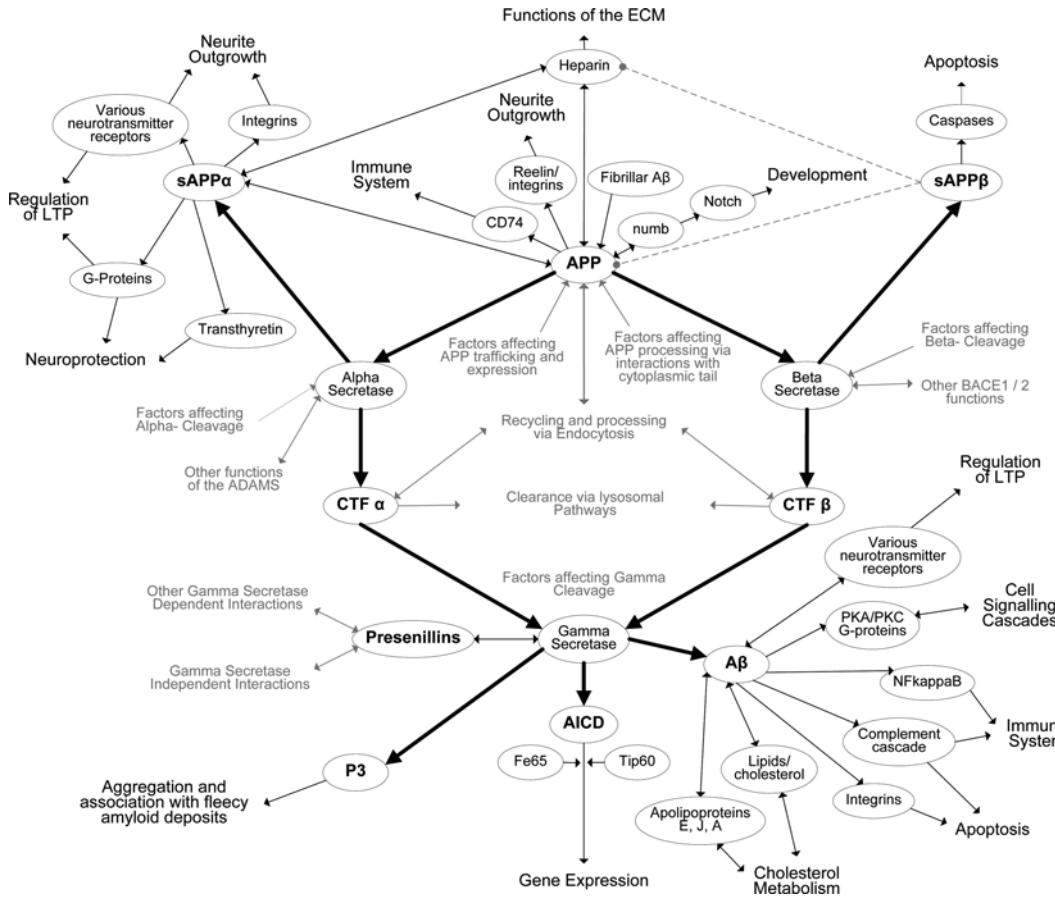


Fig. 2 A simplified view of selected interactions of the APP proteolytic system (adapted from [36]) Nodes represent molecules or molecular assemblies and interactions between them as arrows. Some complex interactions have been collapsed into general processes shown in grey. Multiple sequence variants and conformations of APP and Aβ have been collapsed into a single node for each. Aβ, amyloid beta protein; ADAM, a disintegrin and metalloproteinase domain-containing protein; AICD, APP intracellular domain; APP, amyloid precursor protein; BACE, beta-site amyloid precursor protein cleaving enzyme; CD74, HLA class II histocompatibility antigen gamma chain; CTF, carboxy- terminal fragment; ECM, extracellular matrix; Fe65, Amyloid beta A4 precursor protein-binding family B member 1; LTP, long-term potentiation; PKA, protein kinase A; PKC, protein kinase C; sAPP, secreted amyloid precursor protein; Tip60, Histone acetyltransferase KAT5. With permission from BioMed Central (part of Springer Science + Business Media) under the Open Access License Agreement (http://www.biomedcentral.com/about/license)

A major problem with all molecular map type networks is their inability to include dynamic information relating to the way molecular networks are regulated in living systems. Transient protein modifications, such as phosphorylation, regulate molecular interactions and are central to cellular function are not easily captured, for example, differential phosphorylation of the tyrosine (tyr) residues Tyr_{682} and/or Threonine Thr_{668} of the APP_{695} cytoplasmic domain regulates many interactions with small binding proteins

and kinases [48]. Other dynamic processes that may not be fully represented include transient changes in gene expression via epigenetic mechanisms, changes in protein expression via RNA interference, responses to environmental perturbations such as infection and activity lead changes, such as the up-regulation of synaptic proteins in response to synaptic activity.

Inter-species differences in the way cellular signaling systems are organized, especially in the brain, are well recognized [49, 50] and this should be taken into account when designing animal disease models and building networks. Miller et al. [51] confirm this in the comparison between human and mouse networks, revealing an additional function of PS in oligodendrocytes and myelination in humans that is not seen in the mouse. Given the association of PS mutations in FAD, this difference is likely to impact on the suitability of the mouse as a model for AD.

The development of animal models that represent AD disease processes in humans is crucial in the search for effective therapeutic interventions. Early transgenic mouse models did not completely replicate the neuropathology associated with human disease [52] nor the more fundamental aspects of Aβ biochemistry in humans [50]. Attempts to fully represent AD in humans are on-going with the development of new animal models that can be used to investigate the links between various features of AD. Using multiply transgenic animal models allows the investigation of molecular interactions and signaling pathways involved in different aspects of the disease in a way not possible in humans. For example, the TgF344-AD rat [53] displays oligomeric Aβ species and plaque pathology, tau pathology, behavioral change and neuronal loss, combinations not always present together in other animal models, and this model can be used to study the connections between Aβ and tau pathologies. Different animal models may be used to highlight different aspects of human disease, such as the association between Aβ and cholesterol metabolism in the triple transgenic mouse model over-expressing the sterol regulatory element-binding protein-2 [54] or the relationship between age and cognitive decline in the senescence-accelerated mouse [55].

The success of all animal models depends on being comparable to disease presentation in humans, and this is where the main problems lie. The characteristics of AD in humans are constantly being updated as new disease processes and pathologies are found. Disease processes, such as hippocampal sclerosis [7], or other pathologies such as the Tar-DNA binding protein of 43 kDa (TDP-43) [56] may independently contribute to cognitive status and are yet to be fully characterized in the human population. Population studies highlight the existence of multiple pathologies including contributions from the vascular system in the development of Alzheimer- like dementia in the aging population, with relatively few cases of “pure” AD [23, 57]. Additionally, the

relationship between age, neuropathology and disease is not straight-forward, with many pathologies showing an age related distribution [22].

Fresh human brain tissue that may be of use in functional studies is extremely rare and to a great extent, interaction databases rely on various animal, cell culture and in vitro based models, all of which have yet to be fully characterized with respect to the normal human system. If only animal or cell based systems are used as experimental models, functions that are human specific could be misrepresented or missed completely. The full range of pathologies associated with age and AD in humans still remains to be replicated in any animal model.

5 APP: A Dynamic and Complex Proteolytic System

A review of the complexity of the APP proteolytic system has been described [36]. In summary, APP is a member of a wider family of similar proteins that also includes the APP like proteins (APLP)1 and APLP2 that have significant functional redundancy [58] complicating investigations. It is expressed in various isoforms due to mRNA splicing, with APP_{695} being expressed predominantly in the brain and linked to amyloid deposition. It is a type I, single pass transmembrane protein with diverse functions including associations with cell differentiation [59], neurite outgrowth [60, 61], cell adhesion [62], synapse formation, maintenance and plasticity [62, 63] and many cell signaling pathways [36, 64, 65] including apoptosis [66]. APP is post-translationally glycosylated [67] and phosphorylated [48] at various residues and these modifications may contribute to the regulation of the various APP functions and proteolytic pathways.

Full length APP has a large N-terminal domain that interacts with various components of the ECM including heparin and other proteoglycans [68, 69], other proteins such as reelin [70], DAB1 [71] and also forms homodimers regulated by heparin and Zn^{2+} [72]. The transmembrane region has been implicated in the process of homodimerization and also interacts with various proteins including Notch [73]. The C-terminal domain of full length APP also interacts functionally with a variety of proteins including FE65 [74], the low density lipoprotein receptor protein (LRP) [75, 76], a variety of small binding proteins [48, 77] and several kinases [48, 78, 79] that phosphorylate the residues Y_{682} of the binding and signaling sequence $GY_{682}ENPTY$ and T_{668} of APP_{695} [48, 79, 80]. Phosphorylation regulates the interaction of the C-terminal domain with other proteins [48, 77], may modulate proteolytic processing [80] and allows cross talk between diverse cellular systems [48].

Full length APP can remain at the cell surface, be recycled via endocytosis or proteolytically processed and has a high turnover, with a half-life ranging from 30 min [76, 81–83] to 4 h [84–86]. Unprocessed APP is degraded or recycled via the endosomal or lysosomal pathways and may be recycled back to the membrane and processed within ~30 min [82], with perhaps one third to one half being processed via the cleavage pathways as measured by secreted sAPPα/β [82]. APP is proteolytically processed to more than 40 fragments [87]. There are two main cleavage pathways, α- and β- pathways that then converge on a shared γ-cleavage, summarized in Fig. 2. These cleavages have been well reviewed [88, 89]. Additional cleavage pathways (not shown in Fig. 2) include caspase cleavages producing an alternative C-terminal cytoplasmic fragment C31 that is associated with apoptosis [90, 91] and the alternative cleavages by β-site APP cleaving enzyme (BACE)1, 11 residues within the Aβ sequence [88, 92] leaving a membrane bound fragment C88 and BACE2 at the θ-cleavage site between the phenylalanine residues, F_{615} and F_{616} of APP_{695} downstream of both the Aβ and P3 cleavage sites, producing a membrane bound fragment C80 [93].

5.1 α Cleavage

α-cleavage occurs between residues Lys_{612} and Leu_{613} within the Aβ sequence of APP_{695}, releasing the N-terminal sAPPα and leaving a membrane bound C83 C-terminal fragment [88]. α-secretase activity has been observed by several membrane-anchored zinc-dependent metalloproteinase enzymes including A Disintegrin and Metalloproteinase (ADAM)9, ADAM10, ADAM17 [94–96] and possibly the matrix metalloproteinase (MMP)9 [97]. α-cleavage is both constitutive and regulated, with the various ADAMs responding in different ways depending on many factors [95, 98]. In addition to APP, α-secretases also cleave alternative substrates such as Notch [99], pro-TNF-α and the epidermal growth factor receptor [100] which may lead to competition between different pathways with consequences for many cellular processes including development, synaptic plasticity and the cell cycle and cancer [96, 100, 101]. How the balance between these alternative pathways is regulated is not known.

The soluble N-terminal fragment released by α-cleavage, sAPPα, retains two heparin binding sites and has been shown to bind heparin as a dimer [102]. The ability of sAPPα to disrupt APP dimerization at the cell surface may contribute to its neuroprotective actions [103–105] and may partly explain why sAPPα is ~100× more neuroprotective against excytotoxicity, glucose deprivation and the addition of Aβ in hippocampal cultures than sAPPβ, which lacks the second C-terminal heparin binding site [104]. Additionally, neuroprotective actions of sAPPα may be mediated by its antagonism of stress signaling by the JNK stress signaling pathway [106]. Dementia status has been associated with both reduced sAPPα

levels in human CSF [107] and an increased half-life of sAPPα [86] in transgenic mice, however, as yet, there has been no systematic study of the α-pathway proteolytic fragments in the human population.

5.2 β Cleavage

β cleavage occurs between residues Met_{596} and Asp_{597} of APP_{695} within the second heparin binding site, releasing the N-terminal sAPPβ from the membrane bound C99 C-terminal fragment [88, 92]. Two membrane bound aspartyl proteases are associated with β-cleavage, BACE 1 and to a lesser extent, BACE2 [88, 92]. Additionally, Cathepsins D and B have shown β-cleavage activity to release Aβ [108]. BACE1 and BACE2 are differentially regulated and have different functions [109]. In addition to APP, BACE1 may also cleave alternative substrates including APLP1 and APLP2 [110] and P-selectin glycoprotein ligand-1 [111]. Heparin and heparin sulfates may be involved in regulating APP cleavage by BACE1 [112]. In addition to interactions with sAPPα and APP, the large soluble sAPPβ fragment may be associated with apoptotic signaling and axonal degeneration via the death receptor DR6 and caspase6 [113], though the interactions of sAPPβ are not fully characterized and require further detailed investigation.

5.3 γ Cleavage

Cleavage of APP by the γ-secretase complex occurs within the membrane to release the variable length 38–46 residue Aβ peptide following β-cleavage, the variable length 21–29 residue P3 (Aβ17-X) fragment following α-cleavage, with both pathways releasing the APP intracellular domain, (AICD) [8, 88, 114, 115]. There is some uncertainty as to how γ-cleavage occurs; γ-secretase cleavage may occur via successive ζ and ε cleavages producing progressively shorter Aβ fragments [116–118], though there may also be distinct cleavage mechanisms that may be separately modulated [119].

There are a number of alternative γ-secretase substrates, e.g. APLP1, APLP2, Notch, cadherins, LRP [120, 121], and syndecan-1 [114, 122]. In addition to γ-secretase dependent functions, some PS functions are independent of γ-secretase, so that in effect, γ-secretase may compete for presenilins with other γ-secretase independent PS functions including cell adhesion, trafficking of various proteins [123], and Ca^{2+} homeostasis [114]. How the γ-secretase is regulated between the different substrates is not fully understood but may involve other binding proteins such as numb [65] and Rac1 [124], regulation of PS trafficking, including a possible reciprocal interaction with APP [125] and localization of PS within specific organelles and cellular membrane compartments [126].

Aβ is produced in a range of sequence lengths [87] and can form monomers, dimers, oligomers and fibrils [8] which have been difficult to study due to their dynamic instability [127]. At physiological concentrations Aβ is associated with numerous normal

cellular functions [128] and in AD progression has multiple interactions that have been described as either neuroprotective or neurotoxic [36]. It is deposited in the brain in various pathological forms including CAA, diffuse and cored senile plaques and is often associated with neuritic plaques. Different sequence lengths have different propensities to aggregate [32, 129] and aggregation is also affected by amino acid substitution in mutant forms [130, 131] and various factors such as proximity to membranes [132], and pH or metal ion availability [133]. Different sequence lengths and different aggregation states can have different functional roles [36], making investigations into the exact roles of Aβ in the brain difficult. These associations may be better approached experimentally as a matrix, where the various sequence lengths, aggregation states and mutant forms should be assessed for each interaction.

While it is likely that P3 is produced in alternative sequence lengths following γ cleavage, very little evidence can be found in the literature for the contributions of P3 to disease progression. There is currently little interest in characterizing the contributions of P3 to normal brain function or AD, even though P3 is known to aggregate [134–136], has been associated with in cotton wool type amyloid plaques [137] enhances the aggregation of Aβ1-40 [138] and may have a signaling role in apoptosis via caspase activation [139].

Regulation of expression and proteolysis of APP involves multiple factors, some of which are summarized in Fig. 2 (adapted from [36]). How signals from these multiple factors in various cellular locations are integrated to produce a specific outcome in any one cell is not known. Regulation of APP proteolysis, from both outside and within the APP proteolytic system, can be in response to a wide variety of cellular signals and various modulators including glycosylation, phosphorylation, dimerization, associations with heparin glycoproteins and other binding proteins. Feedback routes can be simple and short range such as the promotion of APP expression associated with fibrillar Aβ and prion protein [140]. Indirect and complex feedback routes also exist, such as the effects of heparin on regulating β-cleavage with low concentration promoting and high concentration inhibiting the activation of BACE1 [141] and the effects of Aβ on heparin. Aβ interacts with heparins in the ECM and at high levels may prevent the catabolism of proteoglycans and promote amyloid formation [142]. Reciprocally heparins modulate many of the interactions involving Aβ such as enhancing both nucleation and elongation processes in the aggregation of Aβ [143], limiting the neurotoxic and pro-inflammatory activity of Aβ in a dose dependent manner [144] and contributing to the uptake of Aβ by a pathway shared with ApoE [145].

6 Modeling the APP Proteolytic System. Practical Considerations

As a summary of interactions, maps, such as Fig. 2, can highlight particular areas that may be of interest such as hubs or regulatory interactions that may be open to modification by medications, or may highlight areas where data are missing, leading to further research. While molecular networks involving APP can be constructed, how these relate to the actual network of molecular interactions in any one human cell type at any one stage of development cannot yet be fully assessed. As reviewed above, different criteria and network construction methods can generate different networks, each with strengths, weaknesses and different behaviors in analysis. The impact of missing data, such as interactions that have not yet been identified, is difficult to assess. For the APP network, the contributions of alternative proteolytic fragments, such as sAPPα, sAPPβ, P3 and the various longer Aβ fragments, e.g. Aβ43, Aβ45, Aβ46 and Aβ48, in various states of aggregation have yet to be fully described. It is still unclear which Aβ sequence or aggregation state is linked to disease progression [146]. These alternative fragments may yet provide further interactions that have the potential to affect network behavior as a whole, as suggested by the predisposition to form Aβ42 from γ cleavage due to the accumulation of γ secretase substrates, C99 and longer Aβ fragments [147].

There are great difficulties in representing an iterative and dynamic proteolytic system, such as APP, as a static network map of connections. One of the first questions raised is what exactly does a static network represent? If a network represents interactions, and these interactions change with protein modifications such as phosphorylation, is it best to represent each functional protein version as a separate node? Should the alternative isoforms of APP be included and if so, should they have separate nodes? How do we best represent Aβ with around 40 possible sequence lengths [87] and various states of aggregation [32, 146]? In Fig. 2, Aβ has been collapsed into a single node for clarity. How would over 40 nodes in this space with potentially different connections affect computational and analytical methods? Given the different conformations [148] and functional actions [149–151] of Aβ(1–40) and Aβ(1–42), a single node for these peptides cannot fully represent the APP functional network.

If the aim is to understand the role of PS in AD, perhaps with a view to developing treatment strategies that modulate its probability to react between its various substrates, then a network of its interactions could be constructed and this could be the basis for a dynamic computational model. This dynamic model would need to include calculations of a protein's probability of reaction, where the basic molecular features described previously, (concentration, half-life etc.), could be represented as values in a computational matrix.

Table 3
Kinetic values for human synthetic wild-type PS1. N/A, not available

Reaction	K_m	V_{max}	Cell/model system	Ref
APP C99 → Aβ and AICD Notch (S3) → NICD	0.40 ± 0.05 μM (C99) 1.08 ± 0.17 μM	175.6 ± 8.4 pM/min (AICD) 95.7 ± 7.5 pM/min (NICD)	Mouse embryonic fibroblasts (MEF) derived membrane cell free assay system	[152]
APP C99 → AICD	874 ± 252 nM (AICD)	15 ± 1.82 nM/h (AICD)	MEF derived membrane cell free assay system	[147]
APP C99 → Aβ	N/A	Maximal activity: 217 ± 110 pM/106 cells (Aβ)	HeLa cells transfected with APP WTC99 cDNA construct	[153]
Aβ42 → Aβ38	370 ± 40 nM	N/A	Hek293 cells transfected with wild type PS1	[118]

This approach could be developed iteratively and different versions of a network could be compared in terms of flow through the network. Experimental data relating to basic biomolecular properties that are relevant to modeling the probability that a reaction will occur can be extracted from the literature, including V_{max}, K_m, and $K_{(cat)}$. However, characterizing enzyme reactions in order to model the probability of reaction is not an easy task as demonstrated in the following example.

Recent studies [118, 147, 152, 153] have looked at γ secretase enzyme kinetics for a variety of PS mutations, substrates and products. Different experimental models have been used and different features of the system have been reported in different formats. Table 3 gives values for K_m, and V_{max} for human synthetic wild-type PS1 and its interaction with various substrates extracted from the associated references.

Values for K_m have been given in μM or nM and values for maximum reaction rate have been given as V_{max} (pM/min or nM/h) or maximal activity (pM/10^6 cells). While manual extraction from the literature could easily convert μM to nM or nM/h to pM/min, automated text based searches could introduce error due to units reported. It is not possible to convert pM/10^6 cells into nM/h or pM/min, making comparisons between these studies difficult. The degree to which the experimental system used affects the values gained is difficult to assess, mouse embryonic fibroblast (MEF) derived membrane cell free assays, Hek293 or HeLa cell based systems are likely to have very different environments and each system will have experimental advantages and disadvantages. None of these systems accurately represent aging in the human brain. Indeed, which values of K_m and V_{max} in Table 3 would be most representative

of the situation in any human neuron? Standard reporting formats for proteomic data exist [154, 155] and are annotated by experimental system used to derive the information such as species used, etc. so that data from different studies can be integrated but it is difficult to choose those values that may best represent the human system as it has not yet been fully characterized.

Attempts to dynamically model the human cognitive system are on-going with a diversity of approaches. For example, Kasabov et al. [156, 157] have combined gene and protein expression networks with a probabilistic spiking neural network and compared this to real human electroencephalograms [158] and used this to investigate pathways involved in AD [157]. In these models, dynamic behavior is captured in the network output, represented as spiking neurons, which can be controlled by networks representing gene and protein expression data. These gene and protein networks are in turn re-modeled iteratively by the spiking neural network. While a computational model of the AD process would be very useful to investigate how the system might be perturbed by changes to gene and protein expression, their current usefulness is open to question. Connectionist network models contain unquantifiable modules, as the weights of connections between the nodes in a network are stochastically modified during the training process. The relationships between the nodes and weighted connections with any feature of the human system are not certain: the nodes do not necessarily represent real human neurons and the connections do not necessarily represent connections between neurons. Populations of trained networks will consist of individual network models, each of which will have different connection weights. The difficulty here is in relating the distributions of the weights in any network to the living human system: the extraction of potentially useful information from the structure of the network is problematic.

7 Applying Systems Biology Approaches in Other Areas

7.1 Pattern Recognition and the Early Diagnosis of AD

Various computational methods such as principle component analysis [159, 160], linear regression methods [161, 162], machine learning methods [163–165] and random forests [166] are being used to investigate automated pattern recognition in magnetic resonance imaging (MRI) image analysis [161] or various imaging methods coupled with multiple biomarker analysis [160, 163, 165, 166] with some success in separating normal aging from mild cognitive impairment (MCI) and AD. Although the use of new computational methods for multiple markers for AD increases the specificity and sensitivity in categorizing normal aging, MCI or AD, there is still no combination of markers that can identify those with MCI that may convert to dementia and AD with certainty and this is an urgent requirement.

7.2 The Human Connectome Project

Beyond mapping gene and protein expression or interaction networks, the effects of the human connectome on dementia risk is another complex area that presents huge challenges. The human connectome project [167] aims to map the human connectome at the macroscopic scale, (~1 mm^3) using a variety of neuroimaging methods. This project aims to create a map of healthy human connectivity. There is great inter-individual heterogeneity, both in the vascular system, that may affect certain imaging methods and in cortical folding, so any resultant map can only be an idealized reference map. Further, how this connectivity changes with progression in dementia may also be highly heterogeneous between individuals and this has yet to be fully investigated.

8 Relating the Systems Biology of APP to Normal Cognition and Disease Progression in AD

For any neuron, signals received via synapses must be integrated into dynamic responses of the cell as a whole and this requires signaling between any specific synapse on a dendrite and its nucleus, possibly located some distance from the synapse. Changes to gene and protein expression in response to synaptic signaling must be transmitted back to the synapse via protein trafficking so that receptors and signaling molecules are in the correct cellular positions. There may be different signals arriving via different pathways, both electrical and metabolic, and these must be integrated into a coherent neuronal response. There is a temporal coherence, where everything must be in the right place at the right time, as the synaptic response builds on the previous state of the synapse. These synapses are further organized within a neural network connectome of different cell types and different functional brain areas from which cognition and human behavior arise that may include inputs from the whole body as it interacts with its environment. Figure 3 illustrates the interdependence of the areas involved in normal brain function, where gene expression may be modified by behavior which in turn may change protein expression and interaction leading to further changes in behavior as the whole system iteratively and stochastically changes over time. Attempts to isolate any specific area, such as protein expression, can be undermined by this interdependence and contributions to cognitive processes may be misrepresented, simply due to the assumptions of independence in experimental design.

In order to understand this coherent system, research has necessarily had to break it into smaller parts giving rise to discrete research fields investigating all the areas involved from genomes and proteomes to interactomes and connectomes. Traditionally, the reductionist approach aims to characterize individual pathways by introducing changes that are meant to impact on specific

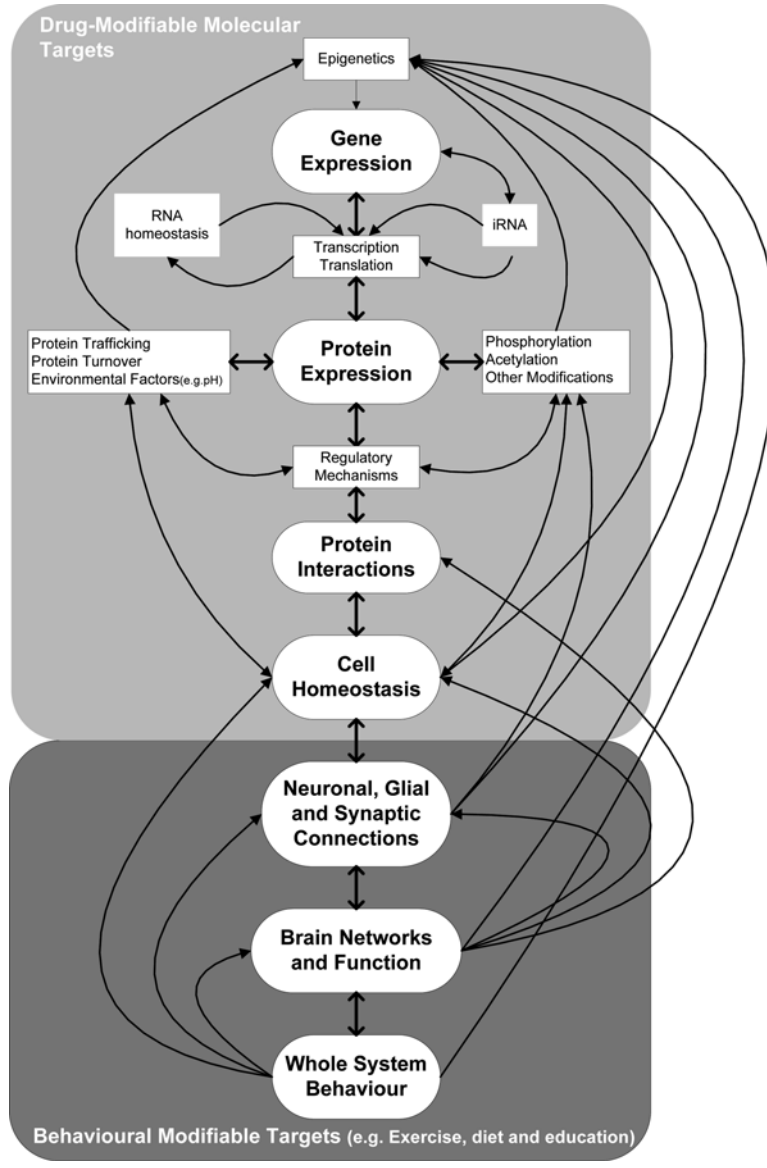


Fig. 3 The interdependence between 'omics research areas. General discrete research areas discussed in the text appear as nodes, selected feedback and feed forward relationships are shown as *arrows*

components in potentially well understood ways. This can lead to a limited view of complex processes, for example, the amyloid cascade hypothesis suggests that Aβ, in some form, is the sole cause of AD and that therefore removal of Aβ should modify the disease course. This can be understood in terms of a more linear infection type model. However, treatments based on this model have been unsuccessful in clinical trials so far and have failed to change the

course of the disease [168], questioning its validity. Population studies highlight complexity in the presentation of AD, bringing wider research areas such as aging, diet, exercise, education, the vascular system and other biochemical pathways into consideration. Few complex biological mechanisms can be reduced to simple in vitro, cell based or animal based experimental models [50] and poorly characterized or unsuitable experimental systems may lead to erroneous interpretations.

In contrast to reductionist approaches, in which molecular systems may be treated as isolated and independent mechanisms, systems biology aims to integrate evidence from diverse areas into a representation of living processes as a whole. Even simple molecular systems present enormous challenges in terms of modeling biological outcomes as theories of molecular communication, i.e. how biological information is transmitted through a molecular network, are still being developed [25] and any computational representations of biological processes are necessarily limited to the data we currently have. In complex maps of protein interactions, many pathways are possible and whether any specific interactions are central, peripheral or involved in only subtypes of disease progression cannot yet be fully assessed.

Integrating networks constructed at the level of gene expression, with networks constructed at the levels of protein expression, protein interaction and cellular behavior is currently difficult as there isn't correspondence between them. As reviewed above, not all genes expressed as mRNA transcripts become functional proteins and not all functional proteins necessarily interact due to dynamic regulation. Additionally, while the human connectome is being mapped at ever increasing resolution [167], how information is represented and stored across the human brain as a dynamic neural system of synaptic connections and how this changes with disease progression is not known.

Given that there is no qualitative marker for AD, diagnosis has relied on various clinical [4, 5] and neuropathological [6, 7] criteria that are quantitative and involve the application of thresholds: no single measure yet defines AD. Further, biomarkers used in the diagnosis of clinical disease remain to be standardized and harmonized [169]. Aβ fragments, commonly employed as biomarkers of disease, may have both protective and aberrant behaviors associated with disease and multiple disease pathways may exist. Additionally, no Aβ fragment has been identified as the "neurotoxic" disease related species [146]. How can poorly defined neurodegenerative diseases be diagnosed at an early stage when treatment strategies could have the best chance of preserving cognitive functions? This has consequences for how we understand AD, whether it is a single process that will respond to a single intervention strategy or whether AD is a syndrome, requiring multiple different interventions depending on disease types, yet to be characterized.

This is of great importance to the design of experimental investigations and clinical trials. Selection of participants and controls relies on how we understand the disease process and how any disease process is reflected in clinical markers. How do we know that in any given clinical trial, the subjects selected represent homogenous disease or non-disease groups? There may be other disease processes, such as hippocampal sclerosis [7] and other pathologies such as TDP-43 [56] that may contribute to disease pathways and are yet to be fully characterized. Additionally, individuals may vary in the degree to which cognitive reserve and compensation to neuronal injury may limit the impact of pathological changes that occur during aging to better preserve cognitive functions [170].

In terms of health care planning, given the lack of progress towards a reliable dementia treatment strategy, in the immediate future perhaps dementia prevention and dementia care are areas where progress can best be made. The association of education [18], exercise [19] and diet [20] throughout life with a lower dementia risk in old age suggests that Public Health strategies devised to promote these activities would be worthwhile. Without a cure or ameliorating treatment, we need to be able to care for dementia sufferers in the most appropriate and efficient manner to maintain an individual's independence and quality of life for as long as possible.

While applying the systems biology approach to represent complex dynamic proteolytic systems such as APP may not yet be entirely feasible, useful perspectives can still be generated. For APP, the complexity of its interactions and regulatory features suggest that multiple initiation and progression pathways are possible: the analysis of networks to highlight those disease pathways that may be most likely to occur in humans presents major challenges. Capturing this complexity in any network model and being able to relate network behavior to real human brains is the ultimate goal. Whether it will ever be possible to build a dynamic model of the AD disease processes at all levels of consideration (genome, proteome, interactome, connectome and whole body) is not clear. There is no best way to build a network and all networks constructed so far are incomplete. Additionally, both AD and normal aging in humans have yet to be fully characterized. How this missing data impacts on the reliable prediction of events from an incomplete network cannot yet be known. However, this chapter suggests some initial steps and proposals on how we could build more sophisticated networks. The challenge to the AD biomedical research community is to iteratively integrate data generated via a variety of approaches, both reductionist and systems biology, and then to use any insights gained to integrate the information and design further experiments to generate new data. It is clear that no single approach, reductionist or systems biology can tackle this problem alone.

Acknowledgements

The authors were supported by the Cambridgeshire and Peterborough Collaboration for Leadership in Applied Health Research and Care (CLAHRC).

References

1. Tacutu R, Budovsky A, Yanai H, Fraifeld VE (2011) Molecular links between cellular senescence, longevity and age-related diseases—a systems biology perspective. Aging 3:1178–1191
2. Arendt T (2003) Synaptic plasticity and cell cycle activation in neurons are alternative effector pathways: the 'Dr. Jekyll and Mr. Hyde concept' of Alzheimer's disease or the yin and yang of neuroplasticity. Prog Neurobiol 71:83–248
3. Ray M, Ruan J, Zhang W (2008) Variations in the transcriptome of Alzheimer's disease reveal molecular networks involved in cardiovascular diseases. Genome Biol 9:R148
4. Morris JC, Heyman A, Mohs RC et al (1989) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 39:1159–1165
5. McKhann GM, Knopman DS, Chertkow H et al (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7:263–269
6. Mirra SS, Heyman A, McKeel D et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41: 479–486
7. Hyman BT, Phelps CH, Beach TG et al (2012) National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 8:1–13
8. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81:741–766
9. Zekry D, Duyckaerts C, Belmin J et al (2003) Cerebral amyloid angiopathy in the elderly: vessel walls changes and relationship with dementia. Acta Neuropathol 10:367–373
10. Attems J, Jellinger K, Thal DR, Van Nostrand W (2011) Review: sporadic cerebral amyloid angiopathy. Neuropathol Appl Neurobiol 37: 75–93
11. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174
12. Pedersen NL, Gatz M, Berg S, Johansson B (2004) How heritable is Alzheimer's disease late in life? Findings from Swedish twins. Ann Neurol 55:180–185
13. Cedazo-Minguez A, Cowburn RF (2001) Apolipoprotein E: a major piece in the Alzheimer's disease puzzle. J Cell Mol Med 5:254–266
14. Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41: 1088–1093
15. Crehan H, Holton P, Wray S et al (2012) Complement receptor 1 (CR1) and Alzheimer's disease. Immunobiology 217:244–250
16. Bertram L, Tanzi RE (2012) The genetics of Alzheimer's disease. Prog Mol Biol Transl Sci 107:79–100
17. Schellenberg GD, Montine TJ (2012) The genetics and neuropathology of Alzheimer's disease. Acta Neuropathol 124:305–323
18. Brayne C, Ince PG, Keage HA et al (2010) Education, the brain and dementia: neuroprotection or compensation? Brain 133: 2210–2216
19. Ma Q (2008) Beneficial effects of moderate voluntary physical exercise and its biological mechanisms on brain health. Neurosci Bull 24:265–270
20. Mattson MP, Chan SL, Duan W (2002) Modification of brain aging and neurodegenerative disorders by genes, diet, and behavior. Physiol Rev 82:637–672
21. Xuereb JH, Brayne C, Dufouil C et al (2000) Neuropathological findings in the very old. Results from the first 101 brains of a population-based longitudinal study of dementing disorders. Ann N Y Acad Sci 903:490–496
22. Savva GM, Wharton SB, Ince PG et al (2009) Age, neuropathology, and dementia. N Engl J Med 360:2302–2309

23. MRC-CFAS (2001) Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). Lancet 357: 169–175
24. Brayne C, Richardson K, Matthews FE et al (2009) Neuropathological correlates of dementia in over-80-year-old brain donors from the population-based Cambridge city over-75 s cohort (CC75C) study. J Alzheimers Dis 18:645–658
25. Nakano T, Moore MJ, Wei F et al (2012) Molecular communication and networking: opportunities and challenges. IEEE Trans Nanobioscience 11:135–148
26. Jensen ON (2006) Interpreting the protein language using proteomics. Nat Rev Mol Cell Biol 7:391–403
27. Ratovitski T, Slunt HH, Thinakaran G et al (1997) Endoproteolytic processing and stabilization of wild-type and mutant presenilin. J Biol Chem 272:24536–24541
28. da Costa CA, Ancolio K, Checler F (1999) C-terminal maturation fragments of presenilin 1 and 2 control secretion of APP alpha and A beta by human cells and are degraded by proteasome. Mol Med 5:160–168
29. Ogishima S, Mizuno S, Kikuchi M et al (2013) A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery. Clin Pharmacol Ther 93:399–401
30. Kimberly WT, LaVoie MJ, Ostaszewski BL et al (2003) Gamma-secretase is a membrane protein complex comprised of presenilin, nicastrin, Aph-1, and Pen-2. Proc Natl Acad Sci U S A 100:6382–6387
31. Kimberly WT, Wolfe MS (2003) Identity and function of gamma-secretase. J Neurosci Res 74:353–360
32. Kirkitadze MD, Condron MM, Teplow DB (2001) Identification and characterization of key kinetic intermediates in amyloid beta-protein fibrillogenesis. J Mol Biol 312: 1103–1119
33. Rubinstein A, Lyubchenko YL, Sherman S (2009) Dynamic properties of pH-dependent structural organization of the amyloidogenic beta-protein (1–40). Prion 3:31–43
34. Smith MA, Hirai K, Hsiao K et al (1998) Amyloid-beta deposition in Alzheimer transgenic mice is associated with oxidative stress. J Neurochem 70:2212–2215
35. Berg JM, Tymoczko JL, Stryer L (2006) Biochemistry, 6th edn. W. H. Freeman and Company, New York, NY
36. Hunter S, Brayne C (2012) Relationships between the amyloid precursor protein and its various proteolytic fragments and neuronal systems. Alzheimers Res Ther 4:10
37. Perreau VM, Orchard S, Adlard PA et al (2010) A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. Proteomics 10:2377–2395
38. Aranda B, Blankenburg H, Kerrien S et al (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Methods 8:528–529
39. Murphy D (2002) Gene expression studies using microarrays: principles, problems, and prospects. Adv Physiol Educ 26:256–270
40. Baranzini SE (2004) Gene expression profiling in neurological disorders: toward a systems-level understanding of the brain. Neuromolecular Med 6:31–51
41. Hu S, Xie Z, Qian J et al (2011) Functional protein microarray technology. Wiley Interdiscip Rev Syst Biol Med 3:255–268
42. Sutandy FX, Qian J, Chen CS, Zhu H (2013) Overview of protein microarrays. Curr Protoc Protein Sci. Editorial board, John E Coligan et al. Chapter 27: Unit 27.1. doi:10.1002/0471140864.ps2701s72
43. Kell DB (2010) Towards a unifying, systems biology understanding of large-scale cellular death and destruction caused by poorly liganded iron: Parkinson's, Huntington's, Alzheimer's, prions, bactericides, chemical toxicology and others as examples. Arch Toxicol 84:825–889
44. Alberghina L, Colangelo AM (2006) The modular systems biology approach to investigate the control of apoptosis in Alzheimer's disease neurodegeneration. BMC Neurosci 7(Suppl 1):S2
45. Soler-Lopez M, Zanzoni A, Lluis R et al (2011) Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. Genome Res 21:364–376
46. Suarez RK, Moyes CD (2012) Metabolism in the age of 'omes'. J Exp Biol 215:2351–2357
47. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82:239–259
48. Tamayev R, Zhou D, D'Adamio L (2009) The interactome of the amyloid beta precursor protein family members is shaped by phosphorylation of their intracellular domains. Mol Neurodegener 4:28
49. Raiteri M (2006) Functional pharmacology in human brain. Pharmacol Rev 58:162–193
50. Kokjohn TA, Roher AE (2009) Amyloid precursor protein transgenic mouse models and

Alzheimer's disease: understanding the paradigms, limitations, and contributions. Alzheimers Dement 5:340–347

51. Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 107: 12698–12703
52. Codita A, Winblad B, Mohammed AH (2006) Of mice and men: more neurobiology in dementia. Curr Opin Psychiatry 19:555–563
53. Cohen RM, Rezai-Zadeh K, Weitz TM et al (2013) A transgenic Alzheimer rat with plaques, tau pathology, behavioral impairment, oligomeric abeta, and frank neuronal loss. J Neurosci 33:6245–6256
54. Barbero-Camps E, Fernandez A, Martinez L et al (2013) APP/PS1 mice overexpressing SREBP-2 exhibit combined Abeta accumulation and tau pathology underlying Alzheimer's disease. Hum Mol Genet 22:3460–3476
55. Butterfield DA, Poon HF (2005) The senescence-accelerated prone mouse (SAMP8): a model of age-related cognitive decline with relevance to alterations of the gene expression and protein abnormalities in Alzheimer's disease. Exp Gerontol 40:774–783
56. Davidson YS, Raby S, Foulds PG et al (2011) TDP-43 pathological changes in early onset familial and sporadic Alzheimer's disease, late onset alzheimer's disease and Down's syndrome: association with age, hippocampal sclerosis and clinical phenotype. Acta Neuropathol 122:703–713
57. Keage HA, Ince PG, Matthews FE et al (2012) Impact of less common and "disregarded" neurodegenerative pathologies on dementia burden in a population-based cohort. J Alzheimers Dis 28:485–493
58. Heber S, Herms J, Gajic V et al (2000) Mice with combined gene knock-outs reveal essential and partially redundant functions of amyloid precursor protein family members. J Neurosci 20:7951–7963
59. De Strooper B, Annaert W (2000) Proteolytic processing and cell biological functions of the amyloid precursor protein. J Cell Sci 113: 1857–1870
60. Small DH, Clarris HL, Williamson TG et al (1999) Neurite-outgrowth regulating functions of the amyloid protein precursor of Alzheimer's disease. J Alzheimers Dis 1:275–285
61. Small DH, Nurcombe V, Reed G et al (1994) A heparin-binding domain in the amyloid protein precursor of Alzheimer's disease is involved in the regulation of neurite outgrowth. J Neurosci 14:2117–2127
62. Priller C, Bauer T, Mitteregger G et al (2006) Synapse formation and function is modulated by the amyloid precursor protein. J Neurosci 26:7212–7221
63. Hoe HS, Rebeck GW (2008) Functional interactions of APP with the apoE receptor family. J Neurochem 106:2263–2271
64. Okamoto T, Takeda S, Murayama Y, Ogata E, Nishimoto I (1995) Ligand-dependent G protein coupling function of amyloid transmembrane precursor. J Biol Chem 270: 4205–4208
65. Roncarati R, Sestan N, Scheinfeld MH et al (2002) The gamma-secretase-generated intracellular domain of beta-amyloid precursor protein binds Numb and inhibits notch signaling. Proc Natl Acad Sci U S A 99: 7102–7107
66. Kogel D, Schomburg R, Schurmann T et al (2003) The amyloid precursor protein protects PC12 cells against endoplasmic reticulum stress-induced apoptosis. J Neurochem 87:248–256
67. Georgopoulou N, McLaughlin M, McFarlane I, Breen KC (2001) The role of post-translational modification in beta-amyloid precursor protein processing. Biochem Soc Symp 67:23–36
68. Bush AI, Pettingell WH Jr, de Paradis M et al (1994) The amyloid beta-protein precursor and its mammalian homologues. Evidence for a zinc-modulated heparin-binding superfamily. J Biol Chem 269:26618–26621
69. Williamson TG, Nurcombe V, Beyreuther K et al (1995) Affinity purification of proteoglycans that bind to the amyloid protein precursor of Alzheimer's disease. J Neurochem 65:2201–2208
70. Hoe HS, Lee KJ, Carney RS et al (2009) Interaction of reelin with amyloid precursor protein promotes neurite outgrowth. J Neurosci 29:7459–7473
71. Hoe HS, Tran TS, Matsuoka Y et al (2006) DAB1 and Reelin effects on amyloid precursor protein and ApoE receptor 2 trafficking and processing. J Biol Chem 281:35176–35185
72. Dahms SO, Hoefgen S, Roeser D et al (2010) Structure and biochemical analysis of the heparin-induced E1 dimer of the amyloid precursor protein. Proc Natl Acad Sci U S A 107:5381–5386
73. Fassa A, Mehta P, Efthimiopoulos S (2005) Notch 1 interacts with the amyloid precursor protein in a Numb-independent manner. J Neurosci Res 82:214–224
74. McLoughlin DM, Miller CC (2008) The FE65 proteins and Alzheimer's disease. J Neurosci Res 86:744–754
75. Rebeck GW, Moir RD, Mui S et al (2001) Association of membrane-bound amyloid precursor protein APP with the apolipoprotein E

receptor LRP. Brain Res Mol Brain Res 87:238–245

76. Pietrzik CU, Busse T, Merriam DE et al (2002) The cytoplasmic domain of the LDL receptor-related protein regulates multiple steps in APP processing. EMBO J 21:5691–5700
77. Taru H, Suzuki T (2009) Regulation of the physiological function and metabolism of AbetaPP by AbetaPP binding proteins. J Alzheimers Dis 18:253–265
78. Muresan Z, Muresan V (2005) c-Jun NH2-terminal kinase-interacting protein-3 facilitates phosphorylation and controls localization of amyloid-beta precursor protein. J Neurosci 25:3741–3751
79. Suzuki T, Nakaya T (2008) Regulation of amyloid beta-protein precursor by phosphorylation and protein interactions. J Biol Chem 283:29633–29637
80. Lee MS, Kao SC, Lemere CA et al (2003) APP processing is regulated by cytoplasmic phosphorylation. J Cell Biol 163:83–95
81. Perez RG, Soriano S, Hayes JD et al (1999) Mutagenesis identifies new signals for beta-amyloid precursor protein endocytosis, turnover, and the generation of secreted fragments, including Abeta42. J Biol Chem 274:18851–18856
82. Loerch PM, Lu T, Dakin KA et al (2008) Evolution of the aging brain transcriptome and synaptic regulation. PLoS One 3:e3329
83. Weidemann A, Konig G, Bunke D et al (1989) Identification, biogenesis, and localization of precursors of Alzheimer's disease A4 amyloid protein. Cell 57:115–126
84. Savage MJ, Trusko SP, Howland DS et al (1998) Turnover of amyloid beta-protein in mouse brain and acute reduction of its level by phorbol ester. J Neurosci 18:1743–1752
85. Lyckman AW, Confaloni AM, Thinakaran G et al (1998) Post-translational processing and turnover kinetics of presynaptically targeted amyloid precursor superfamily proteins in the central nervous system. J Biol Chem 273: 11100–11106
86. Morales-Corraliza J, Mazzella MJ, Berger JD et al (2009) In vivo turnover of tau and APP metabolites in the brains of wild-type and Tg2576 mice: greater stability of sAPP in the beta-amyloid depositing mice. PLoS One 4:e7134
87. Wang R, Sweeney D, Gandy SE, Sisodia SS (1996) The profile of soluble amyloid beta protein in cultured cell media. Detection and quantification of amyloid beta protein and variants by immunoprecipitation-mass spectrometry. J Biol Chem 271:31894–31902
88. Turner PR, O'Connor K, Tate WP, Abraham WC (2003) Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog Neurobiol 70:1–32
89. Selkoe DJ (1994) Normal and abnormal biology of the beta-amyloid precursor protein. Annu Rev Neurosci 17:489–517
90. Pellegrini L, Passer BJ, Tabaton M et al (1999) Alternative, non-secretase processing of Alzheimer's beta-amyloid precursor protein during apoptosis by caspase-6 and -8. J Biol Chem 274:21011–21016
91. McPhie DL, Golde T, Eckman CB et al (2001) beta-Secretase cleavage of the amyloid precursor protein mediates neuronal apoptosis caused by familial Alzheimer's disease mutations. Brain Res Mol Brain Res 97: 103–113
92. Cole SL, Vassar R (2007) The Alzheimer's disease beta-secretase enzyme, BACE1. Mol Neurodegener 2:22
93. Sun X, He G, Song W (2006) BACE2, as a novel APP theta-secretase, is not responsible for the pathogenesis of Alzheimer's disease in Down syndrome. FASEB J 20:1369–1376
94. Slack BE, Ma LK, Seah CC (2001) Constitutive shedding of the amyloid precursor protein ectodomain is up-regulated by tumour necrosis factor-alpha converting enzyme. Biochem J 357:787–794
95. Allinson TM, Parkin ET, Turner AJ, Hooper NM (2003) ADAMs family members as amyloid precursor protein alpha-secretases. J Neurosci Res 74:342–352
96. Yang P, Baker KA, Hagg T (2006) The ADAMs family: coordinators of nervous system development, plasticity and repair. Prog Neurobiol 79:73–94
97. Talamagas AA, Efthimiopoulos S, Tsilibary EC et al (2007) Abeta(1–40)-induced secretion of matrix metalloproteinase-9 results in sAPPalpha release by association with cell surface APP. Neurobiol Dis 28:304–315
98. Deuss M, Reiss K, Hartmann D (2008) Part-time alpha-secretases: the functional biology of ADAM 9, 10 and 17. Curr Alzheimer Res 5:187–201
99. Hartmann D, Tournoy J, Saftig P et al (2001) Implication of APP secretases in notch signaling. J Mol Neurosci 17:171–181
100. Edwards DR, Handsley MM, Pennington CJ (2008) The ADAM metalloproteinases. Mol Aspects Med 29:258–289
101. Arribas J, Bech-Serra JJ, Santiago-Josefat B (2006) ADAMs, cell migration and cancer. Cancer Metastasis Rev 25:57–68

102. Gralle M, Oliveira CL, Guerreiro LH et al (2006) Solution conformation and heparin-induced dimerization of the full-length extracellular domain of the human amyloid precursor protein. J Mol Biol 357:493–508
103. Gralle M, Botelho MG, Wouters FS (2009) Neuroprotective secreted amyloid precursor protein acts by disrupting amyloid precursor protein dimers. J Biol Chem 284: 15016–15025
104. Furukawa K, Sopher BL, Rydel RE et al (1996) Increased activity-regulating and neuroprotective efficacy of alpha-secretase-derived secreted amyloid precursor protein conferred by a C-terminal heparin-binding domain. J Neurochem 67:1882–1896
105. Ghosal K, Vogt DL, Liang M et al (2009) Alzheimer's disease-like pathological features in transgenic mice expressing the APP intracellular domain. Proc Natl Acad Sci U S A 106:18367–18372
106. Copanaki E, Chang S, Vlachos A et al (2010) sAPPalpha antagonizes dendritic degeneration and neuron death triggered by proteasomal stress. Mol Cell Neurosci 44:386–393
107. Sennvik K, Fastbom J, Blomberg M et al (2000) Levels of alpha- and beta-secretase cleaved amyloid precursor protein in the cerebrospinal fluid of Alzheimer's disease patients. Neurosci Lett 278:169–172
108. Hook V, Schechter I, Demuth HU, Hook G (2008) Alternative pathways for production of beta-amyloid peptides of Alzheimer's disease. Biol Chem 389:993–1006
109. Sun X, Wang Y, Qing H et al (2005) Distinct transcriptional regulation and function of the human BACE2 and BACE1 genes. FASEB J 19:739–749
110. Li Q, Sudhof TC (2004) Cleavage of amyloid-beta precursor protein and amyloid-beta precursor-like protein by BACE 1. J Biol Chem 279:10542–10550
111. Lichtenthaler SF, Dominguez DI, Westmeyer GG et al (2003) The cell adhesion protein P-selectin glycoprotein ligand-1 is a substrate for the aspartyl protease BACE1. J Biol Chem 278:48713–48719
112. Scholefield Z, Yates EA, Wayne G et al (2003) Heparan sulfate regulates amyloid precursor protein processing by BACE1, the Alzheimer's beta-secretase. J Cell Biol 163:97–107
113. Nikolaev A, McLaughlin T, O'Leary DD, Tessier-Lavigne M (2009) APP binds DR6 to trigger axon pruning and neuron death via distinct caspases. Nature 457:981–989
114. Vetrivel KS, Zhang YW, Xu H, Thinakaran G (2006) Pathological and physiological functions of presenilins. Mol Neurodegener 1:4
115. Selkoe DJ (1994) Cell biology of the amyloid beta-protein precursor and the mechanism of Alzheimer's disease. Annu Rev Cell Biol 10:373–403
116. Zhao G, Cui MZ, Mao G et al (2005) gamma-Cleavage is dependent on zeta-cleavage during the proteolytic processing of amyloid precursor protein within its transmembrane domain. J Biol Chem 280:37689–37697
117. Qi-Takahara Y, Morishima-Kawashima M, Tanimura Y et al (2005) Longer forms of amyloid beta protein: implications for the mechanism of intramembrane cleavage by gamma-secretase. J Neurosci 25:436–445
118. Okochi M, Tagami S, Yanagida K et al (2013) Gamma-secretase modulators and presenilin 1 mutants act differently on presenilin/gamma-secretase function to cleave Abeta42 and Abeta43. Cell Rep 3:42–51
119. Chen F, Hasegawa H, Schmitt-Ulms G et al (2006) TMP21 is a presenilin complex component that modulates gamma-secretase but not epsilon-secretase activity. Nature 440:1208–1212
120. Lleo A, Waldron E, von Arnim CA et al (2005) Low density lipoprotein receptor-related protein (LRP) interacts with presenilin 1 and is a competitive substrate of the amyloid precursor protein (APP) for gamma-secretase. J Biol Chem 280:27303–27309
121. von Arnim CA, Kinoshita A, Peltan ID et al (2005) The low density lipoprotein receptor-related protein (LRP) is a novel beta-secretase (BACE1) substrate. J Biol Chem 280: 17777–17785
122. Shen J, Kelleher RJ 3rd (2007) The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. Proc Natl Acad Sci U S A 104:403–409
123. Uemura K, Kuzuya A, Shimohama S (2004) Protein trafficking and Alzheimer's disease. Curr Alzheimer Res 1:1–10
124. Boo JH, Sohn JH, Kim JE et al (2008) Rac1 changes the substrate specificity of gamma-secretase between amyloid precursor protein and Notch1. Biochem Biophys Res Commun 372:913–917
125. Liu Y, Zhang YW, Wang X et al (2009) Intracellular trafficking of presenilin 1 is regulated by beta-amyloid precursor protein and phospholipase D1. J Biol Chem 284: 12145–12152
126. Vetrivel KS, Cheng H, Kim SH et al (2005) Spatial segregation of gamma-secretase and substrates in distinct membrane domains. J Biol Chem 280:25892–25900
127. Bitan G, Teplow DB (2004) Rapid photochemical cross-linking–a new tool for studies

of metastable, amyloidogenic protein assemblies. Acc Chem Res 37:357–364

128. Pearson HA, Peers C (2006) Physiological roles for amyloid beta peptides. J Physiol 575:5–10
129. Saito T, Suemoto T, Brouwers N et al (2011) Potent amyloidogenicity and pathogenicity of Abeta43. Nat Neurosci 14:1023–1032
130. Miravalle L, Tokuda T, Chiarle R et al (2000) Substitutions at codon 22 of Alzheimer's abeta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. J Biol Chem 275: 27110–27116
131. Murakami K, Irie K, Morimoto A et al (2002) Synthesis, aggregation, neurotoxicity, and secondary structure of various A beta 1–42 mutants of familial Alzheimer's disease at positions 21–23. Biochem Biophys Res Commun 294:5–10
132. Kakio A, Yano Y, Takai D et al (2004) Interaction between amyloid beta-protein aggregates and membranes. J Pept Sci 10: 612–621
133. Klug GM, Losic D, Subasinghe SS et al (2003) Beta-amyloid protein oligomers induced by metal ions and acid pH are distinct from those generated by slow spontaneous ageing at neutral pH. Eur J Biochem 270:4282–4293
134. Zhao JH, Liu HL, Liu YF et al (2009) Molecular dynamics simulations to investigate the aggregation behaviors of the Abeta(17–42) oligomers. J Biomol Struct Dyn 26:481–490
135. Zheng J, Jang H, Ma B et al (2007) Modeling the Alzheimer Abeta17-42 fibril architecture: tight intermolecular sheet-sheet association and intramolecular hydrated cavities. Biophys J 93:3046–3057
136. Miller Y, Ma B, Nussinov R (2009) Polymorphism of Alzheimer's Abeta17-42 (p3) oligomers: the importance of the turn location and its conformation. Biophys J 97:1168–1177
137. Thal DR, Sassin I, Schultz C et al (1999) Fleecy amyloid deposits in the internal layers of the human entorhinal cortex are comprised of N-terminal truncated fragments of Abeta. J Neuropathol Exp Neurol 58:210–216
138. Liu R, McAllister C, Lyubchenko Y, Sierks MR (2004) Residues 17–20 and 30–35 of beta-amyloid play critical roles in aggregation. J Neurosci Res 75:162–171
139. Wei W, Norton DD, Wang X, Kusiak JW (2002) Abeta 17–42 in Alzheimer's disease activates JNK and caspase-8 leading to neuronal apoptosis. Brain 125:2036–2043
140. White AR, Maher F, Brazier MW et al (2003) Diverse fibrillar peptides directly bind the Alzheimer's amyloid precursor protein and amyloid precursor-like protein 2 resulting in cellular accumulation. Brain Res 966: 231–244
141. Beckman M, Holsinger RM, Small DH (2006) Heparin activates beta-secretase (BACE1) of Alzheimer's disease and increases autocatalysis of the enzyme. Biochemistry 45:6703–6714
142. Bame KJ, Danda J, Hassall A, Tumova S (1997) Abeta(1–40) prevents heparanase-catalyzed degradation of heparan sulfate glycosaminoglycans and proteoglycans in vitro. A role for heparan sulfate proteoglycan turnover in Alzheimer's disease. J Biol Chem 272:17005–17011
143. Klajnert B, Cortijo-Arellano M, Bryszewska M, Cladera J (2006) Influence of heparin and dendrimers on the aggregation of two amyloid peptides related to Alzheimer's and prion diseases. Biochem Biophys Res Commun 339:577–582
144. Bergamaschini L, Donarini C, Rossi E et al (2002) Heparin attenuates cytotoxic and inflammatory activity of Alzheimer amyloid-beta in vitro. Neurobiol Aging 23:531–536
145. Winkler K, Scharnagl H, Tisljar U et al (1999) Competition of Abeta amyloid peptide and apolipoprotein E for receptor-mediated endocytosis. J Lipid Res 40:447–455
146. Benilova I, Karran E, De Strooper B (2012) The toxic Abeta oligomer and Alzheimer's disease: an emperor in need of clothes. Nat Neurosci 15:349–357
147. Svedruzic ZM, Popovic K, Smoljan I, Sendula-Jengic V (2012) Modulation of gamma-secretase activity by multiple enzyme-substrate interactions: implications in pathogenesis of Alzheimer's disease. PLoS One 7:e32293
148. Yan Y, Wang C (2006) Abeta42 is more rigid than Abeta40 at the C terminus: implications for Abeta aggregation and toxicity. J Mol Biol 364:853–862
149. Lee DH, Wang HY (2003) Differential physiologic responses of alpha7 nicotinic acetylcholine receptors to beta-amyloid1-40 and beta-amyloid1-42. J Neurobiol 55: 25–30
150. Zou K, Kim D, Kakio A et al (2003) Amyloid beta-protein (Abeta)1-40 protects neurons from damage induced by Abeta1-42 in culture and in rat brain. J Neurochem 87:609–619
151. Parameshwaran K, Sims C, Kanju P et al (2007) Amyloid beta-peptide Abeta(1–42) but

not Abeta(1–40) attenuates synaptic AMPA receptor function. Synapse 61:367–374

152. Chavez-Gutierrez L, Bammens L, Benilova I et al (2012) The mechanism of gamma-secretase dysfunction in familial Alzheimer disease. EMBO J 31:2261–2274
153. Svedruzic ZM, Popovic K, Sendula-Jengic V (2013) Modulators of gamma-secretase activity can facilitate the toxic side-effects and pathogenesis of Alzheimer's disease. PLoS One 8:e50759
154. Orchard S, Salwinski L, Kerrien S et al (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat Biotechnnol 25: 894–898
155. Taylor CF, Paton NW, Lilley KS et al (2007) The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 25:887–893
156. Kasabov N (2010) To spike or not to spike: a probabilistic spiking neuron model. Neural Netw 23:16–19
157. Kasabov N, Schliebs R, Kojima H (2011) Probabilistic computational neurogenetic modeling: from cognitive systems to Alzheimer's disease. IEEE Trans Auton Ment Dev 3:300–311. doi:10.1109/TAMD.2011.2159839
158. Kasabov N, Benuskova L, Wysoski SG (2005) Biologically plausible computational neurogenetic models: modeling the interaction between genes, neurons and neural networks. J Comput Theor Nanosci 2:569–573
159. Habeck C, Foster NL, Perneczky R et al (2008) Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. Neuroimage 40:1503–1515
160. Haense C, Buerger K, Kalbe E et al (2008) CSF total and phosphorylated tau protein, regional glucose metabolism and dementia severity in Alzheimer's disease. Eur J Neurol 15:1155–1162
161. Desikan RS, Cabral HJ, Hess CP et al (2009) Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132:2048–2057
162. Wang H, Nie F, Huang H et al (2011) Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. Med Image Comput Comput Assist Interv 14:115–123
163. Cui Y, Liu B, Luo S et al (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS One 6:e21896
164. Hinrichs C, Singh V, Xu G, Johnson SC (2011) Predictive markers for AD in a multimodality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55:574–589
165. Zhang D, Wang Y, Zhou L et al (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55:856–867
166. Gray KR, Aljabar P, Heckemann RA et al (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. Neuroimage 65:167–175
167. Van Essen DC, Ugurbil K (2012) The future of the human connectome. Neuroimage 62:1299–1310
168. Castellani RJ, Perry G (2012) Pathogenesis and disease-modifying therapy in Alzheimer's disease: the flat line of progress. Arch Med Res 43:694–698
169. Carrillo MC, Rowe CC, Szoeke C et al (2013) Research and standardization in Alzheimer's trials: reaching international consensus. Alzheimers Dement 9:160–168
170. Stern Y (2012) Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 11:1006–1012

Chapter 4

Effects of Mild and Severe Oxidative Stress on BACE1 Expression and APP Amyloidogenic Processing

Jiangli Tan, Qiao-Xin Li, and Genevieve Evin

Abstract

This chapter describes methods for establishing oxidative stress conditions that do not induce cell death in a neuronal cell culture model. We termed these conditions "mild oxidative stress," as opposed to "severe oxidative stress," which results in significant cell loss. Mild oxidative stress resembles more closely what happens in the aging brain than severe oxidative stress. The protocols we have delineated include the preparation and maintenance of mouse primary cortical cultures, the induction of oxidative stress by treatment with hydrogen peroxide, the assessment of cell viability by the 3-(4,5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) assay, the measurement of free radical production by the 2′,7′-dichlorofluorescein (DCF) assay, and western blot analysis of the amyloid precursor protein (APP) and β-site APP cleaving enzyme, BACE1, two key proteins associated with Alzheimer's disease pathology and oxidative stress.

Key words Oxidative stress, BACE1, Alzheimer's disease, Amyloid precursor protein, Free radicals, Primary cortical cultures, Hydrogen peroxide, MTT assay, DCF assay, Western blotting

1 Introduction

The β-site of APP cleaving enzyme 1 (BACE1) is directly implicated in the pathology of Alzheimer's disease, as its cleavage of the amyloid precursor protein (APP) represents the first step in Aβ amyloid generation [1–5]. Numerous studies have shown that BACE1 expression increases in response to oxidative stress [6–18], which is a process associated with Alzheimer's disease [19–22]. We have recently demonstrated, using murine primary cortical cultures, that changes in BACE1 expression depend on the extent of oxidative stress, and that BACE1 protein levels only increase under severe oxidative stress conditions, which result in cell death [23]. We have also shown that mild oxidative stress, although having no effect on BACE1 cellular levels, caused the subcellular redistribution of the enzyme to promote cleavage of APP, and thereby initiate APP amyloidogenic processing [23].

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_4,

The cellular model of neuronal oxidative stress described in this chapter involves treatment of mouse primary cortical cells with hydrogen peroxide, as the source of reactive oxygen species. The methodology section begins with the preparation of primary cortical cultures from mouse embryos, following with the optimization of hydrogen peroxide concentration to induce mild oxidative stress, and assays to evaluate cell viability and free radical production, and finally analysis of changes in BACE1 protein expression and APP processing.

The method chosen to evaluate the cell viability is the MTT assay, which measures mitochondrial reductase enzyme as a marker of healthy cellular metabolic activity [24]. This enzyme catalyzes conversion of the soluble, cell-permeable, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to its formazan derivative, a purple-colored insoluble salt. Formazan concentration is measured by spectrophotometry.

Free radical production is determined by the DCF assay [25, 26]. This involves 2′,7′-dichlorofluorescin diacetate (DCFH-DA), a stable, non-polar and cell-penetrant compound that is hydrolyzed by intracellular esterases to produce dichlorofluorescin (DCFH), which is retained within the cells. DCFH becomes oxidized in the presence of intracellular free radicals to form the fluorescent compound, 2′,7′-dichlorofluorescein (DCF), which can be measured by fluorescence spectroscopy.

This chapter also describes the applications of SDS-PAGE and immunoblotting to the detection of BACE1 and of APP and its C-terminal fragments, β-CTF (or C99) and α-CTF (or C83).

2 Materials

2.1 Preparation of Murine Primary Cortical Culture Components

2.1.1 Animals

- Animal ethics approval must be obtained prior to the study and animals have to be maintained in an ethically approved facility.
- Pregnant C57BL/6 mouse (embryonic day 14) are used.

2.1.2 Equipment

- Sterile biological safety cabinet.
- Humidified 37 °C, 5 % CO_2 incubator.
- Shaking 37 °C water bath.
- Liquid aspirator (optional).
- Dissection tools: scissors and fine forceps (one straight and one curved).
- Dissection microscope with light source.
- Tissue culture dishes (3 and 10 cm) and tissue culture plates.
- 70 % ethanol.

2.1.3 Buffers and Reagents

- B-27® Serum-Free Supplement 50× liquid (e.g. Invitrogen).
- Krebs/HEPES buffer (pH 7.4): 124.1 mM NaCl, 5.4 mM KCl, 1.0 mM NaH_2PO_4, 14.4 mM D-Glucose, 24.9 mM HEPES, and 0.0027 mM Phenol red. Store at 4 °C.
- Buffer 1: Krebs/HEPES buffer supplemented with 0.3 % (w/v) BSA and 0.031 % (w/v) $MgSO_4$.
- Buffer 2: Krebs/HEPES buffer supplemented with 2.5 mg/mL trypsin (≥7,500 Units/mg). Prepare when needed (*see* **Note 1**).
- Buffer 3: Krebs/HEPES buffer supplemented with 0.8 mg/mL DNase I (>2,000 Units/mg) and 2.6 mg/mL soybean trypsin inhibitor. Prepare when needed (*see* **Note 2**).
- Plating medium: Minimum Essential Medium (MEM), cell culture medium supplemented with 2 mM L-Glutamine, 0.22 % (v/v) bicarbonate ($NaHCO_3$), 0.01 mg/mL Gentamicin, 10 % (v/v) fetal calf serum and 5 % (v/v) horse serum. Store at 4 °C.
- Culture medium: Neurobasal medium supplemented with 0.2 mM L-Glutamine, 0.01 mg/mL Gentamicin and 1× B27® Serum-Free Supplement (*see* **Note 3**).
- 100× stock poly-D-Lysine (0.5 mg/mL) prepared by dissolving 25 mg of poly-D-Lysine (e.g. Sigma) in 50 mL filter-sterilized water. Can be stored as aliquots, at –20 °C.

2.2 Hydrogen Peroxide Treatment Components

- 30 % hydrogen peroxide solution.
- B-27® Serum-Free Supplement Minus Anti-oxidants 50× liquid (e.g. Invitrogen).
- Treatment medium: Neurobasal medium supplemented with 0.2 mM L-Glutamine, 0.01 mg/mL Gentamicin and B27® Serum-Free Supplement Minus Antioxidant supplement.
- Phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 6.5 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4.
- Light-protection (dark) tubes.

2.3 MTT Assay Components

- 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT).
- Dimethyl sulfoxide (DMSO), tissue culture grade.
- Phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 6.5 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4.
- Clear plastic 96-well plate.
- Absorbance microplate reader (spectrophotometer).

2.4 DCF Assay Components

- 2′,7′-dichlorofluorescin diacetate (DCFH-DA).
- DCF assay cell lysis buffer: 0.1 M Tris–HCl pH 7.5 supplemented with protease inhibitor cocktail (Roche Complete) and 1 % Triton X-100.

- 96-well microplate for fluorescence assays (i.e. Perkin Elmer Optiplate).
- Fluorescence microplate reader.

2.5 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Immunoblotting Components

2.5.1 Equipment

- Cell scraper to harvest the cells.
- 1.5 mL microfuge tubes.
- Centrifuge capable of spinning at 14,000 × *g* (either refrigerated at 4 °C, or placed in a cold room).
- Mini gel electrophoresis system (i.e. BioRad Mini Protean, or Life Technologies XCell Sure-Lock Mini Cell).
- Heating block for microtubes.
- BioRad Precision Plus or other pre-stained protein molecular weight markers.
- 8.5 % Tris-Glycine polyacrylamide gels (for BACE1 analysis).
- 10–20 % Tris-Tricine gels (for APP CTF analysis).
- Western transfer apparatus (equipped with transfer cassettes and sponges if using wet system).
- Nitrocellulose membrane (BioRad).
- Filter paper (Whatman).
- Small containers (for incubating membranes with blocking and antibody solutions).
- Rocking or shaking platform.
- Imaging system for enhanced chemiluminescence (ECL) detection.

2.5.2 Buffers and Reagents

- Lysis buffer (RIPA): 50 mM Tris–HCl pH 7.4, 150 mM sodium chloride, 1 % Nonidet-p40, 0.5 % sodium deoxycholate, 0.1 % sodium dodecyl sulfate (SDS). Just before use, add phosphatase inhibitor cocktail (i.e. Roche PhosSTOP), protease inhibitor cocktail (i.e. Roche Complete), and DNAse I (Roche; 0.05 mg/mL) and keep buffer chilled on ice.
- 4x Laemmli sample buffer: 62.5 mM Tris–HCl pH 6.8, 25 % glycerol, 2 % SDS, 0.01 % bromophenol blue, plus 20 % 2-mercaptoethanol (added immediately before use).
- Glycine electrophoresis buffer: 25 mM Tris, 192 mM Glycine, 0.1 % sodium dodecyl sulfate (SDS), pH 8.3.
- Tricine electrophoresis buffer: 100 mM Tris–HCl, 100 mM Tricine, 0.1 % sodium dodecyl sulfate (SDS), pH 8.2.
- Western transfer buffer: 25 mM Tris, 192 mM Glycine, 20 % methanol. Store at 4 °C.
- Tris-buffered saline containing Tween 20 (TBST): 10 mM Tris–HCl, 150 mM NaCl, pH 8, containing 0.05 % Tween 20. Store at 4 °C.

- Skim milk blocking buffer (5 % skim milk): Dissolve skim milk in TBST at a ratio of 5 % (w/v). Store at 4 °C for no more than 3 days (or freeze for longer storage).
- Bovine serum albumin (BSA) blocking buffer (5 % BSA): To be prepared in advance. Dissolve bovine serum albumin (BSA; Fraction V), at a ratio of 5 % (w/v) in TBST, by gently mixing with rocking (avoid energetic mixing methods that will cause frothing). Store at 4 °C for no more than 3 days (or freeze for longer storage).
- Casein blocking buffer (0.5 % casein): To be prepared in advance. Heat 0.5 g of casein in 10 mL of 0.1 N sodium hydroxide, until a clear solution is obtained. Cool to room temperature. Adjust pH to 7.4 with 1 N hydrochloric acid (HCl). Add PBS to make 100 mL final volume. Store at 4 °C for no more than 3 days (or freeze for longer storage).
- Primary antibodies: BACE1 CT antibody (D10E5, Cell Signaling cat #5606), APP CT 369 (directed to the APP cytoplasmic domain; kindly provided by Prof Sam Gandy, Mount Sinai Hospital, New York), rabbit anti-actin antibody (Sigma).
- Secondary horseradish peroxidase (HRP)-conjugated antibodies: Immunopure goat anti-rabbit IgG (#31460, PIERCE Thermo Scientific).
- Enhanced chemiluminescence (ECL) reagent (i.e. Pierce West Dura, Thermo Scientific).

3 Methods

3.1 Preparation of Murine Primary Cortical Cells

3.1.1 Preliminary Steps

The day before the cell preparation, coat tissue culture plates with poly-D-lysine. In a sterile biological cabinet, make a working concentration by diluting 100 times the 0.5 mg/mL stock in sterile water. Pipette poly-D-Lysine solution to cover the bottom of the wells. Wrap the plates in aluminium foil (to protect from UV light) and leave overnight in the tissue culture cabinet. The next day, aspirate the poly-D-Lysine solution, and let the wells dry before plating the cells.

On the day of the procedure, pre-warm the plating medium in a 37 °C water bath. The Krebs/HEPES buffer is kept refrigerated until needed. Thaw aliquots of trypsin and DNase I/Soybean Trypsin Inhibitor solution (*see* **Notes 1** and **2**) on ice or in the refrigerator, and prepare Buffer 2. Place it in the 37 °C water bath.

3.1.2 Procedure for Neuronal Cell Preparation

1. Euthanize a pregnant C57BL/6 mouse at embryonic day E14 by carbon dioxide asphyxiation.
2. Spray scissors, forceps and mouse with 70 % ethanol.
3. Make incision to the abdomen to expose the uterus and spray uterus with 70 % ethanol.

4. Excise uterus and rinse in a 3 cm dish containing cold buffer 1.
5. Fill the base of a 10 cm dish with Buffer 1 and immerse the uterus in the buffer, proceed to liberate the embryonic sac by peeling open the uterus with fine forceps.
6. Gently make an incision in the embryonic sac to expose and extract mouse embryos.
7. Once all embryos have been extracted, transfer them to the lid of a 10 cm dish containing cold Buffer 1 (*see* **Note 3**).
8. Spray fine forceps with 70 % ethanol to remove any animal hair.
9. Under a dissection microscope with a light source, firstly make an incision at the posterior section of the head with fine forceps. Then gently peel off the scalp and skull layers in the direction towards the anterior part of the head to minimize damage to the cortex (*see* **Note 4**).
10. Gently separate the brain from the head at the margin between the midbrain and pons.
11. Remove the olfactory bulbs from the hemispheres.
12. Separate the cerebral cortex from the midbrain and cerebellum.
13. Separate the two hemispheres of the isolated cerebral cortex and remove the meninges with the fine forceps.
14. Transfer isolated cortices to a 3 cm cell culture dish and homogenize with a sterile scalpel blade.

    For all the following steps, buffers and media must be warmed to 37 °C and procedures must be performed in a sterile biological safety cabinet.
15. Mix the homogenized cortical tissue with 1 mL of Buffer 2 and transfer into a 50 mL tube containing 14 mL of the same buffer.
16. Incubate in a 37 °C water bath for 20 min with shaking.
17. During the incubation, prepare 11 mL of Buffer 3 (*see* **Note 2**).
18. After incubation, remove the cortical material from the water bath and add to it 10 mL of the Buffer 3 prepared at **step 17**. Mix by inversion until the viscous precipitate is fragmented into small particles.
19. Centrifuge at ~250 × *g* for 3 min at room temperature and then aspirate the supernatant.
20. Carefully triturate the resultant cell pellet to dissociate the cells with the remaining 1 mL cortical prep Buffer 3 prepared at **step 17**. Triturate slowly no more than 30 times (*see* **Note 5**).
21. Transfer the 1 mL cell suspension into another tube containing 9 mL buffer 3. Avoid the presence of undissolved materials (*see* **Note 6**).

22. Centrifuge at ~250 × *g* for 3 min at room temperature and aspirate the supernatant.
23. Resuspend the cells in 1 mL plating medium by gently triturating three times.
24. Adjust cell suspension volume to 10 mL with plating medium and do cell counting (*see* **Note 7**).
25. Adjust the cell concentration to 8×10^5 cells per mL with warm plating medium and seed 1 mL of this suspension into wells of a 12-well poly-D-Lysine-coated tissue culture plate.
26. Incubate the plate in a humidified 37 °C, 5 % CO_2 incubator for no more than 4 h (*see* **Note 8**).
27. Observe cells under a microscope. At this stage, the neurons will appear as bright round cells attached to the base of the wells; some neurons will have begun to establish processes. Warm up culture medium at 37 °C. Aspirate the plating medium and gently replace with 1 mL of culture medium. Culture cortical cells in a humidified 37 °C, 5 % CO_2 incubator for 6 days.
28. Treatment of primary cortical cells can commence on the sixth day in culture. At this stage, wells should be evenly confluent with a monolayer of cells; the neuronal cell bodies should be clearly visible with extensive branching of axons and dendrites. There should not be uneven cortical cell clusters throughout the wells.

3.2 Induction of Oxidative Stress with Hydrogen Peroxide

3.2.1 Preliminary Steps

Pre-warm the culture medium at 37 °C.

3.2.2 Treatment Procedure

1. Before starting treatment, observe primary cortical cells under the microscope to ensure they are healthy.
2. Prepare hydrogen peroxide dilutions spanning a broad concentration range (i.e. from 1 to 100 μM) in warmed culture medium, in light-protection tubes. Treatment volume per well should be the same as the volume for culturing. Ensure the total volume prepared is sufficient for replicates (*see* **Note 9**).
3. Tilt the tissue culture dish and aspirate conditioned medium.
4. Gently add 1 mL of treatment medium into control wells and 1 mL of medium containing the hydrogen peroxide solutions to the remaining wells (when exchanging media, tilt the plate and gently pipette the media into the wells to avoid lifting the cells). Return the cells to the incubator and allow treatment to

proceed for 6 h (or time predetermined), before lysis for either assessment of cell viability, or measurement of intracellular free radical levels, or for western blotting, as described below.

3.3 Assessing the Cell Viability by the MTT Assay

1. Prepare MTT 100× stock solution at a concentration of 5 mg/mL in PBS. This can be prepared in advance and stored as 1 mL aliquots at −20 °C (*see* **Note 10**). Frozen aliquots should be thawed in a 37 °C water bath before use.
2. To each well, add 10 μL of MTT stock solution into the 1 mL conditioned medium 30 min before the end of the treatment. Return cells to the incubator (*see* **Note 11**)
3. At the end of the treatment period, aspirate and discard conditioned medium.
4. Lyse cells by adding 400–500 μL of DMSO and triturating a few times with a pipette tip. Leave at room temperature until the formazan precipitate is completely dissolved (*see* **Note 12**).
5. Transfer 100–150 μL triplicate aliquots of the DMSO solution into a clear 96-well plate to measure absorbance by spectrophotometry at a wavelength of 540 nm. A decrease in formazan levels indicates reduced mitochondrial function, thus reduced cell viability.
6. Assess the effect of hydrogen peroxide treatment on viability by calculating absorbance readings of the treated cells relative to untreated controls (for an example, refer to Fig. 1).

3.4 Assessing Intracellular Free Radical Generation by the DCF Assay

1. After hydrogen peroxide treatment, aspirate medium and wash the cells with 1 mL PBS.
2. Aspirate PBS and carefully add 1 mL of culture medium supplemented with 50 μM DCFH-DA. Place plate in 37 °C, 5 % CO_2 incubator for 40 min (*see* **Note 13**).
3. Wash cells with PBS (*see* **Note 14**).
4. Disrupt cells by adding 40–50 μL of DCF assay cell lysis buffer to each well.
5. Transfer lysates into a 96-well microplate for fluorescence assay, and measure DCF fluorescence at excitation and emission wavelengths of 485 nm and 535 nm, respectively (*see* **Note 15** and Fig. 2).

3.5 SDS-PAGE and Immunoblotting for BACE1 and APP/APP-CTFs

1. After treatment, aspirate conditioned media and transfer them to microfuge tubes. Store the media at −20 °C for future analysis.
2. Lyse the cells by pipetting 40–50 μL of cold RIPA lysis buffer into each well. Disrupt the cells using a cell scraper. Then triturate several times with a pipette tip and transfer lysate to 1.5 mL microfuge tubes (*see* **Note 16**).

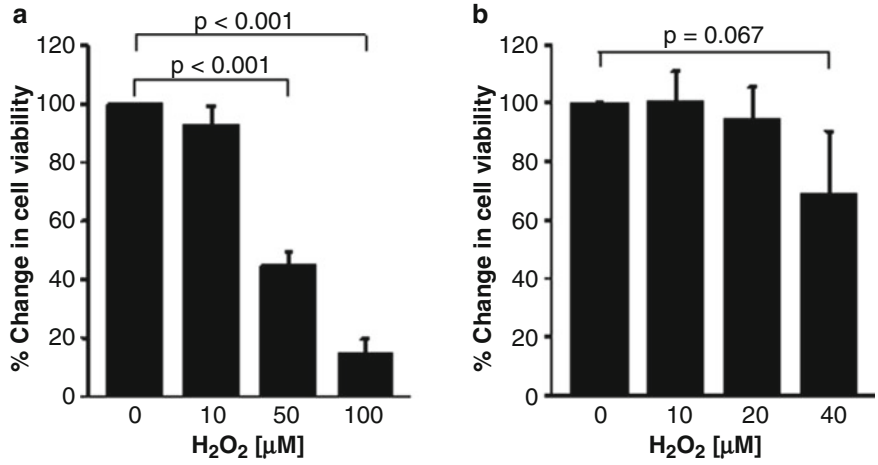


Fig. 1 Use of the MTT Assay for determining non-lethal concentrations of hydrogen peroxide in the treatment of mouse primary cortical cells. Cells at day 6 in culture were treated for 6 h with indicated hydrogen peroxide (H_2O_2) concentrations. (**a**) Treatment with hydrogen peroxide at up to 100 μM ($n=3$). Note that the cell viability decreased significantly with 50 and 100 μM hydrogen peroxide concentrations. (**b**) Treatment with a lower range of hydrogen peroxide concentrations, up to 40 μM ($n=6$). The results indicate no significant loss in cell viability within 10–40 μM hydrogen peroxide concentrations. Data reproduced from previous work [23], with permission under the Creative Commons Attribution License, Open Access License (no permission required; http://www.plosone.org/static/license)

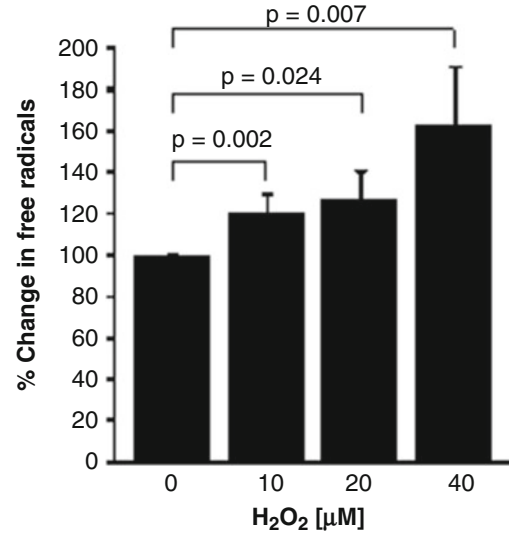


Fig. 2 Use of the DCF assay to demonstrate free radical production in cortical cells treated with low hydrogen peroxide concentrations. Mouse primary cortical cell were treated for 6 h with hydrogen peroxide concentrations, as indicated, and cell lysates were subjected to the DCF assay. The results indicate that levels of intracellular free radicals were significantly increased after treatment with 10–40 μM hydrogen peroxide. Data reproduced from previous work [23], with permission under the Creative Commons Attribution License, Open Access License (no permission required; http://www.plosone.org/static/license)

Table 1
Parameters for SDS-PAGE of BACE1 and APP/APP-CTFs

Protein	Sample heating	Gel system	Electrophoresis	Western transfer
BACE1	55 °C, 10 min	8.5 % Glycine buffer	40 mA, constant 30–40 min	370 mA, constant 60 min
APP/APP CTFs	95 °C, 5 min	10–20 % Tricine buffer	126 V, constant 100 min	26 V, constant 90 min

These parameters are based on electrophoresis using the Biorad Mini Protean III (BACE1) and Novex X Cell II SureLock Mini-Cell (APP-CTFs). Buffer compositions are given in Materials (Subheading 2.5)

3. Centrifuge lysates at 14,000 × *g* for 5 min, at 4 °C.
4. Collect supernatants and transfer to new 1.5 mL microfuge tubes (*see* **Note 17**).
5. Determine the protein concentration in the lysates (e.g. using the Bicinchoninic acid (BCA) assay kit from Pierce, Thermo Fisher or equivalent).
6. Prepare 20 μg protein samples in 1.5 mL microfuge tubes and add Laemmli sample buffer (*see* Subheading 2.5). Place the tubes in a heating block to denature the samples. Heating temperatures and times for specific proteins are indicated in Table 1.
7. Centrifuge the tubes briefly to collect all liquid at the bottom of the tubes.
8. Load samples and pre-stained molecular weight markers on polyacrylamide gels, and carry out electrophoresis (for conditions refer to Table 1).
9. Perform western transfer of proteins onto nitrocellulose membrane (*see* Table 1).
10. After transfer, rinse the membrane with deionized water, and place it into a small container with a minimum volume of blocking buffer to cover the surface of the membrane. Incubate blot at room temperature for 1 h on a rocking platform (for blocking buffer selection, refer to Table 2)
11. Discard blocking buffer and wash membrane three times for 10-min with a minimum volume of cold TBST, with shaking on a rocking platform.
12. Discard washing buffer, and add primary antibody dilution (*see* Table 2). Incubate at 4 °C overnight (or for 2–3 h at room temperature), with shaking on a rocking platform (*see* **Note 18**).
13. Discard antibody dilution (or transfer to a tube for recycling) and wash the membrane three times with cold TBST as in **step 11**.

Table 2
Conditions for immunoblotting of BACE1 and APP/APP-CTFs

Protein	Blocking buffer	Primary antibody	Secondary antibody
BACE1	5 % skim milk	D10E5 (1/1,000 in 5 % BSA)	Anti-rabbit-HRP (1/2,000 in 5 % skim milk)
APP/ APP-CTFs	5 % skim milk	Ab369 (1/2,000 in 0.5 % casein)	Anti-rabbit-HRP (1/5,000 in TBST)

For preparation of blocking buffers and TBST refer to Subheading 2.5

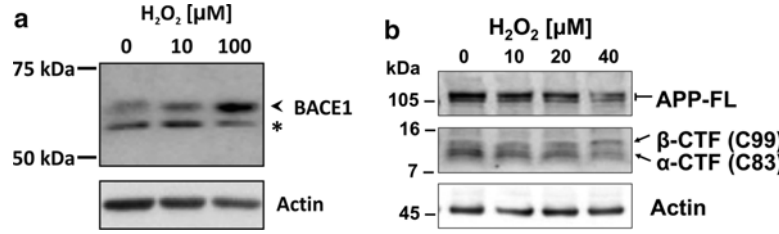


Fig. 3 Immunoblot analysis of BACE1 and APP/APP CTFs in mouse primary cortical cultures treated with hydrogen peroxide. (**a**) Immunoblot of BACE1. 20 μg of protein were separated by SDS-PAGE on an 8.5 % polyacrylamide gel and transferred to nitrocellulose membrane. The blot was probed with BACE1 C-terminal antibody, D10E5. BACE1 was detected as a 70 kDa signal. This signal was increased in cells treated with 100 μM hydrogen peroxide compared to untreated cells. The *asterisk* indicates a non-specific signal. (**b**) Immunoblot of APP and its C-terminal fragments. 20 μg of protein were separated by SDS-PAGE on a 10–20 % Tris-Tricine gel and transferred to nitrocellulose. The membrane was developed with APP C-terminal 369 antibody. APP was detected as a 110 kDa doublet band. β- and α-CTFs were separated as ~12 and 10 kDa bands respectively. The 12 kDa signal increased relatively to the 10 kDa signal in the cells treated with 40 μM hydrogen peroxide

14. Incubate membrane with secondary antibody solution for 1 h at room temperature, on a rocking platform.
15. Wash membrane three times with TBST as in **step 11**.
16. Prepare ECL developing solution according to the manufacturer's recommendations.
17. Delicately place the membrane into a container layered with ECL solution (*see* **Note 19**).
18. Drip excess reagent and place the membrane in an imaging digital camera instrument suitable for chemiluminescence detection. Typical images of blots are shown in Fig. 3.

4 Notes

1. Aliquots of a 20× trypsin solution can be prepared in advance and stored at −20 °C. Dissolve 25 mg of trypsin in 10 mL of Krebs/HEPES buffer; sterilize by passage through a 0.2 μm syringe filter, and distribute into 750 μL aliquots for storage. To prepare Buffer 2, add one thawed trypsin aliquot to 15 mL Krebs/HEPES buffer.
2. Aliquots of DNase I/Soybean Trypsin Inhibitor solution can be prepared and stored at −20 °C. Weigh out 8 mg DNase I and 26 mg Soybean Trypsin Inhibitor in a tube and dissolve in 10 mL Krebs/HEPES buffer; sterilize by passage through a 0.2 μm syringe filter and distribute into 500 μL aliquots for storage at −20 °C. To prepare Buffer 3, add one thawed aliquot of DNase I/Soybean Trypsin Inhibitor solution to 10.5 mL Krebs/HEPES buffer.
3. The volume of Buffer 1 to be used (**step 7**) should be just enough to submerge the embryos. A too large buffer volume will make dissection under a microscope harder as handling the embryos will be challenging, and make microscope observation less clear due to increased light diffraction.
4. When isolating the cortex (**step 9**), the embryo's head may be held with the non-dominant hand using the curved forceps while peeling with the dominant hand using the straight forceps.
5. When dissociating the cell pellet (**step 20**), cell loss is unavoidable due to the fragility of the primary neurons. It is thus advisable to use a pipette tip with a blunt edge to minimize cell rupture during the trituration. For this reason, it is important to keep record of the type of pipette tip used at this step of the procedure. If a low cell yield is obtained, changing to another manufacturer's tip may be trialled, assuming that the other steps of the protocol have been correctly executed. Once a suitable tip brand has been selected, this should always be used for future preparations. Trituration must not exceed 30 strokes to minimize cell loss. Generally, $6–7 \times 10^6$ cells can be obtained from each E14 embryo.
6. Undissolved materials may remain in the 1 mL cell suspension after trituration (**step 21**). This corresponds to pieces of undigested connective tissue and cellular debris. It is essential to avoid carrying this material throughout the next steps to ensure a clean preparation. For this, it may be recommended to filter the cell suspension through a sterile mesh while transferring to the tube containing the 9 mL of buffer 3. Alternatively, the cell suspension may be gently and slowly run down the walls of the tube to retain the undissolved materials.

This can be repeated two to three times on different areas of the tube to further clean up the cell suspension before transfer to a new tube containing Krebs/HEPES buffer 3. This latter approach is often practised in our laboratory.

7. The primary neuron cell density for seeding (**step 24**) will have to be optimized. To foster the viability of the neurons it is crucial that the cell density is not too low. For example, in our laboratory, when using a NUNC 12-well plate, we often seed 8×10^5 cells per well; this density allows a confluent monolayer of cells to be established by the time of treatment (day 6). When seeding cells, one should take care that the cells are evenly distributed to facilitate formation of a monolayer.
8. It is advisable not to maintain the cortical cells in plating medium for more than 4 h (**step 26**), as the serum contents allow the growth and proliferation of glial cells. If the goal is to establish a predominantly neuronal population, then the plating medium should be removed after 4 h (NB: Removing the plating medium sooner may result in losing cortical cells, as these may not have enough time to adhere properly). The serum-free B27 supplement in the Neurobasal medium (**step 26**) promotes neuronal cell growth while minimizing glial proliferation [27, 28],
9. It is recommended to prepare hydrogen peroxide solutions in light-protection (dark) tubes, as hydrogen peroxide is light sensitive.
10. MTT powder may be difficult to dissolve, thus it should be weighed in a screw-capped tube that will allow vigorous vortexing.
11. Although the protocol states that the MTT solution should be added 30 min before the end of the treatment, the experimenter may determine empirically the optimal time to commence MTT treatment so as to avoid under, or over production of formazan. Optimizing the duration of MTT treatment is essential for distinguishing between untreated and treated cell populations, as a too short treatment may only cause a subtle difference in color formation.
12. The volume of DMSO to be used for cell lysis will depend on the diameter of the wells and the intensity of color formation. A volume that can cover the cells completely and easily dissolve the formazan is recommended (for example, for cells seeded at a density of 8×10^5 per well, in a 12-well plate, adding 400–500 μL of DMSO per well is usually sufficient). If the formazan does not dissolve completely, triturating or vortexing several times, followed by resting at room temperature for a few minutes before mixing again may be tried. If formazan granules still persist, allow longer incubation at room temperature

(or add a small volume of DMSO) and mix again. Ensure that the formazan is completely dissolved before spectrophotometry measurement.

13. The final concentration of DCFH-DA (Subheading 3.4, **step 2**) may need to be optimized to allow discriminating between untreated and treated cells. Free radicals are always present within cells, thus using too high a concentration of DCFH-DA will result in saturated signals in both control and treated cells, and make it impossible to assess the increase in free radicals caused by the oxidant. The DCFH-DA concentration of 50 μM suggested here was optimized for our experiments with hydrogen peroxide treatment of murine primary cortical cells. Conditions should be optimized according to the free-radical inducing chemical, the source of the primary cells (embryonic day of harvesting, and days in culture), and the cell number. For example, 5 μM DCFH-DA was determined to be more appropriate than 50 μM for mouse primary cortical cells treated with buthionine sulfoximine.
14. The PBS wash step (**step 3**) is required to remove extracellular DCFH-DA; if needed, this washing step can be repeated before cell lysis. This will ensure that false positive DCF signal is minimized, as unwashed DCFH-DA present will be exposed to intracellular esterases and oxidants when the cells are lysed. This washing step should be done gently to prevent cell loss.
15. The volume of lysis buffer is given for a 12-well plate, and must be adjusted according to the size of the wells (**step 4**). Several lysate dilutions should be tried so as to avoid signal saturation.
16. If a more concentrated lysate is needed, a smaller volume of lysis buffer can be used. It is advisable to perform lysis under cold conditions. Therefore the lysis buffer must be cold and the plate placed on ice while lysing cells; the lysates should be placed on ice immediately after transferring into microfuge tubes.
17. Lysates can be stored at −20 °C as two to three aliquots so as to avoid freezing/thawing cycles that will lead to protein degradation.
18. Antibody dilutions in Table 2 are given as a guide. Optimization may be required, depending on the detection reagent and instrument to be used. Antibody solutions can be reused several times, if adding sodium azide (final concentration of 0.02 % (w/v)) during the initial preparation. Cease reusing when signal development becomes weak and/or background signals and artefacts become an issue.

    Note that sodium azide is a hazardous chemical, thus appropriate safety measures must be followed during its handling.

19. When applying the ECL reagent to the membrane, the entire surface of the membrane should be evenly exposed to the reagent. For this, pipette the ECL solution into a container that is just large enough to accommodate the membrane, and, with a pair of forceps, gently roll the membrane with the protein side down onto the ECL solution; leave for several seconds, and drip off ECL reagent before imaging (excess reagent during imaging may cause blotching). Place the membrane side up on a clear transparent sheet, and load into the imager instrument. Exposure times have to be optimized for the ECL reagent and the instrument used.

Acknowledgments

We thank Dr Joe Ciccotosto and Dr Laura Bica for sharing their expertise in the preparation of primary neuronal cultures, and Prof Sam Gandy for generously providing the 369 antibody. We also acknowledge support from the Judith Jane Mason and Harold Stannett Williams Memorial Foundation.

References

1. Hussain I, Powell D, Howlett DR et al (1999) Identification of a novel aspartic protease (Asp 2) as beta-secretase. Mol Cell Neurosci 14:419–427
2. Lin X, Koelsch G, Wu S et al (2000) Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein. Proc Natl Acad Sci U S A 97:1456–1460
3. Sinha S, Anderson JP, Barbour R et al (1999) Purification and cloning of amyloid precursor protein beta-secretase from human brain. Nature 402:537–540
4. Vassar R, Bennett BD, Babu-Khan S et al (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 286: 735–741
5. Yan R, Bienkowski MJ, Shuck ME et al (1999) Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. Nature 402:533–537
6. Tamagno E, Bardini P, Obbili A et al (2002) Oxidative stress increases expression and activity of BACE in NT2 neurons. Neurobiol Dis 10:279–288
7. Tamagno E, Guglielmotto M, Aragno M et al (2008) Oxidative stress activates a positive feedback between the gamma- and beta-secretase cleavages of the beta-amyloid precursor protein. J Neurochem 104:683–695
8. Tamagno E, Parola M, Bardini P et al (2005) Beta-site APP cleaving enzyme up-regulation induced by 4-hydroxynonenal is mediated by stress-activated protein kinases pathways. J Neurochem 92:628–636
9. Tamagno E, Guglielmotto M, Giliberto L et al (2009) JNK and ERK1/2 pathways have a dual opposite effect on the expression of BACE1. Neurobiol Aging 30:1563–1573
10. Guglielmotto M, Monteleone D, Giliberto L et al (2011) Amyloid-beta(4)(2) activates the expression of BACE1 through the JNK pathway. J Alzheimers Dis 27:871–883
11. Tong Y, Zhou W, Fung V et al (2005) Oxidative stress potentiates BACE1 gene expression and Abeta generation. J Neural Transm 112: 455–469
12. Mouton-Liger F, Paquet C, Dumurgier J et al (2012) Oxidative stress increases BACE1 protein levels through activation of the PKR-eIF2alpha pathway. Biochim Biophys Acta 1822:885–896
13. Quiroz-Baez R, Rojas E, Arias C (2009) Oxidative stress promotes JNK-dependent amyloidogenic processing of normally expressed human APP by differential modification of alpha-, beta- and gamma-secretase expression. Neurochem Int 55:662–670
14. Kao SC, Krichevsky AM, Kosik KS, Tsai LH (2004) BACE1 suppression by RNA interference

in primary cortical neurons. J Biol Chem 279: 1942–1949
15. Kwak YD, Wang R, Li JJ et al (2011) Differential regulation of BACE1 expression by oxidative and nitrosative signals. Mol Neurodegener 6:17
16. Chen CH, Zhou W, Liu S et al (2011) Increased NF-kappaB signalling up-regulates BACE1 expression and its therapeutic potential in Alzheimer's disease. Int J Neuropsychopharmacol 13:77–90
17. Jo DG, Arumugam TV, Woo HN et al (2010) Evidence that gamma-secretase mediates oxidative stress-induced beta-secretase expression in Alzheimer's disease. Neurobiol Aging 31: 917–925
18. Modarresi F, Faghihi MA, Patel NS et al (2011) Knockdown of BACE1-AS nonprotein-coding transcript modulates beta-amyloid-related hippocampal neurogenesis. Int J Alzheimers Dis 2011:929042
19. Perry G, Cash AD, Smith AI (2002) Alzheimer disease and oxidative stress. J Biomed Biotechnol 2:120–123
20. Swomley AM, Förster S, Keeney JT et al (2014) Abeta, oxidative stress in Alzheimer disease: evidence based on proteomics studies. Biochim Biophys Acta 1842:1248–1257
21. Wirz KT, Keitel S, Swaab DF et al (2014) Early molecular changes in Alzheimer disease: can we catch the disease in its presymptomatic phase? J Alzheimers Dis 38:719–740
22. Hardas SS, Sultana R, Clark AM et al (2013) Oxidative modification of lipoic acid by HNE in Alzheimer disease brain. Redox Biol 1:80–85
23. Tan J, Li QX, Ciccotosto G et al (2013) Mild oxidative stress induces redistribution of BACE1 in non-apoptotic conditions and promotes the amyloidogenic processing of Alzheimer's disease amyloid precursor protein. PLoS One 8:e61246
24. Berridge MV, Herst PM, Tan AS (2005) Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. In: El-Gewely MR (eds) Biotech Annu Rev 11:127–152. Elsevier, New York. http://dx.doi.org/10.1016/S1387-2656(05)11004-7
25. Bass DA, Parce JW, Dechatelet LR et al (1983) Flow cytometric studies of oxidative product formation by neutrophils: a graded response to membrane stimulation. J Immunol 130: 1910–1917
26. Wang H, Joseph JA (1999) Quantifying cellular oxidative stress by dichlorofluorescein assay using microplate reader. Free Radic Biol Med 27:612–616
27. Brewer GJ, Torricelli JR, Evege EK, Price PJ (1993) Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. J Neurosci Res 35:567–576
28. Lesuisse C, Martin LJ (2002) Long-term culture of mouse cortical neurons as a model for neuronal development, aging, and death. J Neurobiol 51:9–23

Chapter 5

Advanced Assay Monitoring APP-Carboxyl-Terminal Fragments as Markers of APP Processing in Alzheimer Disease Mouse Models

Ana García-Osta and Mar Cuadrado-Tejedor

Abstract

The 99-amino-acid-long APP-carboxy-terminal fragment, named C99, is a membrane-bound peptide generated from the amyloid precursor protein (APP) by β-secretase cleavage and is the direct precursor of amyloid beta (Aβ). Here we describe a method for the quantification of C99. The amount of C99 is an indicative value of the amyloid pathology in an Alzheimer's disease (AD) model, and could be used as a marker to study AD progression in comprehensive experiments, including screening for new compounds and repurposing of drugs to treat AD.

Key words APP processing, C99, C83, Bis-Tris gels, Tg2576 mouse model

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by the accumulation of extracellular senile plaques, composed of amyloidal-β (Aβ) peptides [1, 2], and intracellular neurofibrillary tangles (NFTs), composed mainly of aggregates of the microtubule-associated protein tau [3].

Blocking Aβ production has been one of the main approaches to treat AD. Aβ is formed from the amyloid precursor protein (APP) processing, which is altered in AD. First, APP is cleaved by either α- or β-secretase, leading to the formation of the 83- or 99-amino acid long APP-carboxy-terminal fragments (APP-CTFs), named APP-C83 and APP-C99, respectively. The APP-C83 fragment is next processed by the intramembrane γ-secretase generating the APP intracellular domains (AICDs), and extracellular nontoxic p3 fragments. The same secretase cuts the APP-C99 fragment generating the Aβ peptides (38–43 amino acids long) [4, 5]. Among the Aβ peptides, Aβ42 has the highest propensity to aggregate and is considered to be the most toxic Aβ species [6].

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_5,

For convenience, we will refer to this form here as 'Aβ'. During decades, scientists have been working trying to prevent the conversion of APP into Aβ.

The transgenic mouse model Tg2576 seems to be a promising laboratory tool to test potential modulators of Aβ formation [7]. This mouse overexpresses the human APP with the Swedish mutation (APPswe) and, with age, develops extracellular amyloid deposits in expected AD-affected brains areas such as the hippocampus and cortex. Indeed, C-terminal products of α-, β-, and γ-secretase cleavage are readily detectable in the brain of these transgenic mice since the age of 7–8 months [7]. Moreover, these animals develop an age-dependent impairment in memory. The production of the C99 fragment from APP is believed to be the rate-limiting step in releasing smaller aggregation-prone peptide fragments, such as Aβ peptide, from the membrane [8]. The accumulation of these APP fragments is thought to ultimately lead to the neurodegeneration that is evident in Alzheimer's disease (AD) [9], and that could contribute to AD pathology independently of Aβ [10]. More recently, the accumulation of intracellular C99 has been suggested an early marker of AD pathology [11]. The determination of C99 levels in AD models presented here can constitute a new strategy to study the onset and progression of AD and the effects of candidate therapeutic compounds in comprehensive drug screening experiments.

2 Materials

Prepare solutions in distilled water (or milli Q water).

2.1 Animal Models

Tg2576 AD transgenic mice are used in this assay. Tg2576 mice express the human 695 amino acid isoform of APP containing the Swedish double mutation (APPswe) [(APP695) Lys670- Asn, Met671-Leu] driven by a hamster prion promoter. These mice were inbred on a C57BL/6JxSJL genetic background. Tg2576 AD mice exponentially accumulate Aβ peptide in the brain between 7 and 12 months of age, and they exhibit memory impairment in the fear conditioning test from 6 months of age [7]. Animals are housed 4–5 per cage with ad libitum access to food and water, and they are maintained in a temperature-controlled environment on a 12 h light/dark cycle. For the assay 12-month-old female Tg2576 mice are used ($n = 10–12$).

2.2 Preparation of Soluble and Insoluble Brain Protein Fractions

- Sodium dodecyl sulfate (SDS) lysis buffer: 2 % (w/v) SDS, 10 mM, Tris–HCl (pH 7.4), protease inhibitors (Complete™ Protease Inhibitor Cocktail, Roche), phosphatase inhibitors (0.1 mM Na_3VO_4, 1 mM NaF).

- Sonicator.
- Standard protein assay such as the bicinchoninic acid (BCA, Thermo Scientific, Pierce) assay, or equivalent.

2.3 Western Blot Analysis

2.3.1 Protein Sample Preparation

- XT sample buffer: 4× premixed protein sample buffer for use with all Criterion XT gels, from Bio-Rad (catalog #161-0791) or equivalent. This ensures lane-to-lane consistency and guarantees reproducible results.
- XT reducing agent: 20× reducing agent for use with all Criterion XT gels, from Bio-Rad (catalog #161-0792) or equivalent. This is a pH neutralized and stabilized solution of tris(2-carboxyethyl) phosphine (TCEP), which is a reducing agent used to break disulfide bonds within and between proteins as a preparatory step for gel electrophoresis.
- Thermoblock.

2.3.2 Electrophoresis

- Criterion XT precast gel system (Bio-Rad) [12].
- Bis-Tris precast gels Criterion XT 4–12 % (12+2 well, Bio-Rad, catalog #345-0123). Criterion XT precast gels are formulated at pH near neutrality to optimize gel matrix stability, significantly delaying acrylamide hydrolysis, which occurs in traditional Laemmli systems. 4–12 % acrylamide allows the separation of small to mid-sized proteins. Bis-Tris gels allow consistent results and have stability for a minimum of 1 year.
- Bis-Tris running buffer XT MES. (Bio-Rad; Catalog #161-0789). MES is the common name for the compound 2-(*N*-morpholino) ethanesulfonic acid).. This is an optimized 20× electrophoresis buffer for use with Criterion XT Bis-Tris gels, which results in consistently resolved bands throughout the life of the gel.

2.3.3 Immunoblotting Components: Transfer to Membrane

- Polyvinylidene difluoride (PVDF) (positively-charged nylon) transfer membrane (0.2 μm removal rating; Hybond LFP, Amersham Biosciences, UK; Catalog #RPN303LPF) or equivalent.
- Transfer buffer: 0.025 M Tris, 0.192 M glycine, 20 % methanol.
- Bio-Rad Criterion Blotter.
- Bio-Ice cooling unit.

2.3.4 Antibodies Incubation

- Tris buffered saline (TBS) (10×): 1.5 M NaCl, 0.1 M Tris–HCl, pH 7.4.
- TBS containing 0.05 % Tween-20 (TBST).
- Blocking solution: 5 % milk in TBS.

- Diluent solution: 2.5 % milk in TBS.
- Rabbit polyclonal anti-APP C-terminal (CT19, amino acids 676–695 of hAPP) (1:2,000, Sigma-Aldrich, St. Louis, MO, USA).
- Mouse monoclonal 6E10 (amino acids 1–17 of Aβ peptide, 1:1,000, Chemicon).
- Mouse monoclonal anti α-tubulin (1:10,000, Sigma).

2.3.5 Imaging and Quantification

- Chemiluminescence system (e.g. ECL, GE Healthcare Bioscience, UK).
- Hyperfilm™ ECL (GE Healthcare Bioscience) or equivalent.
- Quantification, data analysis software. Quantity One™ software v.4.6.3 (Bio-Rad) or equivalent.

3 Methods

3.1 Preparation of Soluble and Insoluble Brain Protein Fractions

Protein sample preparation is critical to differentiate the C99 fragment from other APP-related products that might be detected with the same antibody, and to obtain clear and accurate resolution of protein bands.

1. Animals should be sacrificed by cervical dislocation, removing the brain and quickly dissecting the hippocampus. Homogenize the hippocampus with six volumes of SDS lysis buffer.
2. The homogenates are sonicated for 2 min and centrifuged at 100,000 × *g* for 1 h. Aliquots of the supernatant should be frozen at −80 °C.
3. The protein concentration of the homogenates should be determined using a standard protein assay such as the BCA assay (Thermo Scientific, Pierce) or equivalent, prior to sample preparation (*see* **Note 1**).

3.2 Western Blot Analysis

For Western blot analysis using the Criterion XT precast gel system (Bio-Rad) [12], aliquots of the protein extracts are mixed with XT sample buffer plus XT reducing agent and boiled. Proteins are separated in a Criterion precast gel and transferred to membranes. The membranes are blocked, followed by overnight incubation with antibodies (*see* steps below).

3.2.1 Protein Sample Preparation

Protein samples for the Criterion XT precast gel system (Bio-Rad) are prepared in a reducing buffer.

1. Prepare fresh loading buffer by premixing XT sample buffer with XT reducing buffer.

2. Mix loading buffer (4×) with 40 μg of protein sample and heat 5 min at 95 °C (use a thermoblock) to denature the proteins. The use of TCEP in combination with Bio-Rad's optimized buffers maintains proteins in a fully reduced state during the electrophoresis run, eliminating the need for an anti-oxidant in the upper buffer chamber.

3.2.2 Electrophoresis: Running Conditions

- Power: 150 V, constant.
- Starting current: 90–100 mA (10 min).
- Final current: 150 mA.
- Stop when dye front reaches 1 cm from the bottom of the gel.

3.2.3 Transfer to Membrane

Cut the PVDF membrane to the appropriate size. Then activate it by submerging in methanol for 15 s, place it in purified water and incubate in ice cold transfer buffer for 5 min. The gel also needs to be equilibrated for 3–5 min in ice cold transfer buffer.

Use the Bio-Rad Criterion Blotter for the transfer. Use a frozen Bio-Ice cooling unit, and a magnetic stir bar into the buffer tank. Transfer the protein at 340 mA for 1 h (*see* **Notes 2–4**).

3.2.4 Antibodies Incubation

- After transfer, wash the membrane with TBS for 5 min at room temperature.
- Incubate membrane in 5 % milk in TBS ml for 1 h at room temperature.
- Wash with 15 ml of TBS/T for 5 min.
- Incubate membrane with primary antibody: rabbit polyclonal anti-APP C-terminal (CT19, against amino acids 676–695 of APP, 1:2,000 in 2.5 % milk in TBS), with gentle agitation. Incubate overnight at 4 °C.
- Wash three times with TBS/T, 15 min each time.
- Incubate with HRP-conjugated secondary antibody: (1:5,000 in 2.5 % milk in TBS) for 1 h at room temperature.
- Wash two times with TBS/T, 15 min each, and once with TBS.

3.2.5 Imaging and Quantification

Bands are detected by using an enhanced chemiluminescence system (e.g. ECL, GE Healthcare Bioscience, Buckinghamshire, UK), and autoradiographic exposure to Hyperfilm™ ECL (GE Healthcare Bioscience). Signals quantification performed using Quantity One™ software v.4.6.3 (Bio-Rad).

Analysis of the protein fraction revealed two protein bands migrating close to the 12-kDa molecular weight marker (Fig. 1). These bands corresponded to C99 and C83, based on their apparent molecular masses (*see* **Notes 5** and **6**).

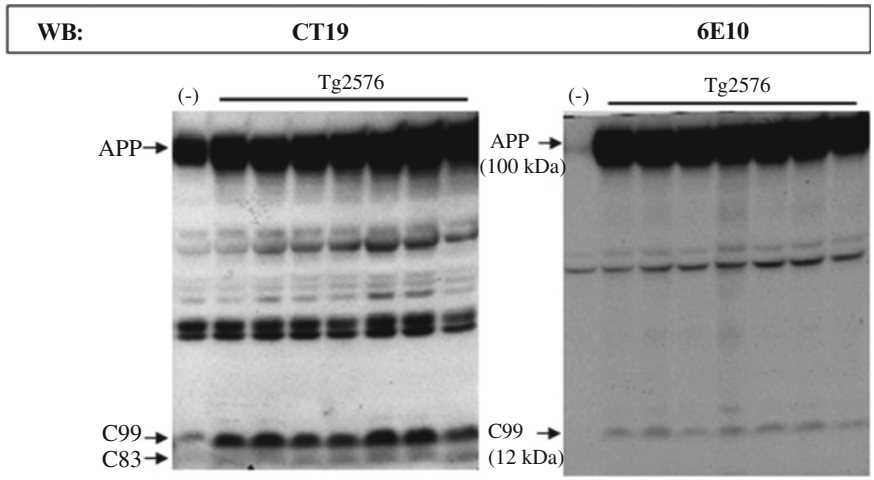


Fig. 1 Western blot of protein extracts obtained from the hippocampus of 12-month-old female Tg2576 mice. As a control, the hippocampus of a non transgenic animal was used (–). In the immunoblot obtained using the CT19 antibody, which specifically recognizes the first 19 amino acids of the carboxy terminal of the APP, the full-length APP (~100 kDa), and both APP derived carboxyl terminal fragments (CTFs) C99 and C83 (~12 kDa) are detected. In the immunoblot obtained using the 6E10 antibody, which specifically recognizes human APP in the fragment corresponding to amino acids 1–17 of Aβ peptide, only APP and C99, in Tg2576 samples, are revealed

4 Notes

1. Loading too much protein will result in poorly resolved bands in the Western blot in overloaded lane and out of shape electrophoretic patterns in adjacent lanes. Underloading the sample will prevent detection of bands that will became too faint for quantification. Therefore, the protein concentration of the homogenates should be carefully determined.
2. Wear gloves for sandwiching the gel and membrane between paper. Air bubbles between the gel and membrane can be removed by rolling them out using a pipette or 15 ml tube.
3. Fill the tank with cold transfer buffer. This will allow efficient and quantitative protein transfer and prevent gel and transfer buffer from overheating.
4. Power settings and transfer time. High electric fields may cause small proteins to be transferred too quickly or, conversely, incomplete separation of large proteins. Performing the run overnight at low voltage (30 V), may result in better separation, quantitative transfer over a broader range of molecular weights.
5. To confirm the band identities developed with the CT19 antibody, the membrane can be stripped and reprobed using 6E10 monoclonal antibody which recognizes C99 but not C83.

6. The 6E10 antibody recognizes only human APP, whereas CT19 recognizes both murine and human APP. A wild type sample can be introduced to recognize the specific band when the membrane is probed with the 6E10 antibody.

Acknowledgements

This study was supported by the foundation for applied medical research-Fundación para la Investigación Médica Aplicada (FIMA, Spain) and the fund for health research-Fondo de Investigación Sanitaria (FIS project 11/02861).

References

1. Glenner GG, Wong CW (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. Biochem Biophys Res Commun 120:885–890
2. Masters CL, Multhaup G, Simms G et al (1985) Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. EMBO J 4:2757–2763
3. Grundke-Iqbal I, Iqbal K, Tung YC et al (1986) Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. Proc Natl Acad Sci U S A 83:4913–4917
4. Haass C, Selkoe DJ (1993) Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide. Cell 75:1039–1042
5. Gandy S (2005) The role of cerebral amyloid beta accumulation in common forms of Alzheimer disease. J Clin Invest 115:1121–1129
6. Jarrett JT, Berger EP, Lansbury PT Jr (1993) The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. Biochemistry 32: 4693–4697
7. Hsiao K, Chapman P, Nilsen S et al (1996) Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. Science 274:99–102
8. Nunan J, Shearman MS, Checler F et al (2001) The C-terminal fragment of the Alzheimer's disease amyloid protein precursor is degraded by a proteasome-dependent mechanism distinct from gamma-secretase. Eur J Biochem 268:5329–5336
9. Lashuel HA, Hartley D, Petre BM et al (2002) Neurodegenerative disease: amyloid pores from pathogenic mutations. Nature 418:291
10. Jin LW, Hua DH, Shie FS et al (2002) Novel tricyclic pyrone compounds prevent intracellular APP C99-induced cell death. J Mol Neurosci 19:57–61
11. Lauritzen I, Pardossi-Piquard R, Bauer C et al (2012) The beta-secretase-derived C-terminal fragment of betaAPP, C99, but not Abeta, is a key contributor to early intraneuronal lesions in triple-transgenic mouse hippocampus. J Neurosci 32:16243–16255a
12. Criterion XT Precast Gel Instruction Guide (Bio-Rad, catalog #345-9898). http://www.bio-rad.com/webroot/web/pdf/lsr/literature/MS4110130B.pdf

Chapter 6

Optical Super-Resolution Imaging of β-Amyloid Aggregation In Vitro and In Vivo: Method and Techniques

Dorothea Pinotsi, Gabriele S. Kaminski Schierle, and Clemens F. Kaminski

Abstract

Super-resolution microscopy has emerged as a powerful and non-invasive tool for the study of molecular processes both in vitro and in live cells. In particular, super-resolution microscopy has proven valuable for research studies in protein aggregation. In this chapter we present details of recent advances in this method and the specific techniques, enabling the study of amyloid beta aggregation optically, both in vitro and in cells. First, we show that variants of optical super-resolution microscopy provide a capability to visualize oligomeric and fibrillar structures directly, providing detailed information on species morphology in vitro and even in situ, in the cellular environment. We focus on *direct* Stochastic Optical Reconstruction Microscopy, *d*STORM, which provides morphological detail on spatial scales below 20 nm, and provide detailed protocols for its implementation in the context of amyloid beta research. Secondly, we present a range of optical techniques that offer super-resolution indirectly, which we call multi-parametric microscopy. The latter offers molecular scale information on self-assembly reactions via changes in protein or fluorophore spectral signatures. These techniques are empowered by our recent discovery that disease related amyloid proteins adopt intrinsic energy states upon fibrilisation. We show that fluorescence lifetime imaging provides a particularly sensitive readout to report on the aggregation state, which is robustly quantifiable for experiments performed either in vitro or in vivo.

Key words Amyloid beta, Amyloid fibrils, In vivo imaging, Super-resolution microscopy, Multi-parametric imaging

1 Introduction

Protein misfolding and aggregation are root causes of neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases. Their pathological hallmarks are caused by the conversion of initially soluble monomeric species into highly ordered, insoluble fibrillar amyloid species, a process affected by many parameters in the complex cellular environment in which it occurs. Key to an understanding of pathological pathways and thus unravelling

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_6,

mechanisms of disease, is an ability to visualize the steps involved in amyloidogenesis at a molecular level.

In Alzheimer's disease, amyloid beta (Aβ) form aggregates, the nature and morphology of which have been the subject of extensive research over recent years [1]. The kinetics of Aβ fibril formation have mostly been studied via experiments performed in vitro [2–4] for example, by monitoring the increase in extrinsic [5] or intrinsic [6] fluorescence upon aggregation. The latter method exploits the fact that the formation of amyloid fibrils leads to the attainment of an intrinsic fluorescence signature in the visible range, which is linked to their rich content of cross β-sheet structure and extensive hydrogen bonding networks (Fig. 1a–c) [6–8]. However, these experiments probe ensemble behavior and thus do not provide detailed insight into mechanisms of aggregation. Moreover, data gained from in vitro samples prepared in the test

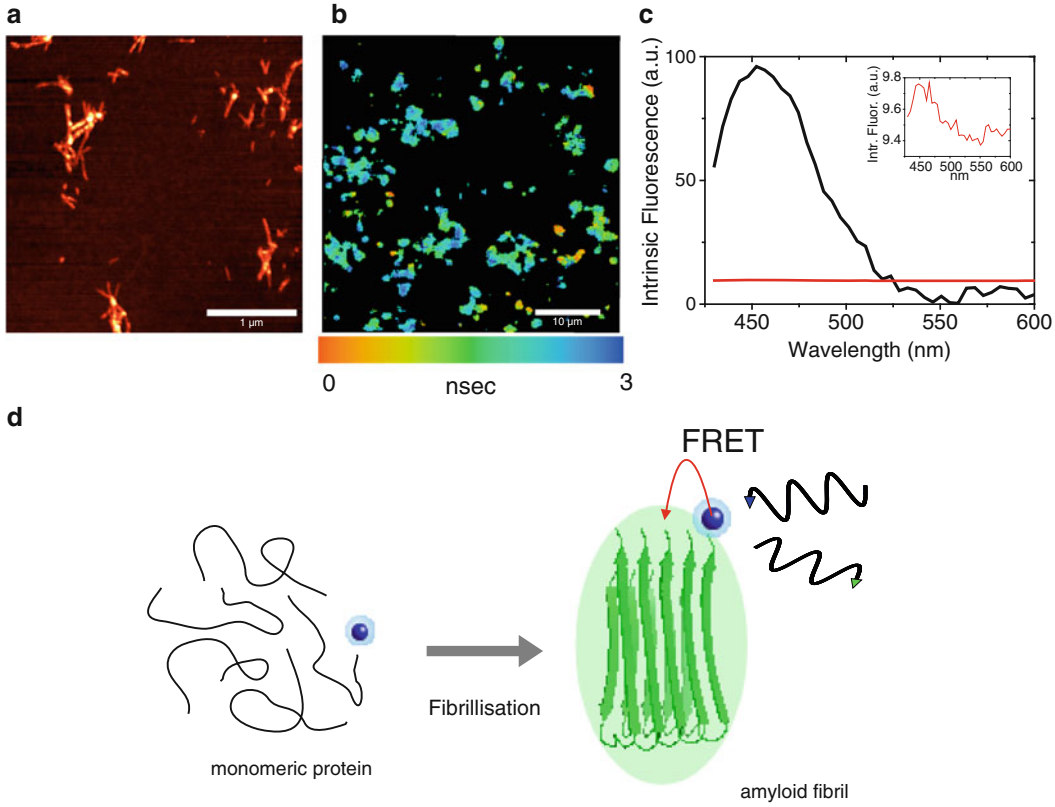


Fig. 1 Aβ42 fibrils formed in vitro develop an intrinsic fluorescence in the visible range. (**a**) Atomic force microscopy (AFM) image of Aβ42 fibrils. (**b**) Intrinsic fluorescence lifetime image of the same amyloid fibrils obtained by confocal microscopy. (**c**) *Black line*: emission spectrum of the intrinsic fluorescence from the aggregates. *Red line*: fluorescence signal from the corresponding monomeric protein. The laser excitation wavelength is at 405 nm. (**d**) Figure reprinted (adapted) with permission from [6] Copyright 2013 Wiley. (**d**) Simple schematic depicting the labeling and FRET interaction of the amyloid fibril formed with the attached fluorophore

tube are often not informative on mechanisms taking place in vivo, where the molecular environment is fundamentally different. There is thus a great need for techniques that monitor the self-assembly reactions of Aβ (and other amyloid proteins) directly and non-invasively.

Fluorescence microscopy is particularly suited for this task. It is relatively non-invasive, thus permitting the probing in the cellular environment; it is specific via molecular labeling techniques; it is dynamic and hence capable of providing real-time information on protein self-assembly reactions. In this chapter, we present two variants of optical microscopy techniques developed in our laboratory for the study of Aβ aggregation both in vitro and in situ in biological environment. Both techniques offer aggregate size specific information on a scale much smaller than the wavelength of light. The first technique, optical super-resolution microscopy provides 'direct photographs' of fibrils at a resolution approaching that of electron microscopy, without the limitations imposed by optical diffraction. The other, referred to as multi-parametric imaging, informs on aggregation reactions via changes in fluorescence properties, which correlate protein aggregation with lifetime changes in reporter fluorophores (Fig. 1d). The latter technique, whilst offering 'indirect' information on aggregate size, is particularly powerful tool for application in live cells and even higher organisms.

1.1 Optical Super-Resolution Imaging

Conventional fluorescence microscopy is limited in resolution by the diffraction of light; in an optical system with finite aperture, this resolution limit is typically in the 250 nm range. An isolated dye molecule, for example, gives rise to a blurred spot in the image plane (e.g. camera chip), the so-called point spread function (PSF) whose lateral dimension is of this size. This precludes its use for the direct observation of aggregate morphology which therefore, until recently, had only been possible with invasive techniques such as electron or atomic force microscopy (AFM) and gel electrophoresis (GE). With the recent advent of optical super-resolution techniques, however, the limitations imposed by optical diffraction have been overcome [9–12]. We have demonstrated that it is possible to obtain structural information on Aβ fibrils in situ, in cells, at a resolution close to that achievable with electron microscopy (Fig. 2) [13]. Here we present protocols for *direct* stochastic optical reconstruction microscopy (*d*STORM) [9] imaging of amyloid fibrils, a variant of the single molecule localization microscopy principle. It relies on the sequential photoswitching of dye labels from a non-fluorescent 'off' state to a fluorescent 'on' state, such that, at any given time, only a sparse subset of all fluorescent labels are active. Images of a suitably sparse fluorophore subset are recorded which consist of spatially distinct PSFs, permitting accurate determination of each fluorophore's position through computer fitting of PSF distributions. Repeating the imaging cycle

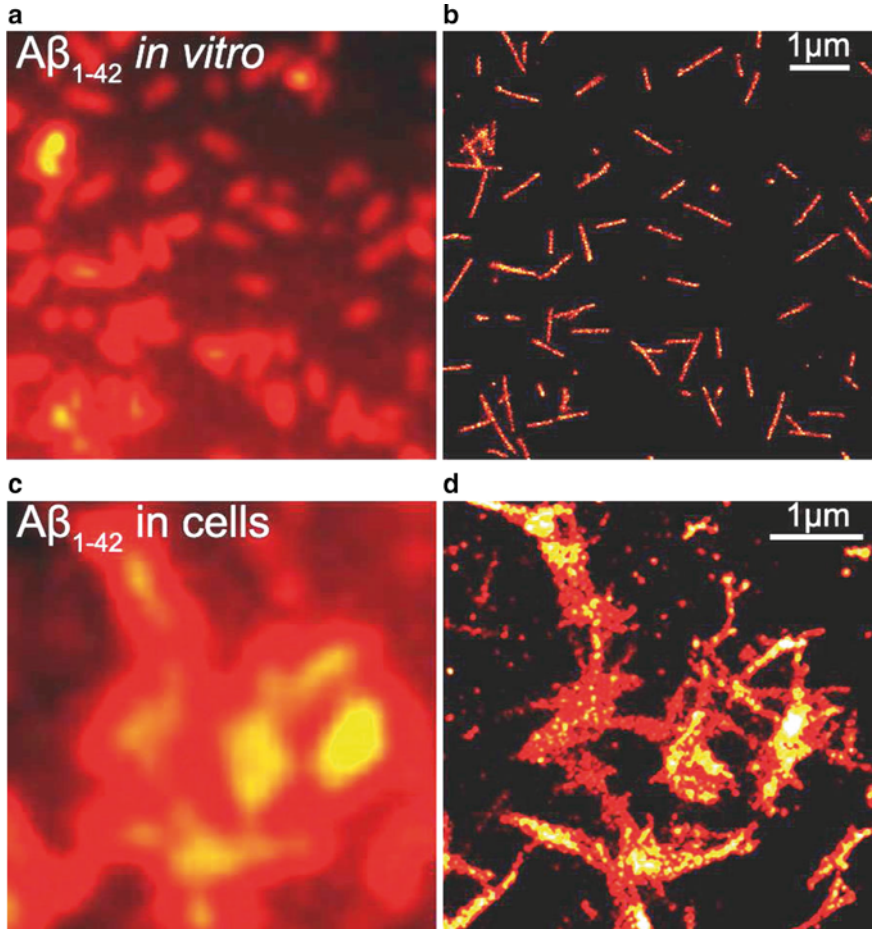


Fig. 2 Standard fluorescence microscopy and *d*STORM images. (**a**) Standard fluorescence microscopy image of $A\beta_{42}$ fibrils formed in vitro on the surface of a coverslip. (**b**) Corresponding *d*STORM image. A comparison with the image obtained with conventional imaging demonstrates the dramatic resolution enhancement provided by *d*STORM. (**c**) Standard fluorescence microscopy image of a section of a cell containing $A\beta_{42}$ fibrils. (**d**) Corresponding *d*STORM image. *d*STORM imaging permits the nature and morphology of intracellular $A\beta_{42}$ aggregates to be probed in situ. Figure reprinted (adapted) with permission from [13]. Copyright 2011 American Chemical Society

thousands of times permits the retrieval of structural information of the labeled sample on a 'super-resolved' scale. *d*STORM imaging is usually performed in total internal reflection (TIR) illumination mode, such that light is collected only from a thin region (ca 100 nm) close to the coverslip surface. The highly inclined illumination (HiLo) mode [14] is a variant which permits imaging deeper into cells (up to ca 5 μm deep). The technique can be extended to provide resolution in 3 dimensions with multiple colors, ideal for example, for co-localization studies of differentially labeled aggregates with cellular substructures. This technique has been applied successfully for other amyloid proteins apart from Aβ [15, 16], such as α-synuclein, and Tau [17, 18].

1.2 Multi-Parametric Imaging

The spectral properties of specific fluorophores is strongly influenced by changes in local (molecular) environment. Since protein aggregation induces modifications in the environment of reporter fluorophores, properties such as spectrum, lifetime and polarization of the emitted light are modified, all of which can be quantified through a range of techniques we refer to as multi-parametric imaging. In this chapter we focus on one variant of these techniques, informing on aggregation of amyloids via changes in fluorescence lifetimes of reporter dyes linked to the protein of interest. Both fluorescent proteins (such as GFP and YFP) and synthetic dyes (such as Alexa Fluor 488) covalently attached to the protein of interest can be used. We have shown previously that the method permits one to distinguish between monomeric, oligomeric and fibrillar species both in vitro and in vivo. In order to perform these measurements, a fluorescence microscope with Fluorescence Lifetime Imaging (FLIM) capability is required [19, 20]. The lifetime change is mediated by a fluorescence resonance energy transfer (FRET)-like process from the reporter fluorophore to the beta sheet rich amyloid scaffold (Fig. 1d). The phenomenon is explained by our recent reports that amyloid proteins develop intrinsic energy states excitable in the visible range upon aggregation [6, 8]. These energy states act as acceptors for suitably chosen fluorescent labels. We have shown that this permits the monitoring of aggregation reactions in vitro, in live cells and even in organisms [7, 21]. It is possible to distinguish between oligomeric and fibrillar species in vivo and to correlate this information with toxic phenotypes. The method is independent of prevailing reporter fluorophore density and works at low dye concentrations thus minimizing potential artifacts caused by steric interference or perturbation of pathological function of forming aggregates. We have applied the technique successfully for a range of amyloid proteins including Aβ, α-synuclein, and Tau both in vitro, in live cells and in organism models of aggregation [7, 18, 21].

The following protocols use Aβ as a representative example for the successful imaging of amyloid aggregation reactions by *d*STORM and lifetime imaging. Protocols need to be suitably adjusted for other amyloid proteins; hints and references to literature are given, where appropriate, to facilitate this adaptation.

2 Materials

1. Unlabeled $A\beta_{40}$ and $A\beta_{42}$ (Bachem GmbH, Weil am Rhein, Germany). $A\beta_{40}$ and $A\beta_{42}$, labeled with Hilyte Fluor™ 488 and Hilyte Fluor™ 647 (fluorescent labeling dye solutions) (Anaspec, Fremont, USA).
2. 1× Phosphate buffered saline (PBS) (e.g. Life Technologies).
3. Milli-Q water.

4. 0.20 μm filters (e.g. Millipore).
5. Triton X-100.
6. Tween-20.
7. Roswell Park Memorial Institute 1640 medium (RPMI 1640 medium) without phenol red (e.g. Life Technologies).
8. 37 % formaldehyde solution.
9. Ammonium hydroxide (NH_4OH).
10. Trifluoroacetic acid (TFA).
11. Hexafluoroisopropanol (HFIP).
12. Sodium azide.
13. Cell culture media and supplements:
    - Minimal essential medium Eagle (MEM) (e.g. Sigma).
    - Nutrient mixture F12 Ham (e.g. Sigma).
    - 1 % MEM non-essential amino acids (e.g. Sigma).
    - 1 %L-glutamine (e.g. Sigma).
    - 1 % antibiotic-antimycotic solution (Life technologies): This solution contains 10,000 U of penicillin, 10,000 μg of streptomycin and 25 μg of amphotericin B per mL (to prevent fungal contamination of cell cultures).
    - 15 % Fetal bovine serum (FBS) solution (e.g. Life Technologies).
    - 2 % B27 supplement medium (for serum free growth medium) (e.g. Life Technologies).
14. Antibodies. Primary antibody: monoclonal anti-Aβ antibody (6E10) (Covance, Leeds, UK).
    Secondary antibody: Alexa Fluor® 647 goat anti-mouse IgG antibody, which is labeled with a tandem dye construct that is excited at 647 nm and has an emission peak at 668 nm (Life Technologies).
15. HeLa or SH-SY5Y human neuroblastoma cells from the European Collection of Cell Cultures (Sigma).
16. LabTek II 8-well chambered coverglass with a thickness of 130 μm (e.g. Fisher Scientific).
17. For preparation of photoswitching buffer for *d*STORM with cyanine dyes (*see* also Subheading 3): Reagents and solutions: Glucose; Glucose oxidase (50 KU) lyophilized powder; Tris (e.g. Sigma); Catalase (10 mg/mL) (Sigma); 1 M Tris–HCl; Mercaptoethylamine (MEA)-HCl (e.g. Sigma) (*see* **Note 1**); 1 M Tris(2-carboxyethyl)phosphine (TCEP) (e.g. Sigma); Glycerine; 1 M KCl solution.
18. Glass Bottom culture dishes (MatTek Corporation, MA, USA).
19. LabTek II (Nunc/Fisher Scientific) (*see* **Note 2**).

20. Inverted widefield microscope with a high numerical aperture (NA) objective lens (oil-immersion objective with NA > 1.4).
21. Immersion oil.
22. Confocal microscope with time correlated single photon counting (TCSPC) module.
23. Different laser lines (excitation at 640, 560, 490 and 405 nm), maximum power of at least 150 mW
24. Pulsed laser, with a laser repetition rate of 40 MHz (for example super-continuum laser SC-450 from Fianium, UK).
25. Emission filters (band-pass) and dichroic beam splitters followed by emission filters (they should be laser flat and multi-edge).
26. Electron Multiplying (EM) charge-coupled device (CCD) camera, (EM-CCD), cooled, with high quantum yield and low read-out and background noise (e.g. iXon Andor, UK).
27. Software to control camera.
28. Photomultiplier tube (PMC-100, Becker & Hickl GmbH, Germany).
29. Acousto-optic tunable filter (AOTFnC-VIS, AA Opto-electronic).
30. Image processing software based on MATLAB computing language (The MathWork Inc., Natick, USA).
31. LabVIEW program (National Instruments, UK).
32. SPC Image software (Becker & Hickl GmbH, Germany).

3 Methods

3.1 Peptide Solutions Preparation

Labeled peptide:

1. Dissolve 1 mg of labeled $A\beta_{42}$ in 200 μL of 1 % ammonium hydroxide at 4 °C.
2. Prepare aliquots, snap-freeze sample in liquid nitrogen and store below −80 °C.
3. Determine the exact concentration of the stock solution with quantitative amino acid analysis and absorption spectroscopy.
4. Use each aliquot after thawing, fresh, only once (*see* **Note 3**).

Unlabeled peptide:

1. Dissolve 1 mg of peptide in trifluoroacetic acid (TFA), keep on ice under fume hood.
2. Sonicate the solution for 30 s on ice and lyophilize overnight.
3. Redissolve lyophilized peptide in 1 mL of cold hexafluoroisopropanol (HFIP).

4. Prepare aliquots and put into centrifugal evaporator.
5. Determine the exact concentration of the stock solution with quantitative amino acid analysis.

3.2 Conversion of Monomeric Protein to Fibrils In Vitro

Dilute the peptide in 50 mM sodium phosphate buffer, pH 7.4, to a final concentration of 50 μM in an Eppendorf tube. Incubate the peptide solutions at 37 °C for 1–7 days.

3.3 Super-Resolution Fluorescence Microscopy by Single Molecule Localization (dSTORM Imaging)

This method is based on inducible photoswitching properties of certain cyanine dyes, which require the use of a photoswitching buffer.

3.3.1 Photoswitching Buffer Preparation

Three stock solutions should be prepared in advance, prior to *d*STORM imaging. These are required to optimize the photoswitching properties of the dyes used:

1. Glucose oxidase/catalase enzyme stock solution (50 mL) (all concentrations are final concentrations): 100 μL Catalase (0.02 mg/mL), 200 μL TCEP (4 mM), 25 mL Glycerine (50 %), 22.5 mL distilled water, 1.25 mL KCl (20 mM), 1 mL Tris–HCl (20 mM), 50 mg Glucose oxidase (1 mg/mL). Prepare 100 μL aliquots and store them at −20 °C (*see* **Note 4**).
2. Glucose stock solution (50 mL): 5 g of glucose (100 mg/mL), 45 mL of distilled water, 5 mL of 10 % Glycerine. Prepare 1 mL aliquots and store at −20 °C.
3. MEA stock solution (10 mL): Dissolve 1.136 g of MEA-HCl in 10 mL of distilled water HCl to produce a 1 M stock solution. Prepare 200 μL aliquots and store at −20 °C.

For each *d*STORM experiment, a fresh photoswitching buffer solution is required. For a LabTek II chambered coverglass prepare in a mixing tube: 50 μL of enzyme stock solution, 400 μL solution of glucose stock solution, 50–100 μL MEA stock solution. Mix gently.

3.3.2 Imaging of Amyloid Fibrils with dSTORM In Vitro

1. Deposit ca. 10 μL of sample on a glass or quartz coverslip glued to the bottom of an imaging chamber, or at the bottom of a LabTek II well. Let fibrils adsorb onto the surface for ca. 30 min.
2. Prepare the photoswitching buffer solution, as described in previous section.
3. Add the photoswitching buffer solution into the LabTek II chamber well and fill it up with additional PBS. Then seal chamber with a coverslip or silicon sheet. Be careful to avoid formation of air bubbles.

4. The final MEA concentration should be 50–100 mM. The final pH should be between 6.0 and 8.5.
5. Place the chamber on the microscope stage and prepare the laser illumination for excitation of the different dyes (*see* **Note 5**).
6. The excitation (at wavelengths depending on the dyes used) and reactivation (at 405 nm) laser beams should be collimated by a beam expanding telescope and combined via dichroic mirrors. They should subsequently be focused onto the back focal plane of the objective.
7. Sample fluorescence should be imaged onto a sensitive electron multiplication gain CCD camera (EM-CCD, e.g. Andor iXon) after passage through dichroic and bandpass filters to reject stray light.
8. Image amyloid fibrils using total internal reflection fluorescence (TIRF) microscopy mode to improve the signal-to-noise ratio. Optimize the labeling density for *d*STORM imaging using mixtures containing different concentration ratios of labeled and unlabeled fibrils respectively (*see* **Notes 5** and **6**).
9. Obtain first a conventional fluorescence image in TIRF mode covering a large field of view, FOV (e.g. 100×100 μm^2 or 20×20 μm^2). Use a low excitation intensity (<0.1 kW/cm^2). Then select a smaller FOV for *d*STORM image acquisition.
10. Zoom in onto a smaller area, reduce the area on the camera chip to ~10×10 μm^2 (*see* **Notes 7** and **8**).
11. Vary the excitation intensity, EM gain and exposure time to optimize signal-to-noise ratio and photoswitching conditions for each of the used fluorophores. The exposure time of the camera should be adapted according to the "on" state time of the fluorescent dye used. Typical "on" times should last 1–4 camera frames. Usually exposure times of 10–12 ms are appropriate for Alexa Fluor® dyes. The reactivation laser should only be turned on when the number of active fluorophores in the field of view has greatly diminished. If necessary, adjust the focus to ensure that the fluorescent spots exhibit symmetric point spread functions (PSF).
12. Increase the excitation intensity to 2–5 kW/cm^2.
13. Acquire between 10,000 and 20,000 fluorescence frames.
14. Save the acquired frames as an image stack in a .tif file. From each image stack a reconstructed *d*STORM image can be generated by using an image processing software (*see* **Note 9**).

3.3.3 Imaging of Amyloid Fibrils with dSTORM in Cells

1. Cell culture and fixation
   (a) Prepare the serum containing culture medium by mixing equal volumes (1:1) of: minimal essential medium, nutrient mixture F12 Ham, 15 % fetal bovine serum,

1 % L-glutamine, 1 % MEM non-essential amino acids and 1 % antibiotic-antimycotic solution.

(b) Maintain the cells in this serum-containing culture medium.

(c) Plate cells 1 day before imaging at a density of 10,000 cells per well, in a Lab-Tek II chamber slide in serum-containing culture medium (*see* Subheading 2).

(d) Replace the serum containing medium with serum-free medium (replace FBS by 2 % B27).

(e) Add the peptide or the fibrils in this serum-free medium at the desired concentration. If labelled peptides or fibrils are used, adjust the labeling density appropriately by mixing labeled with unlabeled fibrils (*see* **Note 5**).

(f) After incubation, wash the cells in each chamber with PBS and fix them with 4 % formaldehyde, added for 10 min at room temperature (*see* **Note 10**).

(g) Wash the cells twice with PBS.

(h) Add 200 μL of PBS and 1 μL of sodium azide in each LabTek II well.

(i) Fixed cells can be stored in PBS-sodium azide at 4 °C, for several days.

2. Immunofluorescence staining

Direct labeling of amyloid protein provides the best resolution for imaging since the dyes are small and covalently attached to the protein of interest but this only works when protein is added exogenously to the cells [18]. On the other hand, immunocytochemisty permits endogenous proteins to be studied at physiological concentration. Here the resolution is reduced by the physical size of antibodies that link peptide and reporter dye. An immunocytochemistry protocol using primary and secondary antibodies for labelling is detailed below. The steps include fixation, permeabilization, blocking, and sequential staining with primary and secondary antibodies.

(a) Remove the fixating solution from cells.

(b) Perform a permeabilization step: Add 0.5 % Tween-20 in PBS to the sample for 20 min at room temperature.

(c) Add the blocking solution to the sample, consisting of, 5 % goat serum and 0.05 % Tween-20 in PBS. Leave for 30 min at room temperature. The blocking solution prevents unspecific binding of antibodies.

(d) Add the monoclonal anti-Aβ antibody (6E10) to the sample in the same blocking solution (1:300) and incubate for 1 h.

(e) Wash the sample twice with 0.05 % Tween-20 in PBS.

(f) Add the secondary antibody (Alexa Fluor®647 goat anti-mouse IgG antibody) to the sample, in a dilution of 1:200 in

the blocking solution and incubate for 1 h at room temperature, protected from light.

(g) Wash the sample twice with 0.05 % Tween-20 in PBS, and once with PBS only.

(h) Add 200 μL of PBS with 0.5 % w/v sodium azide in each LabTek II well.

(i) Fixed and stained cells can be stored in PBS-sodium azide at 4 °C for several days, protected from light.

3. Imaging

(a) Use the fixed cells plated in LabTek II 8-well chamber slides for after immunofluorescence step, or with the labeled fibrils/peptides, after the desired incubation time.

(b) Repeat **steps 2–7** of Subheading 3.3.2.

(c) Imaging of cells: Use TIRF mode if proteins of interest are on the cell membrane close to coverslip, or highly inclined illumination mode if looking at intracellular protein distributions (*see* **Note 11**). A movable mirror before the focusing lens permits easy switching between TIRF and highly inclined or epifluorescence illumination modes [22].

(d) Find an appropriate area with cells to image, using transmitted light and find the appropriate focal plane.

(e) Zoom in on one area enclosing one cell and obtain a differential interference contrast (DIC) image.

(f) Obtain the conventional fluorescence image of the same area. Use low excitation intensity (<0.1 kW/cm^2).

(g) If there are two different labels in the sample, obtain the fluorescence image of one after the other and change the detection filters accordingly. First image the red dye and subsequently the green dye.

(h) In order to get finer details, reduce the FOV on the camera chip to 10×10 μm^2 and perform *d*STORM imaging. Start imaging with the red dye and then proceed with the green dye.

(i) Repeat **steps 11–14** of Subheading 3.3.2.

(j) After image acquisition, take another DIC image of the same smaller area in the cell (the 10×10 μm^2 one). This can be used to produce an overlay with the *d*STORM image. An example is shown in Fig. 3.

3.4 Data Analysis for dSTORM

1. To obtain super-resolved images of the fibrils in cells or in vitro, process the raw image data using appropriate localization algorithms, such as the open source rainSTORM localization microscopy software [23] or the rapidSTORM [24]. The former (rainSTORM) is a MATLAB based software for

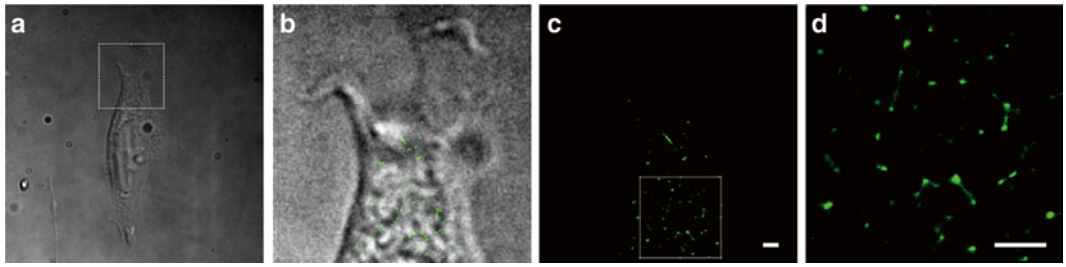


Fig. 3 Super-resolution imaging of $A\beta_{42}$ fibrils in neuroblastoma cells. (**a**) Differential interference contrast (DIC) image of a neuroblastoma cell. (**b**) Zoom-in DIC image of the area inside the square in (**a**), overlaid with super-resolved fluorescence image of the amyloid fibrils. The amyloid fibrils at 50 nM concentration had been added to the cell medium and incubated for 1 h with the cells. (**c**) Super-resolved fluorescence image of the amyloid fibrils. (**d**) Zoom-in on the area inside the square in (**c**) showing the super-resolved fluorescence image of the amyloid fibrils. Scale bars 1 μm. Unpublished data (2013)

Localization Microscopy image processing containing a simple to use graphical user interface (GUI) and a set of MATLAB scripts and functions (*see* **Note 12**).

2. Launch MATLAB. Browse to open the rainSTORM.m file and run it.
3. On the rainSTORM GUI that appears, select the .tif file which contains the raw data that you want to process.
4. Select a fitting algorithm. In most cases the 'Least-Squares Gaussian Halt 3' is the best suited for sparsely blinking datasets.
5. Input the pixel width. This will be dependent on the camera and magnification used on the microscope. Typical values range between 100 and 160 nm.
6. Input the PSF sigma value (the initial guess of the PSF standard deviation in each direction (X and Y) can vary with magnification and wavelength). A value of 1.3 is suitable in most cases.
7. "Radius of region of interest (ROI)" sets the pixel area that the algorithm will search for single molecules. Radius of ROI = 2 or 3 is appropriate for a pixel width of 160 nm. ROI = 3 or 4 is appropriate for a pixel width of 100 nm.
8. Tolerance, signal counts and maximum iterations should be left at default values in almost all cases. If using the "Thorough" algorithm it is advisable to increase the signal counts threshold.
9. Tick the "Display scale bar" and the "Display Sum image" boxes.
10. Click 'Process Images'—a wait bar will appear.
11. An initial super-resolution image will be generated once the localization algorithm process has completed. This is a preview without any further Quality Control factors than are specified

in the main algorithm. Also, a sum (diffraction-limited) image will be generated if selected in the rainSTORM GUI.

12. Click 'Open Reviewer' and a new GUI will appear to perform the Quality Control.
13. Input preliminary review parameters as indicated in the boxes (*see* **Note 13**) and click 'Run Reviewer'.
14. View the histograms and according to these refine the Quality Control parameters. Then click 'Run Reviewer' again.
15. You can adjust the contrast on the images generated by clicking on 'Adjust contrast'.
16. 'Save File'. This will save on screen images that were generated, the histogram images, the sum image (diffraction limited image), the data file and a text file containing the image reconstruction parameters.

3.5 Image Analysis

Use an image analysis software such as ImageJ (NIH, Bethesda, Maryland, USA) to analyse and process images, to determine properties such as the fibril cross sections and to overlay images from different fluorescent channels and DIC.

3.6 Fluorescence Lifetime Imaging

1. Use the $A\beta_{40}$ and $A\beta_{42}$ labeled fibrils or peptides and for excitation use the 490 nm wavelength. Here, any dye that has an excitation peak around 490 nm can be used to label the peptide, due to the required overlap with the intrinsic fluorescence spectrum, for the FRET mechanism to take place (*see* Subheading 1). Such labels are for example Alexa or HiLyte Fluor 488, YFP and GFP.
2. Place in vitro samples of labeled fibrils (~10 μL) on a coverslip.
3. The cells should have been prepared (e.g. 50,000 SH-SY5Y cells in a MatTek dish). After 24 h add the peptide or the amyloid fibrils labeled with the 488 dye to the cells and leave in incubation for the desired time.
4. Potentially perform cell washes with PBS if you are looking at the intracellular aggregation, in order to get rid of any remaining extracellular species.
5. For live cell imaging, incubate the cells in serum-free growth medium in glass bottom culture dishes. During the experiment place the samples in an incubator chamber at 37 °C and 5 % CO_2, mounted onto the microscope stage.
6. Imaging of in vitro and in vivo samples should be performed with a microscopy setup equipped with: a time correlated single photon counting (TCSPC) module [7] and a pulsed laser at 490 nm (for samples labeled with 488 dyes) with a 40 MHz repetition rate. In the present example, the setup includes: an Olympus Fluoview FV300 confocal scan unit and a pulsed

super-continuum source used for excitation, emitting a train of sub 10 ps pulses at 40 MHz repetition rates.

7. In the case of the pulsed super-continuum laser source, the output laser excitation beam should be collimated and should pass through a hot mirror assembly to remove infrared components at wavelengths greater than 700 nm. In order to tune the excitation wavelength, the visible portion of the spectrum passes through an acousto-optic tunable filter whose RF modulator is driven by software developed using LabVIEW.
8. Use a 20/80 broad bandwidth coated beam splitter to reflect the excitation light onto the sample, so that 80 % of the fluorescence signal passes through the confocal pinhole.
9. On the detection path use a 515 nm long-pass filter for the fluorescence light and project this onto a fast photomultiplier tube.
10. Measure first the instrument response function. For that purpose, use the reflected light, tuned at the detection wavelength (fluorescence wavelength) from either a cover slip or a coated mirror.
11. Acquire images for 100–300 s (10 cycles of 10–30 s) and verify that photobleaching is negligible during these acquisition times. Make sure that the excitation intensity is low enough to avoid photon pile-up.
12. Record lifetimes with the TCSPC system. Process all TCSPC images initially using SPC Image software. Fit the data with mono-exponential decay functions, taking into account the instrument response.
13. Use pixel binning until approximately a total of 3,500–5,000 photons are obtained per pixel (typically corresponding to a binning factor of 2 or 3).
14. Export the data and perform further image processing and data analysis using MATLAB.

4 Notes

1. Mercaptoethylamine (MEA) changes the pH slightly. The pH of the buffer is important and should be controlled, optionally adjusted with KOH or NaOH.
2. LabTek II chambered coverglass wells have a defined chamber volume of 1 mL.
3. The exact concentration of the labeled peptide should be determined by absorption spectroscopy of the Hilyte Fluor™ dye using an extinction coefficient of 70,000 M^{-1} cm^{-1} for Hilyte Fluor™ 488 and 250,000 M^{-1} cm^{-1} for Hilyte Fluor™ 647.

4. Glucose oxidase oxidizes glucose to gluconolactone and H_2O_2. It has a broad activity over the pH range of [4–7]. Glucose oxidase does not require any activators. Catalase decomposes H_2O_2. Its activity is constant over the pH range of [4.0–8.5].
5. The microscope setup can consist of a widefield inverted microscope and a high numerical aperture (NA) total internal reflection fluorescence (TIRF) objective. Here, we use the setup described in [13] which is based on a Nikon Eclipse TE 300 inverted widefield microscope, and a 100×, 1.49 NA TIRF objective lens.
6. The labeling ratio (labeled vs. unlabeled protein) should be defined for each experiment. A small labeling ratio (such as 1:20 labeled vs. unlabeled protein) is preferable because it means that the attached dye is less likely to affect the aggregation process.
7. Spatial drift during image acquisition is an important factor that should be taken into account. Typically, the use of fluorescent markers (beads) and/or the use of "autofocus" systems are means for correcting for any drift.
8. For two-color super-resolution microscopy, a step to correct for any chromatic offset or aberrations may be needed [22].
9. Currently, there are several algorithms available online to perform image analysis for the localization microscopy data. Some are published, such as the rainSTORM [23] or rapidSTORM [24]. In this work we show how to use the first software, which is based on MATLAB computing language.
10. The formaldehyde solution should be kept at 4 °C. Fixing the cells preserves the cellular structure thanks to the cross-linking of proteins by the fixative. Wear appropriate protective equipment and avoid contact with skin and eyes.
11. TIRF microscopy is used in fluorescence studies to reduce the background fluorescence and to effectively enhance the signal-to-noise ratio. A specific illumination configuration is applied in order to image a thin layer in the close vicinity of the coverslip/sample interface. A collimated beam with an angle of incidence larger than the critical angle is used, and thus the beam is subjected to total internal reflection, i.e. only the evanescent field, near-field wave with a penetration depth of approximately 100 nm excites the sample. Highly inclined illumination microscopy permits imaging inside the cell, as the illumination beam is "hitting" the sample at an angle different to the total internal reflection angle [14].
12. The rainSTORM software includes the following capabilities: (a) Localization using a "Sparse Segmentation and Gaussian Fitting" algorithm, (b) Quality Control using the Thompson Precision estimate of each localization, (c) Visualization using

"simple histogram image" or "jittered histogram", (d) One-click save of super-resolution images, together with quality-control histograms, and meta-data in a text file, (e) Estimation of the resolution of the super-resolved image and its registration in the text file, using the analysis developed in [23]. The rainSTORM software is available for use by any interested groups and can be uploaded from http://laser.cheng.cam.ac.uk/wiki/index.php/Resources .

13. The Quality Control parameters are defined as follows: (a) 'Updated Signal Counts' is a minimum brightness threshold. The higher this number the brighter a fluorescent molecule/position must be, to be an accepted localization and thus any dim static background signal will be excluded. (b) 'Updated Tolerance' excludes fitted candidates with a high least-squares residual. In practice, it is often best to leave this at 10 %. (c) 'Updated PSF Sigma Range' is the pixel width of each localization that is acceptable. Using Alexa Fluor® 647 or similar dyes with a 160 nm pixel size the theoretical value should be 1.3. Values larger than 1.3 can be a result of defocused fluorophores, multiple overlapping fluorophores or spherical aberration. A restrictive range can exclude slightly out of focus molecules. (d) 'Counts Per Photon' is a calibration value that can be found in the datasheet of the camera. It is dependent on the camera and the gain setting used. Using the correct value is required for accurate assessment of precision and resolution. (e) 'Localization Precision' applies a cutoff to reject localizations with poor localization precision as calculated by the Thompson Formula [25]. Values smaller than 50 nm will generate images with better mean localization precisions, but with fewer localizations in the final image. (f) 'Reconstruction Scale Factor' determines the size of the pixel in the super-resolution image. For example, with a pixel width of 160 nm in the raw data a reconstruction scale factor of 5 will generate super-resolution pixels of 32 nm. (g) 'Limit frame range' determines the subset of raw data to be processed. Often frames early in the sequence can suffer from mislocalizations as the blinking density is too high, or frames later on in the sequence may suffer from focus drift.

Acknowledgements

This work was funded by grants from the Medical Research Council UK (MR/K015850/1 and MR/K02292X/1), Alzheimer Research UK (ARUK-EG2012A-1), the EPSRC (EP/H018301/1) and the Wellcome Trust (089703/Z/09/Z). D.P. wishes to acknowledge support from the Swiss National Science Foundation and the Cambridge Wellcome Trust Senior Internship scheme.

References

1. Fitzpatrick AWP, Debelouchina GT, Bayro MJ et al (2013) Atomic structure and hierarchical assembly of a cross-β amyloid fibril. Proc Natl Acad Sci U S A 110:5468–5473
2. Lomakin A (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein. Proc Natl Acad Sci U S A 94:7942–7947
3. Knowles TPJ, Waudby CA, Devlin GL et al (2009) An analytical solution to the kinetics of breakable filament assembly. Science 326: 1533–1537
4. Cohen SIA, Linse S, Luheshi LM et al (2013) Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. Proc Natl Acad Sci U S A 110:9758–9763
5. Hellstrand E, Boland B, Walsh DM, Linse S (2010) Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. ACS Chem Neurosci 1:13–18
6. Pinotsi D, Buell AK, Dobson CM et al (2013) A label-free, quantitative assay of amyloid fibril growth based on intrinsic fluorescence. Chembiochem 14:846–850
7. Kaminski Schierle GS, Bertoncini CW, Chan FTS et al (2011) A FRET sensor for non-invasive imaging of amyloid formation in vivo. ChemPhysChem 12:673–680
8. Chan FTS, Kaminski Schierle GS, Kumita JR et al (2013) Protein amyloids develop an intrinsic fluorescence signature during aggregation. Analyst 138:2156–2162
9. Heilemann M, van de Linde S, Schüttpelz M et al (2008) Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew Chem Int Ed Engl 47:6172–6176
10. Hell SW, Wichmann J (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. Opt Lett 19:780–782
11. Hess ST, Girirajan TPK, Mason MD (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys J 91:4258–4272
12. Betzig E, Patterson GH, Sougrat R et al (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642–1645
13. Kaminski Schierle GS, van de Linde S, Erdelyi M et al (2011) In situ measurements of the formation and morphology of intracellular β-amyloid fibrils by super-resolution fluorescence imaging. J Am Chem Soc 133: 12902–12905
14. Tokunaga M, Imamoto N, Sakata-Sogawa K (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5:159–161
15. Esbjörner EK, Chan F, Rees EJ, Erdelyi M, Luheshi LM, Bertoncini CW, Kaminski CF, Dobson CM, Kaminski-Schierle GS (2014) Direct observations of the formation of amyloid β self-assembly in live cells provide insights into differences in the kinetics of Aβ(1–40) and Aβ(1–42) aggregation. Chem Biol 21 (6): 732–742
16. Fritschi SK, Langer F, Kaeser SA, Maia LF, Portelius E, Pinotsi D, Kaminski CF, Winkler DT, Maetzler W, Keyvani K, Spitzer P, Wiltfang J, Kaminski Schierle GS, Zetterberg H, Staufenbiel M, Jucker M (2014) Highly potent soluble amyloid-β seeds in human Alzheimer brain but not cerebrospinal fluid. Brain 137(11):2909–2915
17. Pinotsi D, Buell AK, Galvagnion C et al (2014) Direct observation of heterogeneous amyloid fibril growth kinetics via two-color super-resolution microscopy. Nano Lett 14:339–345
18. Michel CH, Kumar S, Pinotsi D et al (2014) Extracellular monomeric tau protein is sufficient to initiate the spread of tau protein pathology. J Biol Chem 289:956–967
19. Frank JH, Elder AD, Swartling J et al (2007) A white light confocal microscope for spectrally resolved multidimensional imaging. J Microsc 227:203–215
20. van Ham TJ, Esposito A, Kumita JR et al (2010) Towards multiparametric fluorescent imaging of amyloid formation: studies of a YFP model of alpha-synuclein aggregation. J Mol Biol 395:627–642
21. Murakami T, Yang SP, Xie L et al (2012) ALS mutations in FUS cause neuronal dysfunction and death in Caenorhabditis elegans by a dominant gain-of-function mechanism. Hum Mol Genet 21:1–9
22. Erdelyi M, Rees E, Metcalf D et al (2013) Correcting chromatic offset in multicolor super-resolution localization microscopy. Opt Express 21:10978–10988
23. Rees EJ, Erdelyi M, Pinotsi D et al (2012) Blind assessment of localisation microscope image resolution. Opt Nanoscopy 1:12
24. Wolter S, Löschberger A, Holm T et al (2012) rapidSTORM: accurate, fast open-source software for localization microscopy. Nat Methods 9:1040–1041
25. Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82: 2775–2783

Chapter 7

Protocols for Monitoring the Development of Tau Pathology in Alzheimer's Disease

Alberto Rábano, Raquel Cuadros, Paula Merino-Serráis, Izaskun Rodal, Ruth Benavides-Piccione, Elena Gómez, Miguel Medina, Javier DeFelipe, and Jesús Avila

Abstract

The microtubule-associated protein tau plays a critical role in the pathogenesis of Alzheimer's disease (AD) and several related disorders collectively known as tauopathies. Development of tau pathology is associated with progressive neuronal loss and cognitive decline. In the brains of AD patients, tau pathology spreads following a predictable, anatomically defined progression pattern that can be followed by immunohistochemistry looking at brain post-mortem samples from Alzheimer patients at different stages of the disease. Furthermore, since it has been proposed that AD may be a synaptopathy and dendritic spines of pyramidal neurons are the major targets of cortical synapses, the analysis of dendritic spines is a useful tool to study the correlation between tau phosphorylation at specific sites, synaptopathy and cognitive impairment. Finally, characterization of phosphorylated tau in detergent-insoluble protein aggregates could also be an indication of the neuropathological staging in AD. Here, we describe these three complementary protocols to follow the development of tau pathology in Alzheimer's disease.

Key words Dendritic spines, Detergent-insoluble aggregates, Gallyas stain, Hyperphosphorylation, Immunohistochemistry, Intracellular injections, Neurofibrillary degeneration, Neuropathology, Staging, Synaptopathy, Tau

1 Introduction

1.1 Tau Neuropathological Staging of Alzheimer's Samples. Staining Techniques

Current neuropathological diagnosis of Alzheimer's disease is closely related to staging of associated histological findings, based on the degree of involvement by characteristic lesions along a known pattern of disease progression within the brain. For a diagnostic work-up of cases recent guidelines put forward by the National Institute on Aging—Alzheimer's Association (NIA-AS) [1] recommend the assessment of (1) beta-amyloid plaque staging [2], (2) Braak and Braak staging of neurofibrillary degeneration [3], and (3) evaluation of neuritic plaque frequency according to the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) protocol [4]

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_7,

(jointly reported as an "ABC score"). In a large majority of cases the pathology of the Alzheimer's type, and in particular neurofibrillary degeneration, follows a highly predictable pattern of progression in the brain that has been classified by Braak and Braak. The staging system originally proposed by these authors defines I–VI stages based on the presence and density of characteristic argyrophylic inclusions (neurofibrillary tangles [NFTs], neuropil threads [NT]) in the medial temporal lobe and several brain isocortical regions. This staging system was subsequently adapted by the authors for routine use in paraffin-embedded tissue based on tau immunohistochemistry [5]. Stages I–II (transentorhinal) correlate with the prolonged preclinical phase of the disease; stages III–IV (limbic) with mild cognitive impairment (loss of episodic memory) or mild dementia; whereas advanced V–VI stages (isocortical) usually correspond to cases with moderate to severe dementia. Stages I–II are defined by the presence of pathological tau-reactive inclusions extending progressively from the transentorhinal to the entorhinal and hippocampal (CA1) cortex. At stages III and IV aberrant tau aggregates can be further observed in the subiculum, amygdala, thalamus and claustrum. Finally, at stages V–VI, tau-positive inclusions are also found in isocortical areas, extending finally from associative to primary motor and sensory cortical areas. Clinicopathological studies show that the Braak and Braak stage is the main neuropathological variable that correlates with cognitive decline in AD patients [6]. Additionally, this staging system allows comparability between case series with *post-mortem* studies by performing either immunohistochemical staining for hyperphosphorylated tau or Gallyas silver staining in tissue sections obtained from standard brain regions at the medial temporal lobe and several isocortical areas, although inter-rater reliability is improved by an easy transformation in a three-stages system, as recommended by the NIA-AS guidelines (B1: stages I–II, B2: stages III–IV, B3: stages V–VI). Accurate staging of AD-related tau-positive pathology may be particularly important in the classification of preclinical disease and in the identification of atypical AD phenotypes. Recent guidelines by the National Institute on Aging—Alzheimer's Association allow standardization of reports for diagnostic and research purposes. The following protocol is focused on the procedure of Braak and Braak staging in *post-mortem* brain.

For routine diagnostic purposes an immunohistochemical stain for hyperphosphorylated tau is preferable, as it can be performed along with other antibodies for a global diagnostic work-up (beta-amyloid, alpha-synuclein, ubiquitin, neurofilaments, TDP-43, etc.). Gallyas silver stain [7] can be used when a more detailed morphological study of tau-positive pathological inclusions is required (i.e., neurofibrillary tangles, argyrophylic grains, coiled bodies, etc.). Gallyas stain is particularly useful for the identification of small slender inclusions, like neuropil threads. However, it is important to

remember that both staining techniques are not fully comparable. Tau immunostaining reveals soluble non-argyrophylic material in neurons ("pretangles") that remains negative for Gallyas. Additionally, Gallyas stain can still identify extraneuronal ghost tangles even after they have lost immunoreactivity to antibodies against hyperphosphorylated tau. As for the choice of a tau antibody, AT8 antibody is usually recommended, as it is highly reliable even in archival samples with a very long fixation time.

1.2 Monitoring Alterations in Dendritic Spines

The dendritic spines on pyramidal cells represent the vast majority of postsynaptic elements of cortical synapses and they are fundamental structures in memory, learning and cognition. Since Alzheimer's disease (AD) has been proposed to be a synaptopathy [8] and the vast majority of degenerating neurons are pyramidal neurons, alterations of dendritic spines represent a major target of study in AD. It has been proposed that disease progression at the cellular level can be tracked via tau phosphorylation sites, first at the site recognized by 'anti-paired helical filaments (PHF) antibody-tauAT8' (antibody PHF-tauAT8), and then at the site recognized by antibody PHF-tauPHF-1. Nevertheless, how this degeneration is triggered and progresses remains unknown [9]. Therefore, the presence of phosphotau is likely to be involved in the changes to dendritic spines and the loss of synapses that leads to cognitive decline in AD. We examined human cortical pyramidal cells with either diffuse phosphotau in a putative pre-tangle state or aggregated tau that forms intraneuronal neurofibrillary tangles in order to analyse possible alterations to their dendritic spines [10]. For this purpose, intracellular injections of Lucifer Yellow (LY) in fixed cortical tissue were used, followed by double immunostaining of the sections with anti-Lucifer yellow and either anti-PHF-tauAT8 or anti-PHF-tauPHF-1.

1.3 Characterization of Phosphorylated Tau in Detergent-Insoluble Protein Aggregates

A way to complement the tau neuropathological staging is to characterize tau aggregates and the tau phosphorylation state from various brain regions from patients with different Braak stages. Frozen tissue is used for characterizing phosphorylated tau by Western blot and for the analysis of detergent-insoluble tau aggregates. Filamentous tau aggregates first appear at the temporal lobe and from there, the pathology spreads to frontal and parietal lobes.

2 Materials

Prepare all solutions with Milli-Q (Millipore) ultrapure water or distilled water as indicated, and analytical grade reagents. Unless stated otherwise, prepare and store all reagents at room temperature (RT). Diligently follow all waste disposal regulations when disposing waste materials.

2.1 Brain Tissue Fixation for Preparation of Sections

- 0.1 M phosphate buffer (PB), pH 7.4: Weigh 10.9 g of Na_2HPO_4 and 3.2 g of NaH_2PO_4. Make up to 1 L with distilled water.
- 4 % paraformaldehyde (PF) phosphate-buffered in PB: 40 g of paraformaldehyde in 1 L of 0.1 M phosphate buffer (PB).
- Vibrating microtome (Vibratome) instrument to produce tissue sections.

2.2 Selection of Tissue Samples for Tau Neuropathological Staging

The procedure is performed as part of the routine neuropathological work-up of *post-mortem* brains, particularly if obtained from aged and/or dementia patients. Samples are obtained from conventional coronal slices of one brain hemisphere (fixed hemisphere in brain bank protocols).
Brain regions used in staging include:

- Entorhinal-perirhinal cortex, medial to the rhinal sulcus, at the coronal level of the amygdala.
- Hippocampal cortex (CA1 sector) at the coronal level of the lateral geniculate nucleus.
- Inferior lateral cortex, lateral to the collateral sulcus.
- Middle frontal gyrus (as defined by CERAD).
- Superior and middle temporal gyri (as defined by CERAD).
- Inferior parietal lobule (as defined by CERAD).
- Occipital cortex (BA 17 and 18).

All cortical regions represented in tissue samples recommended by CERAD guidelines correspond to multimodal associative cortical areas, and are used primarily for the assessment of neuritic plaque density. However, as a staining technique for neurofibrillary degeneration has to be performed in order to identify neuritic plaques (amyloid plaques with dystrophic neurites), this staining can be also used for the assessment of NFTs and NT in these areas (*see* **Notes 1** and **2**).

2.3 AT8 Immunohistochemistry. Phospho-Tau Immunostaining (See Notes 3 and 4)

- 0.1 M sodium citrate pH 6.0 solution.
- 0.3 % hydrogen peroxide (H_2O_2).
- Potassium phosphate buffered saline 0.1 M, pH 7.4 (KPBS).
- Primary antibody: Phospho-PHF-tau pSer202/Thr205 Antibody (AT8) (Thermo Scientific, USA): 1:2,000.
- Antibody diluent.
- Secondary antibody solution (biotinylated secondary anti-mouse antibody solution): 1 μL of secondary antibody in 1 mL of KPBS.
- Avidin-biotin-complex (ABC) reagent (Vectastain, Vector Labs): Add 4.5 μL of A solution and 4.5 μL of B solution per 1 mL of KPBS (*see* **Note 3**).

- Diaminobenzidine (DAB) substrate solution.
- Carazzi's hematoxylin stain.
- Pressure cooker.

2.4 Gallyas Silver Staining (See Notes 5–9)

- 5 % periodic acid solution: Dissolve 10 g of periodic acid in 200 mL of distilled water.
- Alkaline silver iodide solution: Dissolve 8 g of sodium hydroxide and 20 g of potassium iodide in 100 mL of distilled water. Add 7 mL of aqueous solution of silver nitrate. Add distilled water up to a final volume of 200 mL.
- 0.5 % acetic acid: 1 mL of acetic acid in 200 mL of distilled water.
- Stock solution I (*see* **Note 8**): Dissolve 10 g of anhydrous sodium carbonate in 200 mL of distilled water.
- Stock solution II (*see* **Note 8**): Dissolve each reagent consecutively (in series); wait for complete dissolution before adding the next ingredient: 0.4 g of ammonium nitrate, 0.4 g of silver nitrate and 2 g of tungstosilicic acid in 200 mL of distilled water.
- Stock solution III (*see* **Note 8**): Dissolve each reagent consecutively (in series); wait for complete dissolution before adding the next ingredient: 0.2 g of ammonium nitrate, 0.2 g of silver nitrate, 1 g of tungstosilicic acid and 730 μL of 35–45 % formaldehyde in 100 mL of distilled water. Stock solutions are stable and can be stored in dark bottles.
- 0.2 % gold chloride solution: Dissolve 0.4 g of gold chloride in 200 mL of distilled water.
- 1 % sodium thiosulfate solution: Dissolve 2 g of sodium thiosulfate in 200 mL of distilled water.
- 0.1 % nuclear fast red 2.5 % aluminium sulphate.
- Incubator oven at 37 °C (*see* **Note 9**).

2.5 Reagents and Materials for Dendritic Spines and Biochemistry Analysis

- Sodium phosphate buffer 0.1 M, pH 7.4 (PB).
- Paraformaldehyde: 4 % in PB (PF).
- 4,6-Diamidino-2-phenylindole (DAPI) 10^{-5} M.
- 0.1 M Tris–HCl buffer, pH 7.4.
- Lucifer Yellow (LY): 8 % LY in Tris–HCl buffer.
- Stock solution: 2 % bovine serum albumin (BSA), 1 % Triton X-100, 5 % sucrose in PB.

2.5.1 Solutions

2.5.2 Primary Antibodies

- Tau antibody 7.51 is a kind gift of Dr. C. M. Wischik (University of Aberdeen, UK) and recognizes a region included within the microtubule-binding domain of human and murine tau.

- Mouse anti-human paired helical filaments (PHF)-tau monoclonal antibody (clone AT8, PHF-tauAT8; MN1020; Thermo Scientific) which recognizes tau phosphorylated at Ser202/Thr205. 1:2,000 in stock solution.
- Mouse PHF-1 monoclonal antibody (PHF-tauPHF-1, kindly supplied by Dr P. Davies, A. Einstein University, New York), which recognizes tau phosphorylated at Ser396/404. 1:100 in stock solution.
- Rabbit antibody against Lucifer Yellow (LY) produced at the Cajal Institute (Madrid, Spain) [10]. 1:400,000 in stock solution.

2.5.3 Secondary Antibodies

- Biotinylated donkey anti-rabbit secondary antibody (Amersham Pharmacia Biotech). 1:200 in stock solution.
- Alexa fluor 594 anti-mouse and streptavidin coupled to Alexa fluor 488 (Molecular Probes). 1:1,000 in stock solution.
- Autofluorescence Eliminator Reagent.
- Antifade Reagent mounting medium.

2.5.4 Software

- 3D Reconstruction software (e.g. Imaris 7.1. Bitplane AG, Zurich, Switzerland).

2.6 Reagents and Materials for Characterization of Tau Aggregates and Tau Phosphorylation State

- Brain extracts homogenization buffer (1:10 w/v) consisting of: 10 mM Tris–HCl, pH 7.4, 1 mM EGTA, 0.8 M NaCl and 10 % sucrose plus phosphatase inhibitors (10 mM NaF, 1 mM sodium orthovanadate) and protease inhibitors (2 mM phenylmethanesulfonylfluoride (PMSF), 10 μg/mL aprotinin, 10 μg/mL leupeptin, 10 μg/mL pepstatin). PMSF is very unstable and must be added just prior to use.
- 2 % uranyl acetate.
- Extraction buffer for samples homogenization and Western blot analysis consisting of: 20 mM HEPES-NaOH, pH 7.4, 100 mM NaCl, 10 mM NaF, 1 % Triton X-100, 1 mM sodium orthovanadate, 10 mM EDTA and protease inhibitors (2 mM PMSF, 10 μg/ml aprotinin, 10 μg/ml leupeptin and 10 μg/ml pepstatin). PMSF is very unstable and must be added just prior to use.
- Bradford protein assay.
- SDS polyacrylamide (SDS-PAGE) material and reagents.
- Primary antibodies 7.51, AT8, and PHF-1. The tau antibody 7.51 is a kind gift of Dr. C. M. Wischik (University of Aberdeen, UK) and recognizes a region included within the microtubule-binding domain of human and murine tau. The (PHF)-tau monoclonal antibody (clone AT8, PHF-tauAT8; MN1020; Thermo Scientific) recognizes tau phosphorylated

at Ser202/Thr205. The mouse PHF-1 monoclonal antibody (PHF-tauPHF-1), kindly supplied by Dr P. Davies (A. Einstein University, New York) recognizes tau phosphorylated at Ser396/404.

- Secondary antibody. Goat anti-mouse antibody (GIBCO).
- 5 % nonfat dried milk.
- Electrogenerated chemiluminescence (ECL) immunodetection reagents.
- Transmission electron microscope.
- Electron microscopy carbon-coated grids and materials.
- Eikonix IEEE-488 image scanner densitometer.
- DigitalMicrograph 2.1 software (Gatan, Pleasanton, CA).

3 Methods

3.1 Brain Tissue Samples. Fixation and Preparation of Sections

Brain tissue is fixed in 4 % phosphate-buffered formaldehyde (PF) for at least 3 weeks. For dendritic spine analysis (method below), brain samples (containing hippocampus and adjacent cortex) are immediately fixed in cold 4 % PF in PB and cut into small blocks ($10 \times 10 \times 10$ mm). The blocks are postfixed in 4 % PF in PB for 24 h at 4 °C. Vibratome sections of the tissue are then obtained and intracellular injections performed.

3.2 Immunochemistry. Phospho-Tau Immunostaining (See Notes 3, 4, and 10–12)

1. Deparaffinize and hydrate tissue sections.
2. Perform antigen unmasking in pressure cooker, placing sections in 0.1 M sodium citrate pH 6.0.
3. Open the pressure cooker and let the sections temper in the sodium citrate solution for 20 min.
4. Quench endogenous peroxidase by incubating sections for 30 min in 0.3 % H_2O_2.
5. Rinse sections twice with KPBS, 5 min each.
6. Incubate sections with AT8 anti-tau monoclonal antibody in 1/100 dilution in antibody diluent. Place sections in a moisture chamber at room temperature.
7. Rinse sections twice with KPBS, 5 min each.
8. Incubate sections for 30 min with diluted biotinylated secondary antibody.
9. Rinse sections twice with KPBS, 5 min each.
10. Incubate sections with ABC staining reagent during 60 min in a moisture chamber at room temperature (*see* **Note 3**).
11. Rinse sections twice with KPBS, 5 min each (*see* **Note 4**).

12. Incubate sections in DAB until desired stain intensity develops.
13. Rinse sections in tap water.
14. Counterstain sections with Carazzi's hematoxylin, 2 min.
15. Rinse sections in tap water, clear and mount.

3.3 Gallyas Silver Staining (See Notes 5–9, and 13–16)

1. Deparaffinize and hydrate tissue sections.
2. Place sections in 5 % periodic acid, 5 min.
3. Rinse in distilled water, 5 min.
4. Rinse in distilled water, 5 min.
5. During this time, prepare stock solution (developer): Add consecutively 15 mL of stock solution II, 50 mL of stock solution I, and 35 mL of stock solution III. Check that the final solution is crystal clear. If slightly cloudy, discard it (*see* **Note 13**).
6. Place sections in alkaline solution of silver iodide, 1 min.
7. Place sections in 0.5 % acetic acid, 10 min. During this phase, after 3 min, place stock solution (developer) in the oven at 37 °C (*see* **Note 14**).
8. Place sections in stock solution (developer) in the oven at 37 °C. Total time in stock solution may reach 8–10 min. Sections should be checked after 8 min for the appearance of a dark silver shade.
9. Place sections in 0.5 % acetic acid during 3 min (to stop developer).
10. Rinse with distilled water, 5 min.
11. Stabilize with 0.2 % gold chloride during 5 min (*see* **Note 15**).
12. Rinse with distilled water, 5 min.
13. Fix in 1 % sodium thiosulfate, 1 min.
14. Rinse with distilled water, 5 min.
15. Counterstain with nuclear fast red, 1 min.
16. Rinse with distilled water.
17. Dehydrate and cover slip.

3.4 Monitoring Alterations in Dendritic Spines (See Notes 17–20)

Small blocks of the hippocampal and adjacent cortex (10×10×10 mm) are post-fixed in 4 % PF in PB for 24 h at 4 °C. Vibratome sections of the tissue are then obtained and intracellular injections performed. Briefly, individual pyramidal cells in the hippocampal formation and adjacent cortex are injected intracellularly with Lucifer Yellow (below).

3.4.1 Intracellular Injections (See **Note 17**)

Coronal sections (250 μm) are cut with a vibratome and the slices incubated for 10 min in 4,6-diamidino-2-phenylindole (DAPI). These DAPI pre-labeled sections are mounted in the injection

chamber and the nuclei of cells are visualized by UV excitation [10]. Pyramidal cells are then injected individually with Lucifer Yellow by hyperpolarizing current in the cytoarchitectonically identified hippocampal formation and the adjacent cortex (including the entorhinal, EC and parahippocampal cortex, PHC). The sections immediately adjacent were Nissl-stained (50 μm) in order to identify the cortical areas and the laminar boundaries [10].

3.4.2 Reconstruction and Morphometric Analysis of Pyramidal Neurons Labeled with LY

Sections are imaged with a confocal scanning laser microscope attached to a fluorescence microscope. Image stacks of 10–100 image planes (voxel size: 0.057×0.057×0.28 μm; area: 58.36×58.36 μm) obtained with a 63× oil-immersion lens (NA, 1.40; refraction index 1.45) using a calculated optimal zoom factor of 2.3. After acquisition, the stacks are analysed with three-dimensional (3D) image processing software, Imaris 7.1 (*see* **Notes 18** and **19** and ref. 10).

3.4.3 Combination of Iintracellular Injection with Immunohistochemistry and Histochemistry

Following intracellular injection of pyramidal neurons with LY, the sections are processed with a rabbit antibody against LY (1:400,000 in stock solution) and then with the anti-PHF-tauAT8 or anti-PHF-tauPHF-1. Antibody binding is detected with a biotinylated donkey anti-rabbit secondary antibody (1:200 in stock solution), followed by a mixture of Alexa fluor 594 anti-mouse (1:1,000) and streptavidin coupled to Alexa fluor 488 (1:1,000). Thereafter, the sections are washed in PB and treated with Autofluorescence Eliminator Reagent to reduce lipofuscin-like autofluorescence without adversely affecting any other fluorescence labeling in the sections. The sections are then washed and mounted with ProLong Gold Antifade Reagent mounting medium. The stacks containing images of intracellular injections (green) and PHF-tauAT8 or PHF-tauPHF-1 staining (red) are opened with Imaris 7.1 software. The red channel is hidden and the stacks are coded (codes are not broken until the quantitative analysis had been completed). Thereafter, two main types of LY injected neurons can be distinguished: neurons immunostained for PHF-tauAT8 or PHF-tauPHF-1 and neurons that immunostained by either of the two antibodies (*see* **Note 20**).

The density of the dendritic spines is established as the number of dendritic spines found in 10 μm segments along the length of the dendrite. The spine volume is estimated using a reference method [11]. Briefly, 7–10 different intensity threshold surfaces are created for each stack of images, and the solid surface that exactly matched the contour of each dendritic spine is then selected. Each dendrite is rotated in 3D and examined to ensure that the solid surface selected for each dendritic spine is appropriate. The length of dendritic spines is measured individually, from the point of insertion at the dendritic shaft to the distal tip of the spine while rotating the image in 3D [11]. Additional methodological considerations are presented and discussed in Merino-Serrais et al. [10].

3.5 Characterization of Tau Aggregates and Tau Phosphorylation State

Brain cell extracts isolated by homogenizing the brain tissue in cold (ice) homogenization buffer including phosphatase and protease inhibitors (*see* Subheading 2).

3.5.1 Isolation of Brain Cell Extracts

3.5.2 Isolation of Detergent-Insoluble Tau Aggregates

Tau aggregates from brains isolated by reference methods as described for paired helical filaments-associated tau (PHF-associated tau) from Alzheimer disease patients [12, 13]. Briefly, PHF populations extracted from AD brain homogenates with about 50 % of PHF immunoreactivity can be obtained in supernatants following homogenization in buffers containing NaCl. These can be further enriched for PHFs by taking advantage of their insolubility in the presence of detergents and 2-mercaptoethanol, removal of aggregates by filtration and sucrose density centrifugation [12]. Three different sources were used to look for tau aggregates: temporal, frontal and parietal cortex.

3.5.3 Visualization of Insoluble Tau Aggregates by Immunoelectron Microscopy

Detergent-insoluble tau aggregates can be visualized by electron microscopy [14]. Immunoelectron microscopy is performed after adsorption of the samples to electron microscopy carbon-coated grids and incubation with the first antibody (1:100) for 1 h at room temperature. After extensive washing, the grids are incubated with the secondary antibody (1:40) conjugated with 10 nm diameter gold particles. Finally, the samples are stained with 2 % uranyl acetate for 1 min. Transmission electron microscopy can then be performed, e.g. in a JEOL model 1200EX electron microscope operated at 100 kV.

3.5.4 Transmission Electron Microscopy

Microscopy grids with or without uranyl acetate contrasting can be observed in a transmission electron microscopy JEOL 1200 EXII operated at 120 kV. Electron micrographs were obtained at a magnification of ×50,000 on Kodak SO-163 film developed with D19 developer at full strength for 12 min. For measuring purposes, micrographs were digitized using an Eikonix IEEE-488 image scanner densitometer and processed using the DigitalMicrograph 2.1 software (Gatan, Pleasanton, CA).

3.5.5 Western Blot Analysis

Extracts for Western blot analysis are prepared by homogenizing the tissues (or the tau aggregates sample) in ice-cold extraction buffer. The samples are homogenized at 4 °C and protein content determined by Bradford protein assay. Total protein (10 μg) is electrophoresed on 8 % SDS-PAGE gel and transferred to a nitrocellulose membrane. The experiments were performed using the primary antibodies: 7.51, AT8 and PHF-1. The filters are incubated with the antibody at 4 °C overnight in 5 % nonfat dried milk.

A secondary goat anti-mouse (1:5,000) followed by ECL detection reagents was used for immunodetection.

3.6 Conclusions

Two main features have been linked to tau pathology in Alzheimer's disease, tau phosphorylation and tau aggregation. These two modifications have been used as markers to analyze the spreading of the disease that correlates well with that of tau pathology. Thus, the progression of the disease could be followed by looking at the following examples of samples analyzed by Western blot using an antibody raised against tau that reacts with the protein from the aggregates and by electron microscopy of samples obtained from various brain regions, as indicated in Fig. 1.

No significant changes are observed in the microanatomy of dendrites from pyramidal neurons at early stages of neurofibrillary pathology, as defined by the presence of diffuse phosphotau (neurons immunostained for PHF-tauAT8) in a putative pre-tangle state (Fig. 2b). However, once tau aggregates of NFTs had formed (neurons immunostained for PHF-tauAT8 or PHF-tauPHF-1), representing more advanced neurofibrillary alterations, significant reductions in the number, length and volume of spines are evident, suggesting alterations of axospinous synapses (Fig. 2c). The severity of these changes seems to be progressive, from the intermediate/advanced stages to extreme stage of the

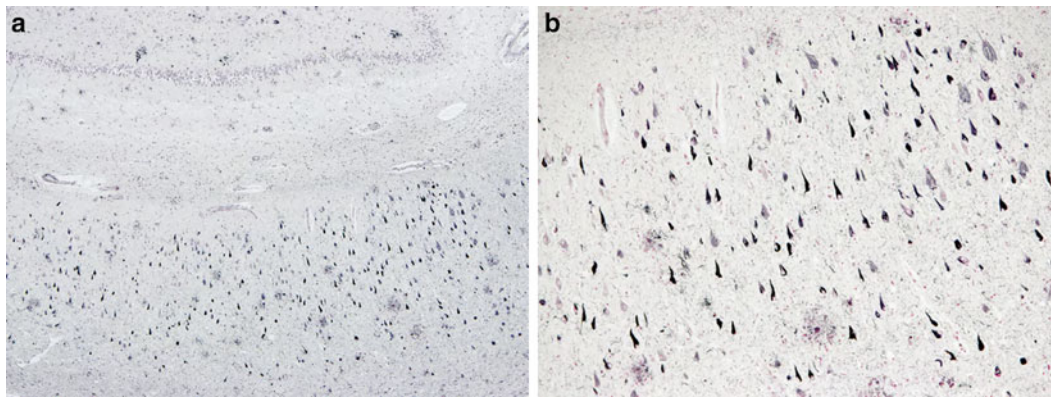


Fig. 1 Paraffin section of parietal cortex immunostained for AT8 anti-hyperphosphorylated tau antibody and Gallyas stained paraffin section of the hippocampus (CA1 sector). (**a**) Very low magnification of a paraffin section of parietal cortex immunostained for AT8 anti-hyperphosphorylated tau antibody. A high density of immunoreactive inclusions (mainly neuropil threads, neurofibrillary tangles and dystrophic neurites of neuritic plaques) stain diffusely the cortex with a band-like pattern at cortical laminae III and V. A Braak stage V–VI (isocortical) can be derived from the image. (**b**) Gallyas stained paraffin section of the hippocampus (CA1 sector) at medium magnification. Abundant flame-like neurofibrillary tangles can be observed occupying the neuronal body of pyramidal neurons. A dense background of neuropil threads is evident between neurons, and some neuritic plaques can also be identified. Note the nuclear counterstain provided by Nuclear Red. A Braak stage equal or superior to III can be established from the image

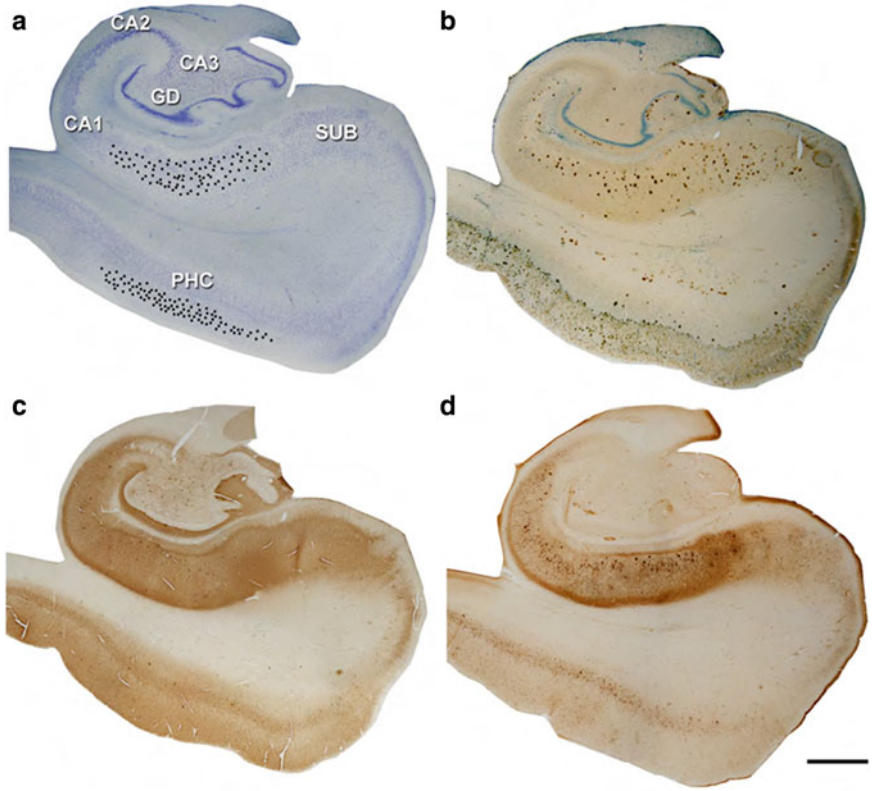
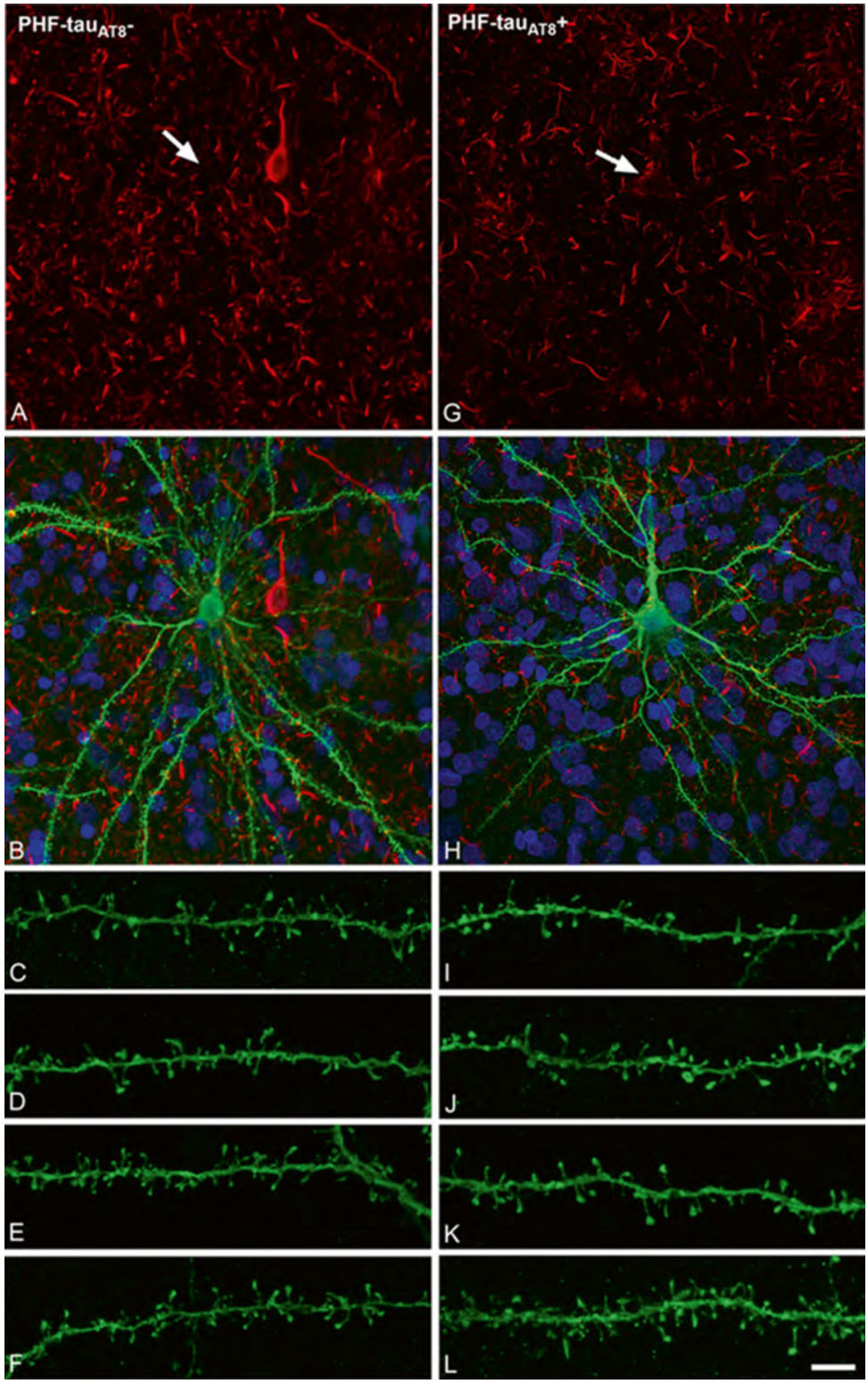


Fig. 2 Photomicrographs of Nissl, Aβ plaques/Nissl, PHF-tau$_{AT8}$ and PHF-tau$_{PHF-1}$ stained sections, neurons and dendrites in the parahippocampal cortex of a patient whose soma is free of PHF-tau$_{AT8}$-ir (PHF-tau$_{AT8}^{-}$; *A–F*) or that contains PHF-tau$_{AT8}$ in the putative pre-tangle state, and alterations of dendrites and dendritic spines in LY injected neurons with different stages of the neurofibrillar pathology. Reproduced from Merino-Serrais et al. [10] with permission (Open Access article distributed under the Creative Commons Attribution Non-Commercial License). (**a**) Low-power photomicrographs of Nissl, Aβ plaques/Nissl (using mouse anti-human beta-amyloid antibody [clone 6F/3D; Dako, Glostrup, Denmark]) (**b**), PHF-tau$_{AT8}$ (**c**) and PHF-tau$_{PHF-1}$ (**d**) stained sections from patient P9 (male, 82 year old, Neurofibrillar/Aβ pathology; Braak stage AD V/C). The *black dots* in *A* show the approximate location of the injected neurons in layer III of the parahippocampal cortex (PHC) and CA1 region. Scale bar (in *D*): 1,600 μm in *A–D*

Fig. 3 Neurons and dendrites in the PHC of a patient injected with LY whose soma is free of PHF-tau$_{AT8}$-ir (PHF-tau$_{AT8}^{-}$; A–F) or that contains PHF-tau$_{AT8}$ in the putative pre-tangle state (Pattern *I*; *G–L*). Stacks of 26 (*A*, *B*) and 28 (*G*, *H*) images, respectively, obtained after combining the channels acquired separately for DAPI (*blue*), LY (*green*) and PHF-tau$_{AT8}$ immunostaining (*red*). *C–L* Stacks of 26–32 confocal optical sections from basal dendrites of PHF-tau$_{AT8}^{-}$ (*C–F*) and immunostained (PHF-tau$_{AT8}^{+}$; *I–L*) LY-injected pyramidal neurons. Scale bar (in *L*): 13 μm in *A*, *B*, *G*, *H*; 2 μm in *C–F*, *I–L*


PHF-tau$_{AT8-}$
PHF-tau$_{AT8+}$
A
G
B
H
C
I
D
J
E
K
F
L

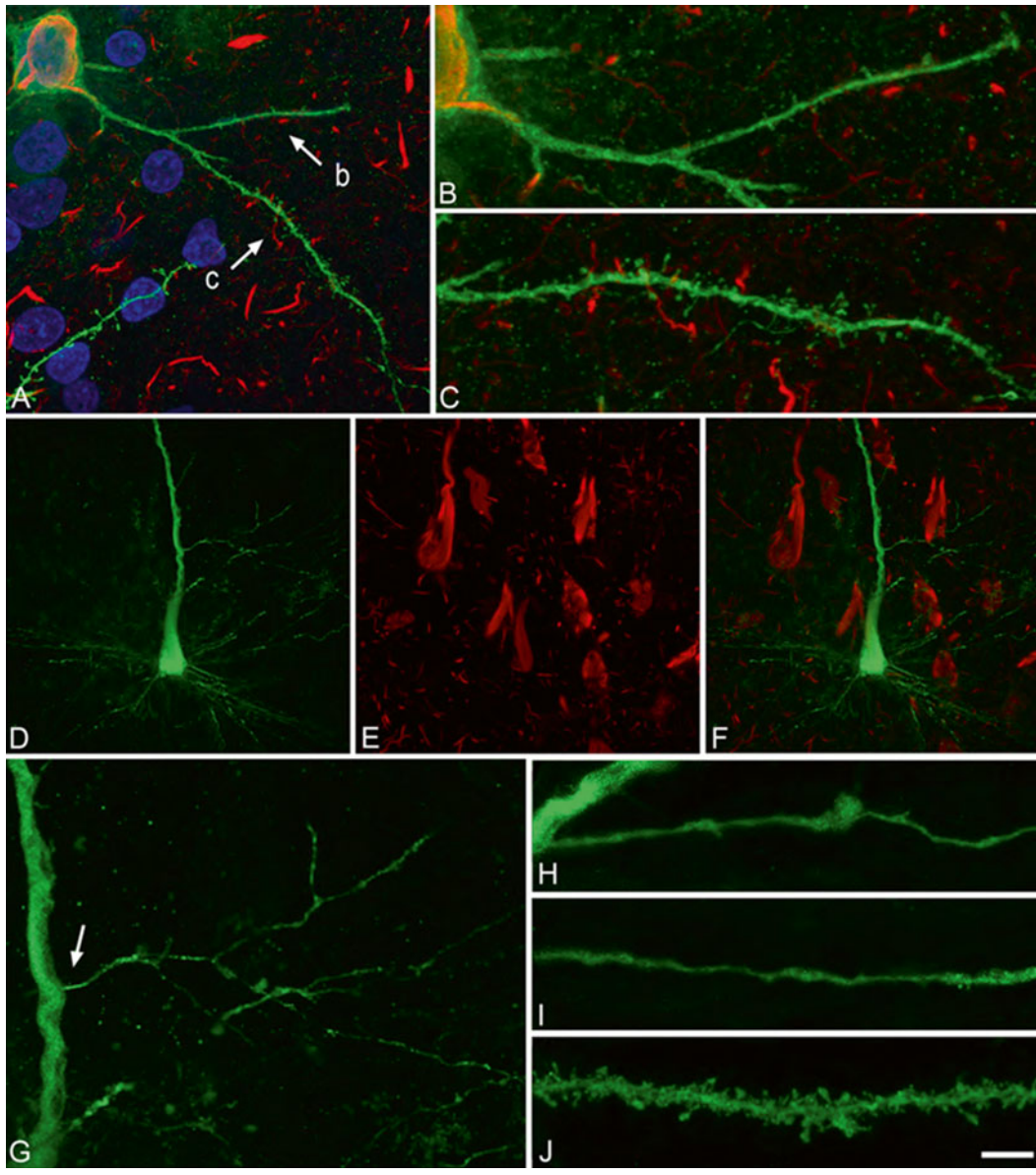


Fig. 4 Alterations of dendrites and dendritic spines in LY injected neurons with different stages of the neurofibrillar pathology. (*A*–*C*) PHF-tauAT8-ir neuron from layer III of the PHC of patient P9 showing an intermediate/advanced stage of the neurofibrillar pathology. (*D*–*L*) PHF-tau$_{PHF-1}$-ir neuron from CA1 of patient P12 (female 82 years old, Neurofibrillar/Aβ pathology; Braak stage and Lewy bodies) showing an extreme stage of the neurofibrillar pathology. *A*, Stack of 27 confocal optical sections obtained after combining the channels acquired separately for DAPI (*blue*), LY (*green*) and PHF-tau$_{AT8}$-ir (*red*), illustrating the cell body and proximal dendrites of the intracellular labelled neuron. *B* and *C*, Higher magnification of *A*, showing the dendrites indicated as b and c, respectively. Note the low density of dendritic spines in dendrite b compared to dendrite c. *D*–*E*, Stacks of 27 confocal optical sections showing the cell body and proximal dendrites of the intracellular labelled neuron (*D*) immunostained for PHF-tau$_{PHF-1}$ (*E*). *F*, image obtained by combining panels *D* and *E*. *G*, Higher magnification of *D*. *H*, *I* stacks of 38–55 confocal optical sections from a collateral apical dendrite (*arrow* in *G*) of the LY-injected pyramidal neuron, showing different segments of the same dendrite (*H*, proximal; *I*, distant). *J*, Stack of 26 confocal optical sections from the collateral apical dendrite of an intracellular labelled neuron that was adjacent to the LY-injected neuron shown in panel *D*, and that was not PHF-tau$_{PHF-1}$-ir. Note the lack of dendritic spines and the thin diameter of the dendrites of the PHF-tau$_{PHF-1}$-ir neuron (*H*, *I*) compared to the dendrite of the PHF-tau^{-} neuron (*J*). Scale bar (in *L*): 10 μm in *A*; 3.5 μm in *B*, *C*; 20 μm in *D*–*F*; 9 μm in *G*; 4.5 in *H*–*J*

neurofibrillary pathology. Thus, the characteristic cognitive impairment in AD is likely to depend on the relative number of neurons that have well-developed tangles. Since the microanatomical alterations could be correlated with changes in tau phosphorylation at specific sites, the methods described here are excellent tools to study the correlation between tau phosphorylation, synaptopathy and cognitive impairment in AD.

4 Notes

1. Staining protocols are here presented for paraffin sections. For both staining techniques, it is possible to use 5–15 μm mounted paraffin sections. Additionally both staining techniques, AT8 immunostaining and Gallyas stain, can be performed on mounted frozen sections and on floating cryostat or vibratome sections, using the same basic protocols and adjusting some staining times.
2. For both staining techniques glass slides should be pretreated for tissue adhesion enhancement, e.g., with poly-L-lysine.
3. Solution volumes for the preparation of the ABC reagent correspond to the Vectastain ABC kit (Vector Labs). There are other suppliers of ABC reagents that can be used with equally good results. Additionally, immunostaining may be enhanced by other methods of detection that result in a higher amplification of the signal and shorter incubation protocols (e.g. polymer detection reagents).
4. The last KPBS rinsing step before DAB color development can be substituted by Tris–HCl buffered saline (TBS) rinse buffer or 0.1 M sodium acetate pH 6.0 rinsing solution. That results in a more thorough elimination of previous reagents.
5. The volume of solutions here presented for Gallyas staining are adequate for use in a Hellendahl jar.
6. All solutions should be stocked in dark tinted bottles.
7. For preparation of Gallyas stain, all glassware must be acid washed. Plastic forceps and gloves should be used when handling all reagents, and no metal instruments should be used in the protocol (staining racks, forceps, etc.). For preparation of solutions avoid the use of metal stir bars of magnetic stirrers.
8. Stock solutions should be prepared in the sequential order of the protocol: II, I and III.
9. The temperature of the oven for incubation in Gallyas stain should be strictly maintained at 37 °C. Even slightly higher temperatures may generate precipitates.
10. The final developer solution should be crystal clear. A cloudy tint indicates the formation of precipitates.

11. Within the oven, the developer solution should turn into a silvery-black colour before stopping the reaction.
12. Decoloration with gold chloride is usually very fast and can be controlled macroscopically or under the microscope (placing sections previously in tap water), particularly if sections turned too dark after incubation in the oven.
13. In Alzheimer's disease brain tissue hyperphosphorylated-tau immunostaining reveals all changes classically described as neurofibrillary degeneration: basically neurofibrillary tangles, neuropil threads and dystrophic neurites incorporated to neuritic plaques.
14. Other inclusions less frequently associated to Alzheimer's pathology can also be observed (e.g., tau-immunoreactive astrocytes and oligodendrocytic coiled bodies), as well as tau-positive inclusions characteristic of tauopathies that may combine with Alzheimer's pathology (e.g. argyrophilic grains, Pick-bodies and different types of astrocytic inclusions).
15. Intensely positive sections can be evaluated even macroscopically due to widespread reactivity of tau inclusions (Fig. 5a).
16. Gallyas stain allows for a more distinct visualization of the morphology of tau-positive inclusions, particularly those formed by small or slender processes, and identifies a subpopulation of highly evolved extracellular tau-negative inclusions (Fig. 5b).
17. Neurons are injected until the individual dendrites of each cell could be traced to an abrupt end at their distal tips and the dendritic spines are readily visible, indicating that the dendrites are completely filled. Only cells identified as pyramidal neurons (recognized by the labeling of the apical dendrite) are included in the analysis (Fig. 2) [10].
18. Horizontally projecting basal dendrites are randomly selected, each one originated from a different pyramidal neuron and acquired at high magnification (Leyca glycerol objective, 63× magnification) to capture the full dendritic depth, length, and width of dendrites [10].
19. For each stack of images, the laser intensity and detector sensitivity are set so that the fluorescence signal from the dendritic spines occupies the full dynamic range of the detector. Therefore, some pixels can be saturated in the dendritic shaft, but no pixels should be saturated within the dendritic spines [10].
20. After acquisition, the stacks of images are opened with 3D image processing software and the red channel (containing PHF-tauAT8 staining) is hidden. Additionally, the stacks are coded (codes not to be broken until the quantitative analysis is completed) [10].

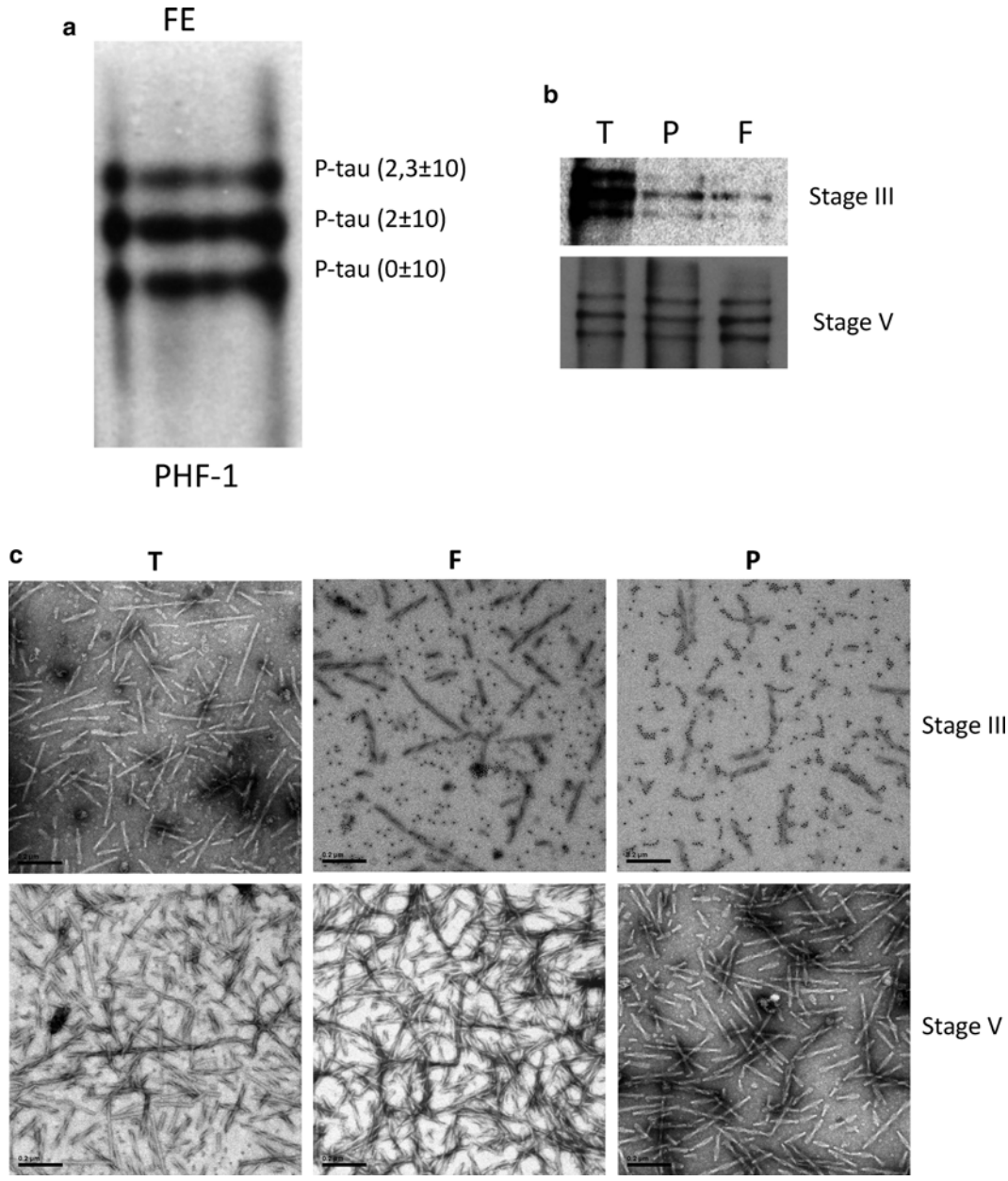


Fig. 5 Hyperphosphorylated tau proteins and Western blot of brain samples. (**a**) Scheme of the nature of the hyperphosphorylated tau proteins observed by Western blot. (**b**) Western blot of brain samples (temporal, frontal and parietal cortex) from three patients (Braak stages III and V) are shown. (**c**) The polymers from those detergent-insoluble tau aggregates obtained from previous patients are shown

References

1. Montine TJ, Phelps CH, Beach TG et al (2012) National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. Acta Neuropathol 123:1–11
2. Thal DR, Rub U, Orantes M, Braak H (2002) Phases of Abeta-deposition in the human brain and its relevance for the development of AD. Neurology 58:1791–1800
3. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82:239–259
4. Mirra SS, Heyman A, McKeel D et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41:479–486
5. Braak H, Alafuzoff I, Arzberger T et al (2006) Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol 112: 389–404
6. Arriagada PV, Growdon JH, Hedley-Whyte ET, Hyman BT (1992) Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. Neurology 42:631–639
7. Gallyas F (1971) Silver staining of Alzheimer's neurofibrillary changes by means of physical development. Acta Morphol Acad Sci Hung 19:1–8
8. Terry RD, Masliah E, Salmon DP et al (1991) Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. Ann Neurol 30:572–580
9. Avila J, Leon-Espinosa G, Garcia E et al (2012) Tau phosphorylation by GSK3 in different conditions. Int J Alzheimers Dis 2012:578373
10. Merino-Serrais P, Benavides-Piccione R, Blazquez-Llorca L et al (2013) The influence of phosphotau on dendritic spines of cortical pyramidal neurons in Alzheimer's disease patients. Brain 136:1913–1928
11. Benavides-Piccione R, Fernaud-Espinosa I, Robles V et al (2012) Age-based comparison of human dendritic spine structure using complete three-dimensional reconstructions. Cereb Cortex 23:1798–1810
12. Greenberg SG, Davies P (1990) A preparation of Alzheimer paired helical filaments that displays distinct tau proteins by polyacrylamide gel electrophoresis. Proc Natl Acad Sci U S A 87:5827–5831
13. Pérez M, Valpuesta JM, de Garcini EM et al (1998) Ferritin is associated with the aberrant tau filaments present in progressive supranuclear palsy. Am J Pathol 152:1531–1539
14. Engel T, Lucas JJ, Gómez-Ramos P et al (2006) Cooexpression of FTDP-17 tau and GSK-3beta in transgenic mice induce tau polymerization and neurodegeneration. Neurobiol Aging 27:1258–1268

Chapter 8

LC3-II Tagging and Western Blotting for Monitoring Autophagic Activity in Mammalian Cells

Anne Streeter, Fiona M. Menzies, and David C. Rubinsztein

Abstract

The autophagosome-associated protein LC3-II is commonly used as a marker of autophagic activity within cells, but its levels are affected by both formation and degradation of autophagosomes. This can make the significance of altered LC3-II levels ambiguous. Here we describe the method of Bafilomycin A_1 blotting, in which the degradation of autophagosomes is prevented in cultured cells, allowing the causes of altered LC3-II levels to be determined.

Key words Autophagy, LC3, Assay, Flux, Bafilomycin A1

1 Introduction

The process of macroautophagy, whereby a portion of cytosol is engulfed into double-membraned autophagosomes and transported to the lysosome for degradation [1], has potential as a therapy for delaying the onset of neurodegenerative diseases. By clearing aggregate-prone proteins from the cytosol, the formation of aggregates within cells can be reduced [2]. This has been demonstrated in experimental models involving mutated huntingtin [3], alpha synuclein [4] and tau [5].

As a specific marker of autophagosomes, LC3-II is a useful and versatile tool in accurate measurement of autophagic flux. LC3 (microtubule-associated protein 1 light chain 3, MAP1-LC3) is the mammalian homologue of the yeast autophagy protein Atg8 [6]. The C-terminus of LC3 is cleaved by the action of Atg4 to give the LC3-I form of LC3, which is then conjugated to the lipid phosphatidylethanolamine to give the membrane-associated LC3-II form [7]. This is found on the inner and outer membranes of autophagosomes. The pool associated with the inner membrane is degraded in the autolysosomes (formed by the fusion of autophagosomes and lysosomes), while that associated with the outer

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_8,

membrane can be cleaved by Atg4 and recycled. LC3-II levels in a cell therefore give an indication of the number of autophagosomes in the cell at that particular time and hence the autophagy state of the cell.

The level of LC3-II in cells, however, needs to be interpreted with care, as it is affected by both formation and degradation processes. An increase in LC3-II levels can be due to increased formation or decreased degradation, while decreased LC3-II could be due to increased degradation or decreased formation. In order to distinguish between these criteria, Bafilomycin A_1 blots of LC3-II levels are used. Bafilomycin A_1 is a drug which inhibits the V-ATPase responsible for acidification of the lysosome, and prevents fusion of autophagosomes and lysosomes [8, 9]. Comparison of LC3-II levels in the absence and presence of Bafilomycin A_1 (or other inhibitors of LC3-II degradation) allows the effects of formation and degradation to be uncoupled [10, 11].

As well as studying the LC3-II levels of cells, it is useful to look at clearance of autophagy substrates such as p62 in order to gain a more comprehensive idea of the autophagy state of the cells. This can be done by Western blotting, though similarly to LC3-II levels, the effect of formation of new p62 (i.e. translation) should be taken into consideration when interpreting the results.

Here, the procedures for carrying out Bafilomycin A_1 blots to determine the autophagic activity of mammalian cells are detailed. Cells in culture are treated with Bafilomycin A_1 and harvested, and the LC3-II levels of the lysates analysed by Western blotting. Interpretation of the blots will be explained.

2 Materials

Make up reagents in distilled water unless otherwise stated.

1. HeLa cells in culture (*see* **Notes 1** and **2**).
2. Treatments being investigated (drugs, culture conditions etc) (*see* **Note 3**).
3. 100 μM Bafilomycin A_1 stock solution in DMSO (*see* **Note 4**).
4. Phosphate-buffered saline (PBS) buffer: 138 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4 at pH 7.4 (*see* **Note 5**).
5. Radioimmunoprecipitation assay (RIPA) buffer: 150 nM NaCl, 1 % NP40, 0.5 % sodium deoxycholate (NaDoC), 0.1 % SDS, 50 mM Tris pH 7.4, 1× protease inhibitor cocktail (Roche) (*see* **Note 6**). Make up fresh each time and use on ice.
6. Standard protein assay kit (e.g. Bio-Rad DC™ Protein Assay).
7. 100 mg/mL BSA solution.

8. 2× Laemmli buffer: 65 mM Tris–HCl pH 6.8, 25 % (w/v) glycerol, 2 % SDS, 0.01 % (w/v) bromophenol blue, 5 % (v/v) β-mercaptoethanol.
9. 30 % Acrylamide/bis-acrylamide solution (37.5:1).
10. Resolving gel buffer: 1.5 M Tris pH 8.8.
11. 10 % SDS solution.
12. 10 % ammonium persulfate solution in H_2O.
13. Stacking gel buffer: 1 M Tris pH 6.8.
14. TEMED (*N*, *N*, *N′*, *N′*—tetramethylethylenediamine).
15. Isopropanol.
16. Pre-stained molecular weight markers (e.g. Invitrogen SeeBlue® Plus2 Pre-Stained Standard).
17. Gel running buffer: 25 mM Tris pH 8.3, 0.192 M glycine, 0.1 % SDS (*see* **Note 7**).
18. Wet transfer buffer: 25 mM Tris pH 8.3, 0.192 M glycine, 20 % methanol (*see* **Notes 7** and **8**).
19. Ponceau S stain solution: 5 % (w/v) in 5 % acetic acid.
20. Milk: 5 % (w/v) milk powder in PBS.
21. PBS-Tween: 0.1 % Tween-20 in PBS.
22. Novus Biologicals rabbit anti-LC3 primary antibody (NB100-2220) (*see* **Note 9**).
23. Sigma rabbit anti-actin primary antibody (A2066) (*see* **Note 9**).
24. IR dye-conjugated anti-rabbit secondary antibody (*see* **Note 10**).
25. Western blot and wet transfer equipment (e.g. PVDF membranes).
26. Licor Odyssey equipment or equivalent (*see* **Note 11**).

3 Methods

The instructions here test the effect of a hypothetical drug ("Drug A") on autophagy in HeLa cells following a 4 h treatment. For assistance on adapting the protocol for other experimental setups, relevant notes are referred to at the appropriate stages.

1. Seed HeLa cells, using approximately 2×10^5 cells per well of a 6-well plate (*see* **Note 12**) to ensure that they will be approaching confluence when harvested. Allow the cells to settle overnight.
2. Treat the cells with Drug A and 400 nM Bafilomycin A_1 for 4 h. This concentration of Bafilomycin A_1 treatment is saturating (*see* **Note 13**). Control conditions should include an appropriate volume of DMSO to control for the DMSO added

in the Bafilomycin A_1 stock solution. The conditions required are: no Drug A; Drug A; Bafilomycin A_1 and no Drug A; Bafilomycin A_1 and Drug A.

3. Harvest the cells. Remove the media from the cells and wash once with PBS (allow approximately 1.5 mL of PBS per well in a six-well plate). Pipette 100 μL of ice-cold RIPA buffer into each well (*see* **Note 14**), and use a cell scraper to lift cells from the bottom of the well. Pipette the lysed cells into labelled Eppendorf tubes and incubate on ice for 10 min. Centrifuge for 10 min at 13,000 × *g* and transfer the post-nuclear supernatant to fresh tubes on ice.
4. Carry out a protein assay. Make standards by making serial dilutions of BSA in the range 5–0 μg/μL in RIPA buffer. Make up Reagent A' by mixing 20 μL of Reagent S with 1 mL of Reagent A. In a 96-well plate, mix 1 μL of sample/standard, 20 μL of Reagent A', and 200 μL of Reagent B. Allow blue color to develop in a shaded place for 15 min. Scan the plate on a plate reader at a wavelength of 750 nm, using the standards to make a standard curve from which to determine the protein concentration in the samples.
5. Adjust the volumes of the samples using RIPA buffer to give the same protein concentration in all samples. Add 2× Laemmli buffer to each sample so that the protein concentrations remain equal between samples (*see* **Note 15**). Boil the tubes for 5 min on a hot block set to 100 °C (*see* **Note 16**). Use a bench-top centrifuge to briefly spin down condensation in the tubes. Samples may be frozen at –20 °C.
6. Mix 3.3 mL of distilled H_2O, 4 mL of 30 % acrylamide/bis acrylamide solution, 2.5 mL of resolving gel buffer, 100 μL of SDS and 100 μL of ammonium persulfate. Add 4 μL of TEMED, mix well, and pour the gel in a 1.5 mm × 10.1 cm × 7.3 cm cast. Leave space for a 1 cm depth of stacking gel (*see* **Note 17**), and pipette approximately 250 μL of isopropanol on top of the gel (*see* **Note 18**). Allow to set (around 15–20 min, depending on room temperature).
7. Mix 3.4 mL of distilled H_2O, 830 μL of 30 % acrylamide/bis acrylamide solution, 630 μL of stacking gel buffer, 50 μL of SDS and 50 μL of ammonium persulfate. Pour off the isopropanol from the top of the resolving gel, and rinse with distilled water. Use blotting paper to soak up residual water. Add 5 μL of TEMED to the stacking gel mixture and pour the gel immediately. Insert a 10-well comb to form the wells. Allow the gel to set for approximately 1 h (*see* **Note 19**).
8. Remove the comb from the gel, and set up the gel-running tank and fill with the appropriate volume of running buffer. Ensure the buffer is not leaking from the gel chamber. Load the gel with 20 μL of each of the samples (*see* **Note 20**) and 3 μL

of molecular weight markers. Run the gels slowly at 0.2 mA per gel, until the blue dye front is 3–5 mm from the end of the gel (approximately 95 min). Do not run further, as this may result in losing the small LC3 proteins from the gel.

9. Assemble a cassette for wet transfer of proteins to the polyvinylidene difluoride (PVDF) membrane (*see* **Note 21**). Soak the PVDF membrane in methanol for 1 min, rinse in H_2O, and then soak in transfer buffer before use. Working in a tray containing a shallow depth of transfer buffer, soak sponges and blotting paper in transfer buffer and place in a cassette. Prise apart the gel casting plates, and trim off the stacking gel. Carefully place the gel on the blotting paper, taking care not to stretch or damage the gel. Place the PVDF membrane on top of the gel, and then another layer of blotting paper. Ensure that no bubbles are trapped in the layers by smoothing the stack with a test tube. Add pre-soaked sponges and close the cassette. Ensure it is loaded into the frame with the membrane between the gel and the positive electrode. Keep the apparatus cool with an ice pack while transferring for 1 h at 90 V.
10. Disassemble the cassette and check for successful transfer of protein markers to the PVDF membrane. Stain the membrane with Ponceau S stain to confirm transfer of proteins to the membrane (*see* **Note 22**), agitating for 10 min at room temperature. Wash in water to develop the bands. Cut the PVDF membrane at around 30 kDa, using the markers as a guide. Block the membrane in milk, agitating for 1 h at room temperature (*see* **Note 23**).
11. Add 10 μL of anti-LC3 primary antibody to 10 mL of milk (giving a 1 in 1,000 dilution), and do the same for anti-actin antibody. Incubate the upper part of the membrane in anti-actin primary antibody, and the lower half in anti-LC3 primary antibody overnight, agitating at 4 °C. Primary antibody in milk can be frozen at –20 °C and reused.
12. Wash the membrane three times for 5 min each time in PBS-Tween. Make up secondary antibody against the primary antibodies used at a concentration of 1 in 3,000 (3 μL of secondary antibody in 9 mL of milk) (*see* **Note 24**). Incubate the membrane parts in secondary antibody for 1 h, agitating at room temperature. Wash the membrane three times for 5 min each time in PBS-Tween.
13. View membranes on the Licor Odyssey imager, placing them face down to get the best signal (*see* **Note 25**). Quantify the bands, and normalize the LC3-II signal using the actin bands as a loading control. The actin levels should be fairly even between the lanes, so this should only be a slight adjustment.

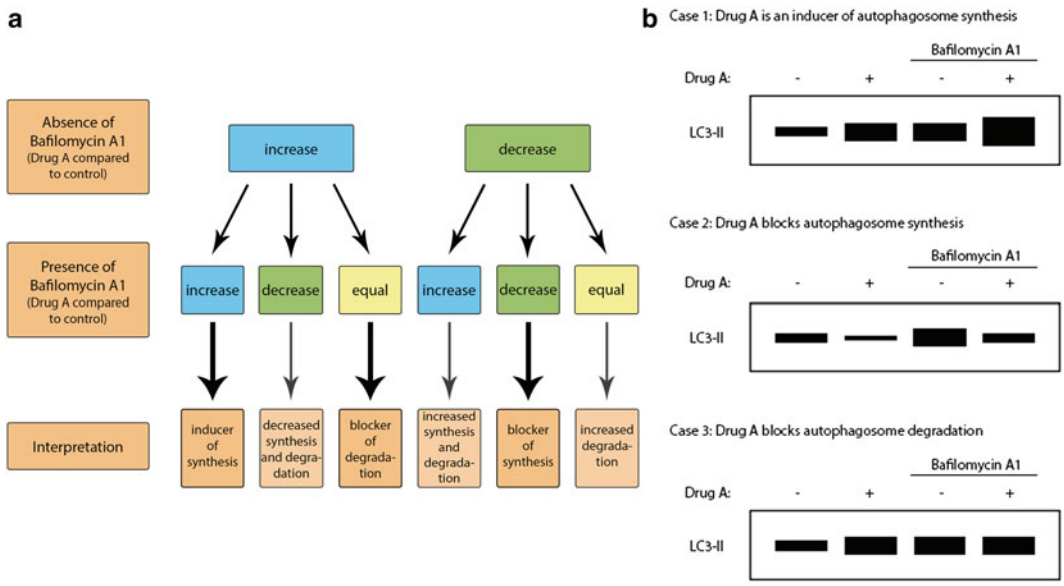


Fig. 1 Interpretation of Bafilomycin A_1 western blots. (**a**) Flow-chart indicating interpretations of LC3-II levels in blots. More common scenarios are indicated by *bold arrows*. (**b**) Schematics of LC3-II levels in Bafilomycin A_1 blots. Protein loading, as measured by actin, is assumed to be even across the lanes. Situations are shown in which hypothetical treatment Drug A is an inducer of autophagosome synthesis (Case 1), a blocker of autophagosome synthesis (Case 2), or a blocker of autophagosome degradation (Case 3)

14. Interpret the results of the blot (Fig. 1).
    - Compare cells treated with Drug A and untreated control cells to see the effect of Drug A on the size of the LC3-II pool in the cells.
    - Compare the cells treated with Drug A and Bafilomycin A_1 with control cells treated with Bafilomycin A_1 to see the contribution of LC3-II formation. If the LC3-II levels increase with Drug A in the presence of Bafilomycin A_1, then Drug A has increased formation of LC3-II (Case 1 in Fig. 1b). A decrease in LC3-II levels with Drug A in the presence of Bafilomycin A_1, together with a decrease with Drug A in the absence of Bafilomycin A_1, indicates that Drug A blocks autophagosome synthesis.
    - Compare LC3-II levels between cells treated with Drug A and cells treated with both Drug A and Bafilomycin A_1 to see the effect of Drug A on degradation of LC3-II. If Drug A blocks degradation of LC3-II, there will be no increase in LC3-II levels on treatment with Bafilomycin A_1, as Bafilomycin A_1 will not exert an additional effect on LC3-II levels (Case 3 in Fig. 1b). If Drug A does not affect degradation of LC3-II, there will be an increase in LC3-II levels when cells are treated with Bafilomycin A_1.

4 Notes

1. A wide variety of cells may be used for Bafilomycin A_1 blots. The method described here looks at the effect of a hypothetical drug ('Drug A') on autophagy in HeLa cells, but is easily adapted for other cell types and treatments.
2. Comparing LC3-II levels directly between cell types or lines is not advised, as levels of LC3-I and LC3-II vary widely between cell types, and the extent of the response to autophagy-modulating treatments can differ significantly. Therefore, we recommend that treated cells are compared to untreated controls of the same cell line.
3. The treatment of cells will be determined by the interests of the researcher. Bafilomycin A_1 blots are compatible with many types of treatment, but the incubation of cells with Bafilomycin A_1 will not necessarily occur for the entire length of time that the investigated treatment is being carried out.
4. Bafilomycin A_1 is made up into 100 μM stock solution by adding 1.61 mL of DMSO to 100 μg of powder (available from Enzo Sciences). This stock should be made into aliquots and stored in the dark at −20 °C.
5. For 1 L of PBS, use 8.1 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4 (final concentration 10 mM) and 0.24 g of KH_2PO_4 (final concentration 1.76 mM), at pH 7.4.
6. We use RIPA buffer for harvesting cells, but in theory other buffers can be used instead.
7. We have a 10× solution consisting of 250 mM Tris pH 8.3, 1.92 M glycine. To make up 1 L of running buffer, take 100 mL of 10× solution, add 10 mL of 10 % SDS and make up to 1 L with distilled water. When adding the water, run it down the side of the container to avoid excessive amounts of bubbles being formed by the SDS, or add the SDS last. To make up 1 L of transfer buffer, take 100 mL of 10× solution, and add 200 mL of methanol and 700 mL of distilled water.
8. Methanol in the transfer buffer improves the transfer of small proteins to the PVDF membrane by giving a charged surface the protein can bind to. It slightly fixes the gel however, which can limit the transfer of large proteins. As LC3-I and LC3-II are small (running at 18 kDa and 16 kDa, respectively), methanol transfer buffer is used in this protocol.
9. Other antibodies can be used, but the concentrations used here are specific to the named antibodies and may require adjustment when using other brands.
10. The details in this protocol are for visualization of the Western blot using Licor Odyssey equipment, which visualizes infrared

dye-conjugated secondary antibodies and allows quantification of the signal. This means that LC3-II signal can be normalized to actin and the results viewed quantitatively rather than qualitatively. Quantification should not be used as a replacement for even loading. It is important to normalize LC3-II to actin rather than to LC3-I, as LC3-I levels vary relative to LC3-II depending on cell types and treatment, and the dynamics of the pool are poorly understood. In addition, the affinity of anti-LC3 antibody to LC3-I and LC3-II is different [7].

11. Western blots can be developed using enhanced chemiluminescence (ECL) instead of the Licor system. We prefer to use the Licor system because it allows quantification of protein across a wider linear range than is possible with ECL. If ECL is being used, use HRP-conjugated secondary antibodies. For more information on Licor equipment, visit: http://www.licor.com/bio/applications/quantitative_western_blots/.
12. The volumes of cell suspension given here are for HeLa cells which are to be subjected to a 4 h treatment with a hypothetical drug. With different cell lines and treatments, the seeding density will need to be varied accordingly, ensuring that the cells will be approaching confluence when harvested. For example, a longer treatment time means that fewer cells need to be seeded. The rate of growth of the cells will also affect the number of cells which are seeded. For very long treatments, such as some small interfering RNA (siRNA) knockdown procedures, it may be necessary to split the cells part-way through the treatment.
13. Bafilomycin A_1 causes a defect in lysosomal acidification, which causes a block in fusion with autophagosomes [9]. Prolonged treatment with Bafilomycin A_1 can affect other protein degradation mechanisms [12]. We use 400 nM Bafilomycin A_1 (4 μL of stock solution in 1 mL final volume) for 4 h, or reduced concentrations if the treatment needs to be longer. Bafilomycin A_1 treatments longer than 16 h are best avoided where possible. Where the conditions being investigated require longer times in culture, we recommend adding Bafilomycin A_1 for the final 4 h only.
14. In order to optimize LC3 stability, cells can be lysed directly in 200 μL of Laemmli buffer in each well. This however means that a protein assay cannot be carried out. If, by eye, there appear to be similar numbers of cells per well, then the protein assay is not essential and this method can be used instead. With longer treatments or ones that affect cell survival or growth rates, the number of cells in each well can vary quite significantly, so we recommend that a protein assay is used.
15. For example, if a sample has been lysed in 100 μL of RIPA buffer, and 15 μL of RIPA has been added to adjust the concentration to match the other samples, add 115 μL of 2× Laemmli buffer.

16. If the samples have a high DNA content, making them hard to load accurately into gels (particularly an issue when lysing directly into Laemmli buffer), boiling for slightly longer (e.g. 7 min) at this stage can reduce the viscosity. Care should be taken not to damage the samples by boiling for too long however. Alternatively, samples can be sonicated briefly using a probe sonicator.
17. Having a reasonably sized stacking gel layer improves band resolution.
18. The layer of isopropanol prevents contact with the air, which inhibits polymerization of the acrylamide. In addition, this layer bursts bubbles and gives an even top to the resolving gel.
19. The stacking gel will be set after approximately 10 min, but in our experience allowing longer gives a better quality Western blot, with improved band resolution.
20. Aim to load approximately 15 μg of protein per well. When using a 15-well comb we load around 10 μg in each well. Even protein loading is very important, as analysis of the blots involves comparison of the lanes. Actin is used as a loading control in this procedure. Loading an equal (or at least similar) volume of sample in each lane, as achieved by ensuring equal protein concentrations in the samples, is important to make sure that the gel runs evenly.
21. Semi-dry transfer can be used instead of wet transfer; in our experience, the wet transfer method gives slightly better transfer results for LC3.
22. With practise, an idea of whether a transfer has been successful can be obtained by holding the membrane at an angle to the light and seeing if the lanes can be seen. This method also shows whether there were bubbles in the cassette.
23. Blocking is a very flexible step, and can last between 30 min and several hours.
24. If the primary antibodies used were both raised in the same animal, then the parts of the membrane can be incubated together in secondary antibody. With the Licor system, it is possible to color the bands red or green, thus distinguishing them by color as well as size, by using IR680- or IR800-conjugated secondary antibody respectively.
25. To view the membranes using enhanced chemiluminescence (ECL), mix equal volumes of ECL reagents, allowing 1 mL of solution per membrane. Pipette ECL reagents onto the surface of the membrane which was against the gel during transfer, ensuring the solution is evenly spread over the membrane. Allow to develop for approximately 30 s. Using forceps, touch the edge of the membrane against tissue to remove excess ECL

reagent. Develop the blot using photographic film. Take care not to overdevelop the blot. The bands should be clearly seen, but writing should be visible through the bands when the film is held against text. If films are overdeveloped, the sensitivity of the blot is reduced as differences in band intensity are harder to observe.

Acknowledgements

We are grateful for funding from the Wellcome Trust to D.C.R. (Principal Research Fellowship) and A.S. (PhD Studentship).

References

1. Ravikumar B, Sarkar S, Davies JE et al (2010) Regulation of mammalian autophagy in physiology and pathophysiology. Physiol Rev 90: 1383–1435
2. Ravikumar B, Duden R, Rubinsztein DC (2002) Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. Hum Mol Genet 11: 1107–1117
3. Williams A, Sarkar S, Cuddon P et al (2008) Novel targets for Huntington's disease in an mTOR-independent autophagy pathway. Nat Chem Biol 4:295–305
4. Webb JL, Ravikumar B, Atkins J et al (2003) Alpha-synuclein is degraded by both autophagy and the proteasome. J Biol Chem 278: 25009–25013
5. Berger Z, Ravikumar B, Menzies FM et al (2006) Rapamycin alleviates toxicity of different aggregate-prone proteins. Hum Mol Genet 15:433–442
6. Kabeya Y, Mizushima N, Ueno T et al (2000) LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. EMBO J 19:5720–5728
7. Kabeya Y, Mizushima N, Yamamoto A et al (2004) LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. J Cell Sci 117:2805–2812
8. Yamamoto A, Tagawa Y, Yoshimori T et al (1998) Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. Cell Struct Funct 23:33–42
9. Klionsky DJ, Elazar Z, Seglen PO, Rubinsztein DC (2008) Does bafilomycin A1 block the fusion of autophagosomes with lysosomes? Autophagy 4:849–850
10. Mizushima N, Yoshimori T (2007) How to interpret LC3 immunoblotting. Autophagy 3: 542–545
11. Rubinsztein DC, Cuervo AM, Ravikumar B et al (2009) In search of an "autophagomometer". Autophagy 5:585–589
12. Korolchuk VI, Mansilla A, Menzies FM, Rubinsztein DC (2009) Autophagy inhibition compromises degradation of ubiquitin-proteasome pathway substrates. Mol Cell 33: 517–527

Chapter 9

Advanced Mitochondrial Respiration Assay for Evaluation of Mitochondrial Dysfunction in Alzheimer's Disease

Amandine Grimm, Karen Schmitt, and Anne Eckert

Abstract

Alzheimer's disease (AD) is characterized by the presence of amyloid plaques (aggregates of amyloid-β [Aβ]) and neurofibrillary tangles (aggregates of tau) in the brain, but the underlying mechanisms of the disease are still partially unclear. A growing body of evidence supports mitochondrial dysfunction as a prominent and early, chronic oxidative stress-associated event that contributes to synaptic abnormalities, and, ultimately, selective neuronal degeneration in AD. Using a high-resolution respirometry system, we shed new light on the close interrelationship of this organelle with Aβ and tau in the pathogenic process underlying AD by showing a synergistic effect of these two hallmark proteins on the oxidative phosphorylation capacity of mitochondria isolated from the brain of transgenic AD mice. In the present chapter, we first introduce the principle of the Aβ and tau interaction on mitochondrial respiration, and secondly, we describe in detail the used respiratory protocol.

Key words Mitochondria, Alzheimer's disease, Amyloid-β, Tau, Oxygraph, High-resolution respirometry (HRR), Oxidative phosphorylation

1 Introduction

With the increasing average life span of humans, Alzheimer's disease (AD) is the most common neurodegenerative disorder among elderly individuals. It accounts for up to 80 % of all dementia cases and ranks as the fourth leading cause of death amongst those above 65 years of age [1]. Although the hallmark lesions of the disease were already described by Alois Alzheimer in 1906—amyloid-β (Aβ) containing plaques and microtubule-associated protein tau-containing neurofibrillary tangles (NFTs)—the underlying molecular mechanisms that cause the formation of these end-stage lesions are still poorly understood. However, a growing body of evidence supports mitochondrial dysfunction as a prominent and early chronic oxidative stress-associated event that contributes to synaptic abnormalities and, ultimately, selective neuronal degeneration in AD [2, 3]. Within the last few years, several cell culture models

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_9,

as well as single, double, and more recently triple transgenic mouse models have been developed to reproduce diverse aspects of AD. These models help in understanding the pathogenic mechanisms that lead to mitochondrial failure in AD, and in particular the interplay of AD-related cellular modifications within this process [4]. In this chapter, we highlight the critical key role of mitochondria and the close inter-relationship of this organelle with the two main pathological features in the pathogenic process underlying AD. Particularly, we will emphasize on the recent insights showing independent as well as synergistic effects of Aβ peptide and hyperphosphorylated tau on mitochondrial function by using a high-resolution respirometry system (Oxygraph-2k).

1.1 Aβ and Tau Induce Mitochondrial Toxicity

Mitochondria play a pivotal role in cell survival and death by regulating both energy metabolism and apoptotic pathways. They are the "powerhouses of cells" providing energy via ATP generation which is accomplished through oxidative phosphorylation (OXPHOS) from nutritional sources [5] (Fig. 1). Neurons have particularly high numbers of mitochondria which are especially enriched in synapses. Due to the limited glycolytic capacity of neurons, those cells are

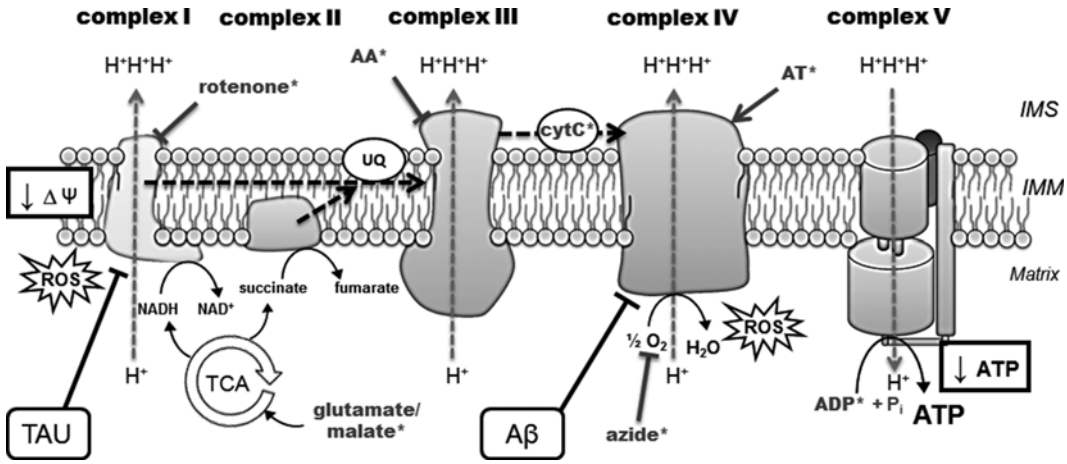


Fig. 1 The mitochondrial electron transport chain: impact of Aβ peptide, tau protein, and effects of mitochondrial substrates used during the measurement protocol with Oxygraph. Complexes I (NADH: ubiquinone oxidoreductase) and II (succinate dehydrogenase, belongs to the tricarboxylic acid (TCA) cycle) receive electrons from NADH and $FADH_2$, respectively. Electrons are then driven from complexes by the mobile carrier molecules coenzyme Q/ubiquinone (*UQ*) and cytochrome c (*Cyt c*) to the final acceptor, molecular oxygen (O_2). Electron flow is coupled to proton movement across the inner mitochondrial membrane (*IMM*) in complexes I, III and IV. The resulting proton gradient is harvested by complex V to generate ATP. In Alzheimer's disease, abnormal mitochondrial electron activities have been observed, predominantly in complexes I and IV, leading to impaired mitochondrial membrane potential, decreased production of ATP (complex V), and increasing reactive oxygen species (*ROS*) levels. Interestingly, deregulation of complex I is mainly tau-dependent, while deregulation of complex IV is amyloid-β (Aβ)-dependent, at both the protein and activity levels. The targets of the different substrates used during the Oxygraph measurement are marked with an *asterisk* and their specific actions are summarized in Table 1. *AA* antimycin A, *AT* ascorbate/TMPD, *IMS* intermembrane space

highly dependent on mitochondrial function for energy production [6]. Thus, deregulation of mitochondrial function leads to synaptic stress, disruption of synaptic transmission, apoptosis and ultimately, systemic neurodegeneration [7, 8].

Evidences from cellular and animal AD models indicate that Aβ triggers mitochondrial dysfunction through a number of pathways such as impairment of OXPHOS, elevation of reactive oxygen species (ROS) production, interaction with mitochondrial proteins, and alteration of mitochondrial dynamics [9, 10]. Success in developing mouse models that mimic diverse facets of the disease process has greatly facilitated the understanding of pathophysiological mechanisms underlying AD. In 1995, Games and collaborators established the first amyloid precursor protein (APP) mouse model (called PDAPP) bearing the human "Indiana" mutation of the *APP* gene (V171F). They observed the accumulation of Aβ in the brain and subsequent amyloid plaque formation, as well as astrocytosis and neuritic dystrophy [4]. Interestingly, in most of the APP mouse models, the cognitive impairment begins concomitantly with Aβ oligomer formation in the brain (around 6 months of age), while neuritic amyloid deposits become visible only between 12 and 23 months and the amount of deposits increases in parallel [11]. Thus, memory deficits seem to correlate directly with the accumulation of intracellular Aβ oligomers and not with amyloid plaque formation. When those mice were crossed with those bearing a mutation in presenilin 1 gene (PS1), coding for a gene involved in APP processing, an earlier onset of amyloid plaques was observed, alongside a stronger decrease of mitochondrial membrane potential as well as ATP level [12].

Mitochondrial dysfunctions occur at a very early disease stage in AD transgenic mouse models. For example, in the APPsw transgenic strain Tg2576 (Swedish mutation), an upregulation of genes related to mitochondrial energy metabolism and apoptosis was observed already at 2 months of age. Alterations in composition of the mitochondrial respiratory chain complexes I and III protein subunit as well as impairment of mitochondrial respiration were detected around 6 months, when soluble Aβ accumulated in the brain without plaque formation [13, 14].

Consistent with this observation, in APPsw/presenilin 2 (PS2) double-transgenic mice, mitochondrial impairment was first detected at 8 months of age, before amyloid plaque deposition, but after soluble Aβ accumulation [15]. Taken together, these findings are consistent with the recently proposed hypothesis of an age-related Aβ toxicity cascade that suggests that the most toxic Aβ species that cause majority of molecular and biochemical abnormalities are in fact intracellular soluble oligomeric aggregates rather than the extracellular, insoluble plaques [16].

How does tau, the second hallmark lesion in AD, interfere with mitochondrial function? In its abnormally hyperphosphorylated form, which forms the neurofibrillary tangles (NFTs), tau has

been shown to block mitochondrial transport. This results in energy deprivation and oxidative stress at the synapse, and, consequently, neurodegeneration [17, 18]. Until now, no mutations in microtubule-associated protein tau (*MAPT*) coding genes have been detected in relation to familial forms of AD. However, in familial frontotemporal dementia (FTD) with parkinsonism, mutations in the microtubule-associated protein tau gene (*MAPT*) were identified on chromosome 17. This was the basis for creating a robust mouse model for tau pathology in 2001. These P301L tau-expressing pR5 mice show an accumulation of tau as soon as 3 months of age and develop NFTs around 6 months of age [19]. A mass spectrometric analysis of the brain proteins from these mice (aged from 8.5 to 10 months) revealed mainly a deregulation of mitochondrial respiratory chain complex components (including complex V), antioxidant enzymes, and synaptic protein space [20]. The reduction in mitochondrial complex V levels in the P301L tau mice was also confirmed in human P301L FTDP-17 (FTD with parkinsonism linked to chromosome 17) brains. The functional analysis demonstrated age-related mitochondrial dysfunction, together with reduced NADH ubiquinone oxidoreductase (complex I) activity as well as age-related impaired mitochondrial respiration and ATP synthesis in a pR5 mouse model. Mitochondrial dysfunction was also associated with higher levels of ROS in aged transgenic mice. Increased tau pathology resulted in modification of lipid peroxidation levels and the upregulation of antioxidant enzymes in response to oxidative stress [20]. Thus, this evidence demonstrated for the first time that not only Aβ but also tau pathology weakens gradually mitochondrial function in a rather specific way leading to metabolic impairment and oxidative stress in AD.

1.2 Synergistic Mode of Action of Aβ and Tau

Although Aβ and tau pathologies are both known hallmarks of AD, the mechanisms underlying the interplay between plaques and NFTs (or Aβ and tau, respectively) have remained unclear. However, a close relationship between mitochondrial impairment and Aβ on the one hand and tau on the other hand has been already established. How do both AD features relate to each other? Several studies suggest that Aβ aggregates and hyperphosphorylated tau may block the mitochondrial transport to the synapse leading to energy deficiency and neurodegeneration [21].

Remarkably, intracerebral Aβ injections amplify a pre-existing tau pathology in several transgenic mouse models [22, 23], whereas lack of tau abrogates Aβ toxicity [18, 24]. Our findings indicate that in tau transgenic pR5 mice, mitochondria display an enhanced vulnerability toward Aβ insult in vitro [2, 25], suggesting a synergistic action of tau and Aβ pathology on this organelle. Thus, these studies provide the first evidence for the existence of a complex interplay between Aβ and tau in AD whereby these two molecules damage mitochondria in multiple ways, but what about their specific effects on mitochondrial respiration?

1.3 High-Resolution Respirometry in Isolated Mitochondria to Evaluate OXPHOS Capacity

To address this question, we used a high-resolution respiratory system to evaluate the capacity of the entire oxidative phosphorylation system (OXPHOS) of cerebral mitochondria from mice bearing either an APP/PS2 mutation, P301L mutation (pR5 mice), or the triple mutation APP/PS2/P301L (^{triple}AD mice) compared to wild-type mice [26]. Measurement of oxygen (O_2) flux and consumption was performed at 37 °C using an Oroboros Oxygraph-2k system on freshly isolated mitochondria from cortical brains of age-matched wild-type, APP/PS2, pR5 and ^{triple}AD mice as follows. After detection of endogenous respiration, glutamate and malate were added to induce state 4 respiration (Figs. 1 and 2a), then ADP was added to stimulate state 3 respiration. After determining coupled respiration, a mitochondrial uncoupler (FCCP, *see* below) was added and the maximal respiratory capacity measured in the absence of a proton gradient. Cytochrome c (cyt c) injection was used to demonstrate mitochondrial membrane integrity. To inhibit activities of complexes I–III, rotenone (rot) and antimycin A (AA) were added. Complex IV activity was stimulated by ascorbate/TMPD (A/T) before terminating mitochondrial respiration by adding sodium azide (azide). Oxygen (O_2) consumption was normalized to the corresponding citrate synthase activity [3, 26].

We determined flux control ratios to obtain information on metabolic states of respiration. The respiratory control ratio (RCR3/4) is an indicator of the state of coupling of mitochondria. State 3 is the rate of phosphorylating respiration in the presence of exogenous ADP, and state 4 is associated with proton leakage across the inner mitochondrial membrane in the absence of ADP. Our findings suggest a pronounced decrease of RCR3/4 in mitochondria from APP/PS2 and ^{triple}AD compared with age-matched wild-type mice already at 8 months of age. This decrease was also found in the oldest mice (12 months of age). When we examined the ETS/ROX (electron transport system/residual oxygen consumption) ratio, which yields an index of the maximum oxygen consumption capacity relative to the magnitude of residual oxygen consumption, we found that it was also decreased in APP/PS2 and ^{triple}AD compared with age-matched wild-type mice at 8 and 12 months of age. Interestingly, in a previous study, the decreased respiration of mitochondria from pR5 mice compared with wild-type controls was not detectable before the age of 24 months [20]. In contrast, APP/PS2 mitochondria showed a decrease in OXPHOS compared with wild-type already at the age of 8 months. At this age, OXPHOS of brain mitochondria from ^{triple}AD mice did not differ compared with that of age-matched APP/PS2 mitochondria, but it was significantly decreased in ^{triple}AD mice at the age of 12 months (Fig. 2b). Taken together, with increasing age, the global failure of the mitochondrial respiratory capacity deteriorated the strongest in mitochondria from ^{triple}AD mice, suggesting a synergistic destructive effect of tau and Aβ on mitochondria.

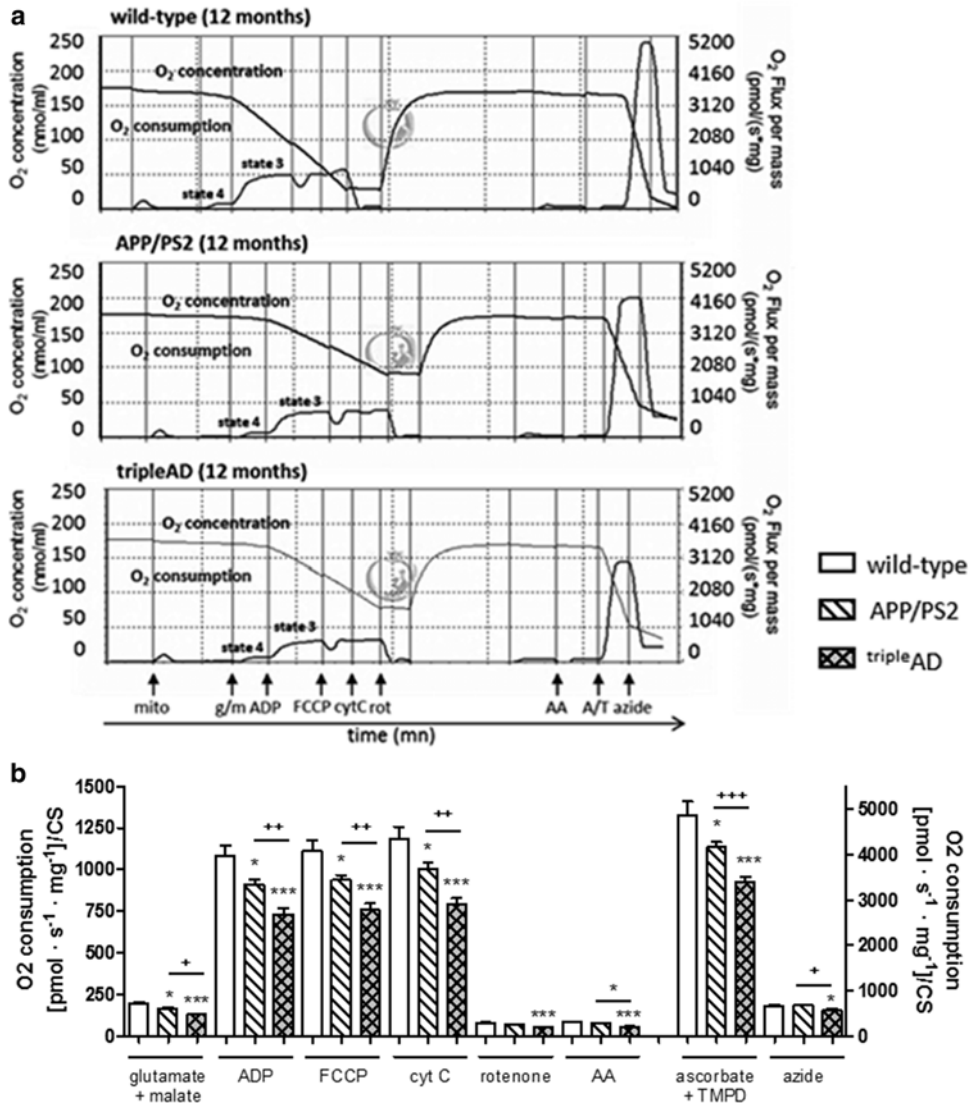


Fig. 2 Synergistic effects of Aβ and tau on mitochondrial respiration. (a) Representative diagrams of O_2 flux and consumption in mitochondria from 12-month-old wild-type, APP/PS2, and ^{triple}AD transgenic mice in response to titrated substrates and inhibitors of mitochondrial complexes. (b) Two-way ANOVA revealed a significant effect of on the respiratory rates of mitochondria between 12-month-old wild-type and APP/PS2 mice, and this impaired respiration was even more pronounced in ^{triple}AD mice. Two-way ANOVA post-hoc Bonferroni. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$ vs. wild-type; $^{+}P<0.05$; $^{++}P<0.01$; $^{+++}P<0.001$ vs. APP/PS2 ($n=7$–12 animals/group). Modified from Rhein et al., PNAS (2009) [26] with permission

In conclusion, our studies highlight the key role of mitochondria in AD pathogenesis and the close interrelationship of this organelle and the two main pathological features of the disease. We showed that disturbances in the respiratory and energy system of ^{triple}AD mice seem to be due to a convergence of Aβ and tau on

mitochondria, accelerating defects in respiratory capacity, which consolidates the idea that a synergistic effect of tau and Aβ increase the pathological deterioration of mitochondria.

Now we will describe in detail the protocol which we followed previously [25]. After listing the material needed, we will describe the isolation of mitochondria from mouse brains and the steps required to measure the mitochondrial respiration. It is important to note that this protocol assumes that the Oroboros Oxygraph-2k system is routinely used in the laboratory and does not include technical details about oxygraph maintenance or calibration, but only experimental procedure regarding the assessment of mitochondrial respiration.

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C).

2.1 Solutions for Isolated Mitochondria Preparation

1. Medium 1: 138 mM NaCl, 5.4 mM KCl, 0.17 mM Na_2HPO_4, 0.22 mM KH_2PO_4, 5.5 mM glucose·H_2O, 58.4 mM sucrose, pH 7.35. To prepare 1 L of Medium 1, weigh 8 g of NaCl, 0.4 g of KCl, 0.024 g of Na_2HPO_4, 0.03 g of KH_2PO_4, 1.1 g of glucose·H_2O, 20 g of sucrose. Add water to a volume of 900 mL and mix with magnetic stir bar at room temperature until all powders are dissolved. Adjust pH and make up to 1 L with water. Store at 4 °C.
2. Isolated Mitochondria Buffer: 210 mM mannitol, 70 mM sucrose, 10 mM HEPES, 1 mM EDTA (tritriplex III), 0.45 % BSA, pH 7.4. To prepare 200 mL of buffer, weigh 7.65 g of mannitol, 4.79 g of sucrose, 477 mg of HEPES, 74.4 mg of EDTA (tritriplex III) and 0.9 g of BSA. Add water to a volume of 190 mL and mix with magnetic stir bar at room temperature until all powders are dissolved. Adjust pH and make up to 200 mL with water. Prepare aliquots of 10 mL and keep at −20 °C (*see* **Note 1**).
3. 1 M dithiothreitol (DTT) stock solution in water.

2.2 Solutions for Mitochondrial Respiration Assay

1. Mitochondrial Respiration Buffer: 65 mM sucrose, 10 mM KH_2PO_4, 10 mM Tris–HCl, 10 mM $MgSO_4·7H_2O$, 2 mM EDTA (tritriplex III)·$2H_2O$, pH 7. To prepare 200 mL, weigh 4.45 g of sucrose, 0.272 g of KH_2PO_4, 0.315 g of Tris–HCl, 0.493 g of $MgSO_4·7H_2O$, and 0.149 g of EDTA (tritriplex III)·$2H_2O$. Add water to a volume of ca. 190 mL and mix with magnetic stir bar at room temperature until all powders are dissolved. Adjust pH and make up to 200 mL with water. Prepare aliquots of 20 mL and keep them at −20 °C.

2. Mitochondrial Respiration Medium (MiR05): 0.5 mM EGTA, 3 mM $MgCl_2 \cdot 6H_2O$, 60 mM K-lactobionate, 20 mM taurine, 10 mM KH_2PO_4, 20 mM HEPES, 110 mM sucrose, 1 g/L BSA, pH 7.1. First, prepare the 0.5 M K-lactobionate stock solution dissolving 35.83 g of lactobionic acid in 100 mL of H_2O, adjusting the pH to 7.0 with KOH and bringing the volume to 200 mL. To prepare 1 L of MiR05, weigh 0.190 g of EGTA, 0.610 g of $MgCl_2 \cdot 6H_2O$, 2.502 g of taurine, 1.361 g of KH_2PO_4, 4.77 g of HEPES, 37.65 g of sucrose, and 1 g of BSA. Add ca. 750 mL of water and 120 mL of 0.5 M K-lactobionate stock solution. Mix with magnetic stir bar at room temperature, adjust the pH to 7.1 with 5 N KOH and make up to 1 L with water. Divide into 20 mL aliquots and store them frozen at –20 °C (*see* **Note 2**).
3. Substrates: The substrates (stock solutions) employed and details of preparation are summarized in Table 1 (*see* **Notes 3–13**).

2.3 Oxygraph System

Oroboros Oxygraph-2k system for high resolution respirometry (HRR) studies (http://www.oroboros.at/?Oxygraph).

3 Methods

Before experiment, perform an instrumental and chemical setup with the oxygraph.

3.1 Isolated Mitochondria Preparation

Prepare isolated mitochondria buffer (*see* **Note 1**) and keep on ice. Turn on the centrifuge (4 °C).

1. Kill the mice by decapitation and dissect one brain hemisphere on ice. Wash in 10 mL of ice-cold medium 1.
2. Put the preparation in the Potter-tube to homogenize in 1 mL of isolated mitochondria buffer. Pipette 10–15 times to homogenize the preparation (*see* **Note 14**).
3. Wash the Potter's plug three times with 150 µL of isolated mitochondria buffer and put the preparation in a 2 mL tube. Wash the Potter's tube three times with 150 µL of isolated mitochondria buffer and put the preparation in the same 2 mL tube. Vortex (*see* **Note 15**).
4. Centrifuge at 1,450 × *g* (4 °C) for 7 min and recover the supernatant in a new 2 mL tube. This step removes nuclei and tissue particles.
5. Centrifuge at 1,450 × *g* (4 °C) for 3 min and recover the supernatant again in a new 2 mL tube.
6. Centrifuge at 10,000 × *g* (4 °C) for 5 min. Throw away the supernatant and recover the pellet.

Table 1
Preparation and function of the substrates used to investigate the mitochondrial respiration using the Oroboros Oxygraph-2k system

		Name/formula	MW (g/mol)	Stock solution		Function	Note #
Mitochondrial substrates	Glutamate	Na-glutamate $C_5H_8NO_4Na$	169.1	2 M	3.382 g/10 mL H_2O	Induces state 4 respiration	4
	Malate	L-Malic acid $C_4H_6O_5$	134.1	0.8 M	1.073 g/10 mL H_2O	Induces state 4 respiration	5
	ADP	K-adenosine 5′diphosphate dihydrate salt ($C_{10}H_{15}N_5O_{10}P_2K{\cdot}2H_2O$)	501.32	0.5 M	0.501 g/2 mL H_2O	Induces state 3 respiration	6
	cytC	Cytochrome c	12,500	4 mM	50 mg/1 mL H_2O	Demonstrates mitochondrial membrane integrity	7
	Ascorbate	Ascorbate sodium salt $C_6H_7O_6Na$	198.1	0.8 M	1.584 g/10 mL H_2O	Stimulates complex IV activity	8
	TMPD	*N*,*N*,*N*′,*N*′-Tetramethyl-p-phenylenediamine dihydrochloride $C_{10}H_{16}N_2{\cdot}2HCl$	237.2	0.2 M	47.4 mg/1 mL H_2O	Stimulates complex IV activity	9
Mitochondrial inhibitors	Rotenone	Rotenone $C_{23}H_{22}O_6$	394.4	1 mM	3.94 mg/10 mL Ethanol	Inhibits complex I activity	10
	AA	Antimycin A	540	5 mM	11 mg/4 mL Ethanol	Inhibits complex III activity	11
	Azide	Sodium azide NaN_3	65.01	1 M	65 mg/1 mL H_2O	Inhibits oxygen consumption	12
Mitochondrial uncouplers	FCCP	Carbonyl cyanide *p*-(trifluoromethoxy) phenylhydrazone $C_{10}H_5F_3N_4O$	254.2	0.32 mM	2.54 mg/10 mL ethanol results in a 1 mM solution. From here, dilute 1:3.125 in ethanol	Determines uncoupled respiration in absence of a proton gradient	13

7. Put the pellet (mitochondria) in 1 mL of isolated mitochondria buffer and mix 15 times using the pipette.
8. Repeat **steps 7** and **8** to obtain the mitochondrial fraction and put the pellet in 100 µL of isolated mitochondria buffer. Keep on ice until the measurement (*see* **Note 16**).

3.2 Mitochondrial Respiration Measurement: Preparations

Before experiment prepare the substrates (stock solutions) (Table 1) and the oxygraph (*see* **Note 17**).

1. Add 50 µL of isolated mitochondria preparation to each chamber and close the chamber (*see* **Note 18**). Mark it as (01-state 1).
2. Add 10 µL of 2 M glutamate/5 µL of 0.8 M malate (respirometry assay final concentrations will be 10 mM and 2 mM respectively). Mark it as (02-GM2).
3. Add 8 µL of 0.5 M ADP/chamber (final assay concentration, 2 mM). Mark it as (03-GM3).
4. Add 2.5 µL of 0.32 mM FCCP/chamber (assay concentration, 0.4 µM). Mark it as (04-GP3u).
5. Add 5 µL of 4 mM of Cytochrome c/chamber (assay concentration, 10 µM). Mark it as (05-GM3c).
6. Preparing from stock solution (Table 1), add 5 µL of 0.2 mM rotenone/chamber (assay concentration, 0.5 µM). Mark it as (06-rot).
7. Preparing from stock solution (Table 1), add 5 µL of 1 mM antimycine A/chamber (assay concentration, 2.5 µM). Mark it as (07-AA).
8. Add 5 µL of 0.8 M sodium ascorbate/chamber (assay concentration, 2 mM) and 5 µL of 0.2 M TMPD/chamber (assay concentration, 0.5 mM). Mark it as (08-AT).
9. Add 20 µL of 1 M sodium azide/chamber (assay concentration, 10 mM). Mark it as (09-azide).

3.3 Mitochondrial Respiration Measurement: High-Resolution Respirometry

Mitochondrial oxygen consumption is measured by high-resolution respirometry (HRR) at 37 °C using an Oroboros Oxygraph-2k system (http://www.oroboros.at/?Oxygraph) following the Gnaiger method [27].

3.4 Analysis

After the measurement, extract the raw data from the oxygraph software (DatLab) to an Excel file. Normalize the data on citrate synthase activity, which correlates with mitochondrial content (*see* **Note 19**). Perform the statistical analysis using GraphPad Prim software (or equivalent) and a two-way ANOVA followed by Bonferroni post hoc tests to compare the different groups. Consider statistically significant only P values < 0.05. Represent data as means ± SEM.

4 Notes

1. Just before starting the experiment, warm up the isolated mitochondria buffer. For two brain hemispheres, add one tablet of Complete^{R} Mini (protease inhibitor cocktail tablet) and 5 μL of 1 M DTT to 10 mL of buffer (final concentration, 0.5 mM DTT). Prepare fresh, less than 3 h before use.
2. The MiR05 medium is stable for about 2–3 months. The K-lactobionate must be prepared fresh.
3. Manipulation of solutions at low temperature (4 °C). After rewarming, mix carefully since phase separation may occur and compounds may precipitate in cold mixtures. During the course of the experiment keep stock solutions on ice. Note: Solutions which contain ethanol may have a problem of evaporation and subsequent increase of concentration.
4. Glutamate solution (*see* Table 1). Adjust pH to 7.0 with 37 % HCl and divide into 0.5 mL aliquots. Store frozen at −20 °C.
5. Malate solution (*see* Table 1). Neutralize (adjust to pH 7.0) with 10 N KOH and divide into 0.5 mL aliquots. Store frozen at −20 °C.
6. ADP solution (*see* Table 1). Neutralize with 5 N KOH and divide into 100 μL aliquots. Store at −80 °C.
7. Cytochrome c solution (*see* Table 1). Divide into 0.2 mL aliquots. Store frozen at −20 °C. Protect from light.
8. Ascorbate solution (*see* Table 1). To prevent autoxidation, prepare 0.8 M ascorbic acid solution (137.6 mg/mL, pH ca. 2). Adjust the pH of the sodium ascorbate solution to ca. 6 with ascorbic acid. Divide into 0.2 mL aliquots. Store frozen at −20 °C protected from light (light sensitive).
9. TMPD solution (*see* Table 1). To prevent autoxidation, neutralize with the ascorbate salt solution. Dilute 1:80 to result in a solution with 10 mM ascorbate final concentration. Divide into 0.2 mL aliquots. Store frozen at −20 °C.
10. Rotenone solution (*see* Table 1). Difficult to dissolve. Divide into 0.2 mL aliquots, store at −20 °C protected from light. Note: Light sensitive; very toxic. Handle with care.
11. Antimycin A (AA) solution (*see* Table 1). Divide into 0.2 mL aliquots, store frozen at −20 °C. Note: Very toxic. Handle with care.
12. Azide solution (*see* Table 1). Divide into 0.2 mL aliquots, store frozen at −20 °C. Note: Very toxic. Handle with care.
13. FCCP solution (*see* Table 1). Divide into 0.5 mL aliquots, store frozen at −20 °C.
14. Pipette gently up and down to avoid bubble formation and strong oxygenation of the sample.

15. If you have several mice, stop the process at this step, put the preparation on ice and use the next mouse to perform the centrifugation steps with all the samples at the same time. Since the Oxygraph contains two chambers, it is possible to investigate the mitochondrial respiration for only a few animals per day (6–8 mice/day).
16. A volume of 50 μL of the preparation will be used for the Oxygraph measurement. For protein determination, dilute 3 μL of isolated mitochondria in PBS (dilution 1:5) and perform the protein assay (e.g. Biorad DC™ Protein Assay and bovine serum albumin (BSA) for the standard curve).
17. The experiment requires an instrumental and chemical background following the protocol of the company (http://www.oroboros.at/?Oxygraph). Careful calibration will determine the "air saturation" (R1) and the "zero saturation" (R0) values.
18. When the oxygraph chambers are closed, check no air bubbles are left inside.
19. Citrate synthase activity is frequently used to normalize other mitochondrial enzymatic activities and mitochondrial respiration because it correlates to mitochondrial content. Citrate synthase activity can be measured following the reduction of 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) by citrate synthase at 412 nm (extinction coefficient of 13.6 mM^{-1} cm^{-1}) in a coupled reaction with coenzyme A (CoA) and oxaloacetate [3, 26].

Acknowledgements

This work was supported by grants from the Swiss National Science Foundation (Grant SNF 310000–108223), Synapsis and Novartis Foundation for medical-biological research.

References

1. Gibson GE, Huang HM (2002) Oxidative processes in the brain and non-neuronal tissues as biomarkers of Alzheimer's disease. Front Biosci 7:d1007–d1015
2. Eckert A, Hauptmann S, Scherping I et al (2008) Soluble beta-amyloid leads to mitochondrial defects in amyloid precursor protein and tau transgenic mice. Neurodegener Dis 5:157–159
3. Rhein V, Baysang G, Rao S et al (2009) Amyloid-beta leads to impaired cellular respiration, energy production and mitochondrial electron chain complex activities in human neuroblastoma cells. Cell Mol Neurobiol 29: 1063–1071
4. Games D, Adams D, Alessandrini R et al (1995) Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. Nature 373:523–527
5. Scheffler IE (2001) A century of mitochondrial research: achievements and perspectives. Mitochondrion 1:3–31
6. Reddy PH (2007) Mitochondrial dysfunction in aging and Alzheimer's disease: strategies to protect neurons. Antioxid Redox Signal 9: 1647–1658

7. Caspersen C, Wang N, Yao J et al (2005) Mitochondrial Abeta: a potential focal point for neuronal metabolic dysfunction in Alzheimer's disease. FASEB J 19:2040–2041
8. Santos RX, Correia SC, Wang X et al (2010) Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. Int J Clin Exp Pathol 3:570–581
9. Pagani L, Eckert A (2011) Amyloid-beta interaction with mitochondria. Int J Alzheimers Dis 2011:925050
10. Eckert A, Schmitt K, Gotz J (2011) Mitochondrial dysfunction—the beginning of the end in Alzheimer's disease? Separate and synergistic modes of tau and amyloid-beta toxicity. Alzheimers Res Ther 3:15
11. Schmitt K, Grimm A, Kazmierczak A et al (2012) Insights into mitochondrial dysfunction: aging, amyloid-beta, and tau-A deleterious trio. Antioxid Redox Signal 16: 1456–1466
12. Blanchard V, Moussaoui S, Czech C et al (2003) Time sequence of maturation of dystrophic neurites associated with Abeta deposits in APP/PS1 transgenic mice. Exp Neurol 184:247–263
13. Gillardon F, Rist W, Kussmaul L et al (2007) Proteomic and functional alterations in brain mitochondria from Tg2576 mice occur before amyloid plaque deposition. Proteomics 7: 605–616
14. Reddy PH, McWeeney S, Park BS et al (2004) Gene expression profiles of transcripts in amyloid precursor protein transgenic mice: up-regulation of mitochondrial metabolism and apoptotic genes is an early cellular change in Alzheimer's disease. Hum Mol Genet 13: 1225–1240
15. Richards JG, Higgins GA, Ouagazzal AM et al (2003) PS2APP transgenic mice, coexpressing hPS2mut and hAPPswe, show age-related cognitive deficits associated with discrete brain amyloid deposition and inflammation. J Neurosci 23:8989–9003
16. Fernandez-Vizarra P, Fernandez AP, Castro-Blanco S et al (2004) Intra- and extracellular Abeta and PHF in clinically evaluated cases of Alzheimer's disease. Histol Histopathol 19: 823–844
17. Gotz J, Ittner LM, Fandrich M, Schonrock N (2008) Is tau aggregation toxic or protective: a sensible question in the absence of sensitive methods? J Alzheimers Dis 14:423–429
18. Ittner LM, Ke YD, Delerue F et al (2010) Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models. Cell 142:387–397
19. Gotz J, Chen F, Barmettler R, Nitsch RM (2001) Tau filament formation in transgenic mice expressing P301L tau. J Biol Chem 276:529–534
20. David DC, Hauptmann S, Scherping I et al (2005) Proteomic and functional analyses reveal a mitochondrial dysfunction in P301L tau transgenic mice. J Biol Chem 280:23802–23814
21. Gotz J, Ittner LM, Kins S (2006) Do axonal defects in tau and amyloid precursor protein transgenic animals model axonopathy in Alzheimer's disease? J Neurochem 98: 993–1006
22. Gotz J, Chen F, van Dorpe J, Nitsch RM (2001) Formation of neurofibrillary tangles in P301l tau transgenic mice induced by Abeta 42 fibrils. Science 293:1491–1495
23. Gotz J, Schild A, Hoerndli F, Pennanen L (2004) Amyloid-induced neurofibrillary tangle formation in Alzheimer's disease: insight from transgenic mouse and tissue-culture models. Int J Dev Neurosci 22:453–465
24. Ittner LM, Gotz J (2011) Amyloid-beta and tau—a toxic pas de deux in Alzheimer's disease. Nat Rev Neurosci 12:65–72
25. Eckert A, Hauptmann S, Scherping I et al (2008) Oligomeric and fibrillar species of beta-amyloid (A beta 42) both impair mitochondrial function in P301L tau transgenic mice. J Mol Med 86:1255–1267
26. Rhein V, Song X, Wiesner A et al (2009) Amyloid-beta and tau synergistically impair the oxidative phosphorylation system in triple transgenic Alzheimer's disease mice. Proc Natl Acad Sci U S A 106:20057–20062
27. Gnaiger E (2008) Polarographic oxygen sensors, the oxygraph, and high resolution respirometry to assess mitochondrial function. In: Dykens JA, Will Y (eds) Drug-induced mitochondrial dysfunction. Wiley, Hoboken, NJ, pp 327–348

Chapter 10

Analysis of Microglial Proliferation in Alzheimer's Disease

Diego Gomez-Nicola and V. Hugh Perry

Abstract

The expansion and activation of the microglial population is a hallmark of many neurodegenerative diseases. Despite this fact, little quantitative information is available for specific neurodegenerative disorders, particularly for Alzheimer's disease (AD). Determining the degree of local proliferation will not only open avenues into understanding the dynamics of microglial proliferation, but also provide an effective target to design strategies with therapeutic potential. Here we describe immunohistochemical methods to analyse microglial proliferation in both transgenic murine models of AD and in human post-mortem samples, to provide a broad picture of the microglial response at the different experimental levels. The application of a common and universal method to analyse the microglial dynamics across different laboratories will help to understand the contribution of these cells to the pathology of AD and other neurodegenerative diseases.

Key words Alzheimer's disease, Microglia, Proliferation, Bromodeoxyuridine (BrdU), Ki67, Phospho Histone H3, CSF1R, PU.1, Immunohistochemistry

1 Introduction

Alzheimer's disease is a chronic neurodegenerative disease and the most common form of dementia in the Western countries. Despite much interest in the inflammatory response in AD, and the extensive research focused on understanding the role of microglia in this disease, the scientific community has failed to shed clear and uniform light into their contribution to the disease [1–3]. The neuropathology of AD shows a robust innate immune response characterized by the presence of activated microglia, with increased or de novo expression of diverse macrophage antigens [3, 4], and at least in some cases production of inflammatory cytokines [5, 6]. Microglial activation in neurodegeneration is accompanied by an increase in their density. In addition, other brain macrophages, perivascular macrophages (PVMs) and meningeal macrophages (MMs), play a critical role in signaling from the periphery to the brain. Recent studies report a minor or even absent contribution of circulating progenitors to the microglial population in a mouse

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_10,

model of AD [7], pointing to in situ microglial proliferation as the mechanism regulating microglial turnover, with little or no contribution of circulating progenitors [8, 9]. Microglia are maintained and function largely independently of circulating progenitors in health [10] and disease [7, 11, 12]. Therefore, the analysis of PVMs, MMs and microglial proliferation under pathological conditions with widespread chronic neurodegeneration, as is the case of Alzheimer's disease, is critical for understanding how innate inflammation contributes to disease onset and progression.

Although proliferation was assumed to be responsible for the increased number of microglial cells observed in AD samples, direct evidence of proliferating microglial cells (Ki67 expression in Iba1+ cells) was reported only recently, together with the upregulation of the transcription factor PU.1 and the mitogen IL-34, key components of the pathway regulating microglial proliferation [13]. An important signaling pathway for microglial proliferation, the CSF1-receptor (CSF1R) pathway, has also been shown to be upregulated in microglial cells during AD, indicating prominent activity of this pathway [14]. The expansion of the microglial population has been consistently documented in transgenic mouse models of AD, mainly accumulating around plaques [15, 16]. However, direct evidence of microglial proliferation (incorporation of bromodeoxyuridine (BrdU) in Iba1+ cells) was only recently reported, suggesting a direct effect of the plaque microenvironment over the regulation of microglial mitogenesis [17].

These studies pinpoint the importance of the control of microglial proliferation during AD, offering new avenues for the regulation of the innate immune response in the brain. Establishing reproducible and universal methods to monitor microglial proliferation in models mimicking aspects of AD and in post-mortem AD brains will provide the scientific community with valuable tools to better compare results across experimental models or cohorts of patients, contributing to a better understanding of the pathophysiology of AD.

2 Materials

The immunohistochemical identification of proliferating microglial cells can be performed using the following materials.

2.1 Tissue Samples

2.1.1 Mouse/Rat Tissue Samples

To provide a reliable correlate of cell proliferation we recommend the use of thymidine analogues such as bromodeoxyuridine (BrdU), which gets incorporated into the nuclear DNA in dividing cells (*see* **Note 1**). The use of fixed tissue obtained from intracardiac perfusion (4 % paraformaldehyde; *see* **Note 2**) is highly encouraged, although the methods are also applicable to the use of fresh-frozen brain tissue. We also encourage the use of reporter mice with

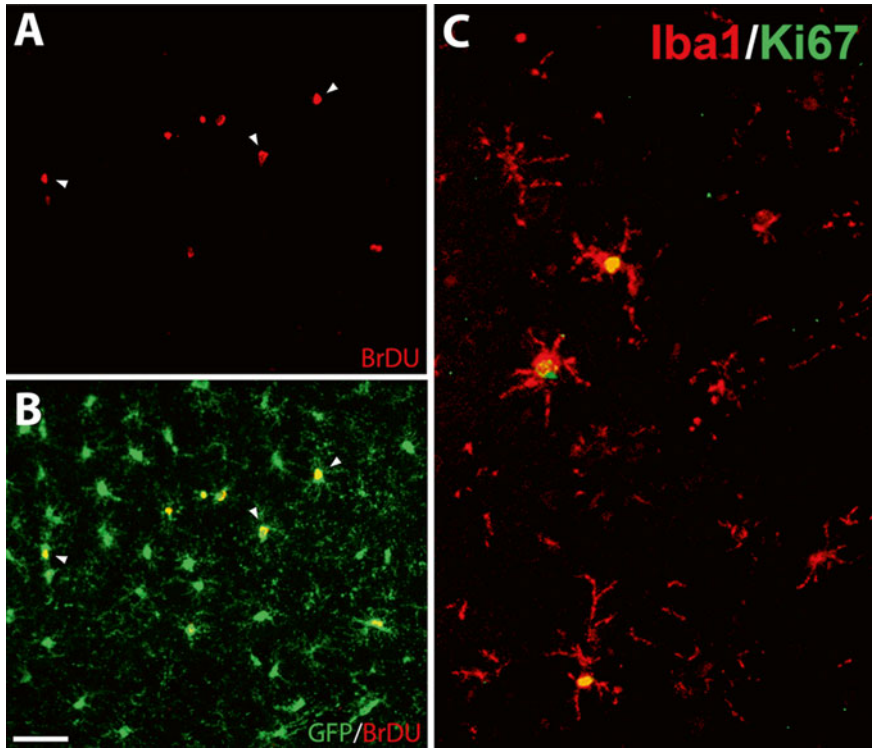


Fig. 1 Microglial proliferation in murine and human chronic neurodegeneration. (**a** and **b**) Representative image of the immunohistochemical detection of bromodeoxyuridine (BrdU) (**a**, **b**; *red*) in microglial cells (**b**; c-fms-EGFP+, *green*) in the hippocampus of a mouse having prion disease (ME7 model). (**c**) Representative image of the immunohistochemical detection of Ki67 (*green*) in microglial cells (Iba1+, *red*) from the temporal cortex of an AD patient. Scale bar in (**a** and **b**) 20 μm. In (**c**) 100 μm. Reproduced from Gomez-Nicola et al. [13], with permission from Journal of Neuroscience. Society for Neuroscience (www.jneurosci.org; reuse of own material)

fluorescent microglia/macrophages, such as c-fms EGFP mice (macgreen) [18] or CX3CR1 EGFP mice [19], to facilitate the detection of microglial cells in the brain (Fig. 1) (*see* **Note 1**).

2.1.2 Human Tissue Samples

Samples from post-mortem human tissue are usually obtained from brain banks as paraffin-embedded tissue. Tissue obtained from any brain bank should have appropriate consent and ethical permission to use the tissue. It is the responsibility of the experimenter to ensure that this is in place when tissue is obtained from a source.

For human tissues: The method can be used with wax-embedded or fresh-frozen tissue. Brain sections can be sectioned at a range of thickness from 5 to 30 μm, depending on the experimental needs, although the use of 30 μm sections combined with free-floating immunohistochemistry (*see* Subheading 3) is highly encouraged.

2.2 Buffers and Solutions

1. Citrate buffer: Mix 2.1 g of citric acid in 1 L of distilled water (dH_2O). Adjust pH to 6.0 with NaOH. Store at 4 °C.
2. Phosphate buffered saline with Tween 20 (PBST 0.1/0.2): Prepare a stock solution of PBS (10×) by dissolving 80 g of NaCl, 2 g of KCl, 26.8 g of $Na_2HPO_4{\cdot}7H_2O$ and 2.4 g of KH_2PO_4 in 800 mL of dH_2O. Adjust volume to 1 L with dH_2O. Adjust pH to 7.4 with HCl or NaOH when diluted to (1×) (PBS 1×). Add 0.1 or 0.2 % (v/v) of Tween 20 to the PBS solution and mix gently to get the final 'PSBT0.1' and 'PBST0.2' solutions. Store at room temperature (RT).
3. Mowiol/DABCO mounting medium for immunofluorescence: Combine 2.4 g of Mowiol 4-88 (e.g. Sigma-Aldrich), with 6 g of glycerol and 6 mL of H_2O. Mix for approx. 3 h. Add 12 mL of 0.2 M Tris–HCl (pH 8.5). Incubate with mixing at 50 °C until it dissolves. Centrifuge at 5,000 × *g* for 15 min to pellet insoluble material. Add 1,4-diazabicyclo-[2,2,2]-octane (DABCO) as antibleaching agent (to reduce fading of fluorophores) to a final concentration of 2.5 % (w/v). Store in 500 μL aliquots at −20 °C.
4. Fluorescence quenching solution: 0.1 % (w/v) Mix Sudan Black in 70 % ethanol. Mix and filter. Store solution at RT, protected from light.

2.3 Reagents and Other Components

1. Liquid chemicals: Ethanol; xylene; 2 N HCl.
2. Solid chemicals: Bovine serum albumin (BSA); DAPI (4′,6-diamidino-2-phenylindole dihydrochloride).
3. Serum from the host animal of the secondary antibody to be used (*see* Subheading 3).
4. Blocking solution: 5 % serum, 5 % BSA in PBST0.2.
5. Incubation chamber or tray.
6. Free-floating incubation plate: Starting from a plastic cell culture plate, divide each well into two chambers with a stainless metallic mesh adhered to the bottom and the sides of the well. The tissue sections are incubated in free-floating in one chamber. Washes and incubations are done through the communicating chamber (*see* **Note 3**).
7. ImmEdge Hydrophobic Barrier Pen (Vector Labs), to provide a heat-stable, water-repellent barrier that keeps reagents localized on tissue specimens (*see* **Note 3**).
8. Glass slides coated with gelatin or APES (3-aminopropyltriethoxysilane). Alternatively, use ionized slides (*see* **Note 3**).

2.4 Primary and Secondary Antibodies

2.4.1 Primary Antibodies (Recommended)

1. Microglial markers: Rabbit anti-Iba1 (Wako); goat anti-Iba1 (Abcam); rat anti-CD11b (ABD Serotec); rabbit anti-PU.1 (Cell Signaling).
2. Proliferation markers: Mouse anti-BrdU (Developmental studies Hybridoma Bank); rat anti-BrdU (Santa Cruz Biotechnologies); rabbit anti-PCNA (Abcam); rabbit anti-phospho Histone H3 (Cell Signaling); rabbit anti-Ki67 (Abcam).
3. Other: Chicken anti-GFP (Abcam).

2.4.2 Secondary Antibodies

Biotinylated, affinity purified, secondary antibodies (Vector Labs), and fluorescence-conjugated (Alexa 405, 488 or 594 recommended) secondary antibodies, or streptavidin (Life technologies).

3 Methods

Unless otherwise specified, carry out all procedures at room temperature.

3.1 Immunohistochemical Detection of Microglial Proliferation in AD Mouse Models

1. Wash sections three times with PBST0.1 buffer, 5 min each (*see* **Note 3**).
2. [*Only if detecting BrdU*]. *DNA denaturation step for BrdU detection*: Incubate with 2 N HCl for 30 min at 37 °C. This step will provide access of the anti-BrdU antibodies to its epitope in the DNA (*see* **Note 4**).
3. [*Only if detecting BrdU*]. Wash with PBST0.1 buffer three times, 5 min each.
4. *Blocking*: Incubate with blocking solution (5 % serum, 5 % BSA in PBST0.2) for 1 h. This incubation will prevent unspecific binding of the primary or secondary antibodies to the tissue. Note it is not necessary to wash after the incubation.
5. *Primary antibodies*: Incubate with primary antibodies (choose one microglial marker (i.e. Iba1) and one marker of proliferation (i.e. BrdU), from different hosts) at manufacturer's recommended dilution in blocking solution, at 4 °C overnight (*see* **Note 5**).
6. Wash with PBST0.1 buffer three times, 5 min each.
7. *Secondary antibodies*: Incubate with appropriate fluorescent secondary antibodies at manufacturer's recommended dilution in blocking solution for 1 h. From this step, sections will need to be protected from light (*see* **Note 6**).
8. Wash with PBST0.1 buffer three times, 5 min each.
9. *Counterstain* with DAPI: Incubate with DAPI (1:2,000) in PBST0.1, 10 min if the blue channel is available (*see* **step 7** and **Note 6**). Nuclear staining will provide anatomical reference

and will also define the nuclear compartment to better identify proliferation-related markers.

10. Wash with PBST0.1 buffer three times, 5 min each.
11. *Mounting and coverslipping*: Use mowiol/DABCO mounting medium (*see* **Note 7**). Store slides at 4 °C, protected from light until imaging.

3.2 Immunohistochemical Detection of Microglial Proliferation in Post-mortem Tissue from AD Patients

Samples from post-mortem human tissue are usually obtained from brain banks as paraffin-embedded tissue. In case human sections are obtained by alternative preservation methods please omit **steps 1** and **2**. Tissue samples obtained from any brain bank should have appropriate consent and ethical permission to be used.

1. *Dewaxing and rehydration*. Transfer the slides with the samples to a rack and incubate 40 min at 60 °C in an oven. After heating, directly transfer slides to xylene (15 min), followed by sequential incubation in the rehydrating solutions (100, 95, 80 and 75 % ethanol, ending with dH_2O; 5 min each). Wash three times in PBS, 5 min each.
2. *Antigen retrieval*: Transfer slides to a plastic rack and cover with excess citrate buffer (to prevent drying due to evaporation). Heat at full power in a microwave for 25 min. Then, transfer quickly to cold running tap water.
3. Wash with PBST0.1 buffer three times, 5 min each (*see* **Note 3**).
4. *Blocking*: Incubate with blocking solution for 1 h. This incubation will prevent unspecific binding of the primary or secondary antibodies to the tissue. Note it is not necessary to wash after the incubation.
5. *Primary antibody*: Incubate with primary antibodies (choose one microglial marker (i.e. Iba1) and one marker of proliferation (i.e. Ki67), from different hosts) at manufacturer's recommended dilution in blocking solution. Incubate overnight at 4 °C.
6. Wash with PBST0.1 buffer three times, 5 min each.
7. *Quenching autofluorescence step*: Incubate with fluorescence quenching solution (Sudan Black) for 10 min. This step is particularly important in the case of AD human tissue, as the occurrence of autofluorescent artefacts (e.g. lipofuscin granules) is very frequent and can confound the interpretation of results.
8. Wash with PBST0.1 buffer three times, 5 min each.
9. *Secondary antibodies*: Incubate with appropriate fluorescent secondary antibodies at manufacturer's recommended dilution in blocking solution for 1 h. From this step, sections will need to be protected from light (*see* **Note 6**).
10. Wash with PBST0.1 buffer three times, 5 min each.

11. *Counterstain* with DAPI: Incubate with DAPI (1:2,000) in PBST0.1, for 10 min if blue channel is available (*see* **step 9** and **Note 6**). Nuclear staining will provide anatomical reference and will also define the nuclear compartment to better identify proliferation-related markers.
12. Wash with PBST0.1 buffer three times, 5 min each.
13. *Mounting and coverslipping*: Use mowiol/DABCO mounting medium (*see* **Note 7**). Store slides at 4 °C, protected from light until imaging.

4 Notes

1. If allowed by the experimental conditions, the use of birthdating studies with thymidine analogues (e.g. tritiated-thymidine birthdating) is highly recommended. To date, we recommend one single window of proliferation using BrdU (50 mg/kg body weight, in 0.9 % (w/v) sterile solution of NaCl (sterile saline)), administered by intraperitoneal injection. Each dose of BrdU will label approximately 2–3 h of proliferation, so we recommend using cumulative dosage paradigms (i.e. three to four consecutive injections at 3 h intervals), to maximize the readout of proliferating microglia and facilitate the analysis and quantification. Multiple windows of proliferation can be differentiated by sequentially administering complementary analogues, such as CldU, IdU or EdU, with detection methods similar to that of BrdU [20, 21].
2. In case of using BrdU for the detection of proliferation, avoid long post-fixation times (no longer than 2 h at 4 °C) which might interfere with the accessibility to the BrdU epitope in the DNA. If long post-fixation is necessary due to experimental needs, add a step of antigen retrieval (*see* Subheading 3.2) to the method, before DNA denaturation.
3. Immunohistochemical detection of microglial proliferation can be performed on sections mounted on glass slides or on free-floating sections in incubation plates (encouraged). In the first case, start by tracing an area around the section with ImmEdge pen, to limit diffusion of buffers (*see* Subheading 2).
4. When using tissue from transgenic reporter mice (i.e. macgreen or CX3CR1-EGFP), the DNA denaturation step required for BrdU detection might eliminate the native fluorescence from the enhanced green fluorescent protein (EGFP). We suggest using anti-GFP primary antibodies combined with secondary antibodies coupled to green fluorescence to retrieve the EGFP signal (Fig. 1).
5. As a complementary study, we strongly recommend analysing the expression of the different components of the main pathway

regulating microglial proliferation: the activation of CSF1R. The expression of the transcription factor PU.1 is specific for microglia, and correlates with the proliferative status. Also, analysing the expression levels of CSF1R (c-fms), CSF1 or IL34 by immunohistochemistry will inform about the proliferative activity of microglia [13].

6. The immunohistochemical method can be adapted to the specific experimental aims allowing, for example, the simultaneous detection of up to four epitopes using conventional imaging methods. In these cases, matching each primary antibody with a specific color of the fluorescent-coupled secondary antibody will depend on the expected intensity for each epitope. Thus, green fluorescence is usually better registered by conventional microscopes, so it will be used for the antigen expected to have the worse signal. Signals expected to be optimal and intense will be assigned to the red or blue channels. If required, biotin-conjugated secondary antibodies could be used, bridging to fluorescence with the use of fluorescent-coupled streptavidin conjugates (streptavidin-biotin binding enables detection of biotinylated antibodies).
7. If using free-floating immunohistochemistry, sections will need to be previously transferred to gelatine-coated or ionized glass slides will the help of a paintbrush.

Acknowledgements

The research was funded by the European Union Seventh Framework Programme under grant agreement IEF273243, from Alzheimer Research UK and from the Medical Research Council (MRC). The authors have no conflicting financial interests.

References

1. Ransohoff RM, Perry VH (2009) Microglial physiology: unique stimuli, specialized responses. Annu Rev Immunol 27:119–145
2. Heneka MT, O'Banion MK (2007) Inflammatory processes in Alzheimer's disease. J Neuroimmunol 184:69–91
3. Akiyama H, Barger S, Barnum S et al (2000) Inflammation and Alzheimer's disease. Neurobiol Aging 21:383–421
4. Edison P, Archer HA, Gerhard A et al (2008) Microglia, amyloid, and cognition in Alzheimer's disease: an [11C](R)PK11195-PET and [11C]PIB-PET study. Neurobiol Dis 32:412–419
5. Dickson DW, Lee SC, Mattiace LA et al (1993) Microglia and cytokines in neurological disease, with special reference to AIDS and Alzheimer's disease. Glia 7:75–83
6. Fernandez-Botran R, Ahmed Z, Crespo FA et al (2011) Cytokine expression and microglial activation in progressive supranuclear palsy. Parkinsonism Relat Disord 17:683–688
7. Mildner A, Schlevogt B, Kierdorf K et al (2011) Distinct and non-redundant roles of microglia and myeloid subsets in mouse models of Alzheimer's disease. J Neurosci 3:11159–11171
8. Lawson LJ, Perry VH, Gordon S (1992) Turnover of resident microglia in the normal adult mouse brain. Neuroscience 48:405–415
9. Prinz M, Mildner A (2011) Microglia in the CNS: immigrants from another world. Glia 59:177–187

10. Ginhoux F, Greter M, Leboeuf M et al (2010) Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science 330:841–845
11. Ajami B, Bennett JL, Krieger C et al (2007) Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. Nat Neurosci 10:1538–1543
12. Mildner A, Schmidt H, Nitsche M et al (2007) Microglia in the adult brain arise from Ly-6ChiCCR2+ monocytes only under defined host conditions. Nat Neurosci 10:1544–1553
13. Gomez-Nicola D, Fransen NL, Suzzi S, Perry VH (2013) Regulation of microglial proliferation during chronic neurodegeneration. J Neurosci 33:2481–2493
14. Akiyama H, Nishimura T, Kondo H et al (1994) Expression of the receptor for macrophage colony stimulating factor by brain microglia and its upregulation in brains of patients with Alzheimer's disease and amyotrophic lateral sclerosis. Brain Res 639: 171–174
15. Bolmont T, Haiss F, Eicke D et al (2008) Dynamics of the microglial/amyloid interaction indicate a role in plaque maintenance. J Neurosci 28:4283–4292
16. Frautschy SA, Yang F, Irrizarry M et al (1998) Microglial response to amyloid plaques in APPsw transgenic mice. Am J Pathol 152: 307–317
17. Kamphuis W, Orre M, Kooijman L et al (2012) Differential cell proliferation in the cortex of the APPswePS1dE9 Alzheimer's disease mouse model. Glia 60:615–629
18. Sasmono RT, Oceandy D, Pollard JW et al (2003) A macrophage colony-stimulating factor receptor-green fluorescent protein transgene is expressed throughout the mononuclear phagocyte system of the mouse. Blood 101: 1155–1163
19. Jung S, Aliberti J, Graemmel P et al (2000) Analysis of fractalkine receptor CX(3)CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. Mol Cell Biol 20:4106–4114
20. Gomez-Nicola D, Valle-Argos B, Pallas-Bazarra N, Nieto-Sampedro M (2011) Interleukin-15 regulates proliferation and self-renewal of adult neural stem cells. Mol Biol Cell 22:1960–1970
21. Llorens-Martin M, Trejo JL (2011) Multiple birthdating analyses in adult neurogenesis: a line-up of the usual suspects. Front Neurosci 5:76

Part III

Comprehensive Disease Models Recapitulating Alzheimer's Disease Features: From Cellular Models to Human

Chapter 11

Yeast as a Model for Alzheimer's Disease: Latest Studies and Advanced Strategies

Mathias Verduyckt, Hélène Vignaud, Tine Bynens, Jeff Van den Brande, Vanessa Franssens, Christophe Cullin, and Joris Winderickx

Abstract

The yeast *Saccharomyces cerevisiae*, a unicellular eukaryotic model, has enabled major breakthroughs in our understanding of a plethora of cellular and molecular processes. Today, a 're-invention' of its use in fundamental and applied research is paving the way for a better understanding of the mechanisms causing neurodegeneration. The increasing emergence of neurodegenerative disorders is becoming more and more problematic in our ageing society. Most prevalent is Alzheimer's disease (AD), affecting more than 35 million people worldwide (Abbott, Nature 475, S2–S4, 2011) and causing an enormous burden on a personal and communal level. The disease is characterized by two major pathological hallmarks: extracellular amyloid plaques consisting mainly of deposits of amyloid β (Aβ) peptides, and intracellular neurofibrillary tangles (NFTs), consisting mainly of aggregates of hyperphosphorylated tau protein. Despite the huge importance of thoroughly understanding the underlying molecular mechanisms of neurodegeneration, progress has been slow. However, multiple complementary research methods are proving their value, particularly with the work done with *S. cerevisiae*, which combines well-established, fast genetic and molecular techniques with the ability to faithfully capture key molecular aspects of neurodegeneration. In this review chapter, we focus on the considerable progress made using *S. cerevisiae* as a model system for Alzheimer's disease.

Key words *Saccharomyces*, Yeast, Model, Alzheimer's disease, Amyloid beta, Tau

1 Yeast as a Model System

Saccharomyces cerevisiae, also known as baker's and budding yeast, has historically proven to be instrumental in deciphering mechanisms underlying a variety of central, conserved, cellular and molecular eukaryotic processes. These include the regulation of the cell cycle, the secretory pathway, mitochondrial biology, gene interactions and recombination, among others. In 1996, *S. cerevisiae* was the first eukaryote to have its genome fully sequenced [1], with a very condensed genome, 12.1 Mbp with ca. 6,600 open reading frames (ORFs) annotated to date [1–3]. By comparison, the human genome features three to five times as many genes, but it is more

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_11,

than 250 longer (3,200 Mbp; 3.2 billion base pairs). Sixty percent of yeast genes show significant homology to human genes, or have at least one conserved domain, often implicated in signal transduction or specific metabolic processes [4]. Furthermore, approximately 30 % of genes known to be implicated in human diseases have a yeast ortholog [5]. Since the publication of the genome sequence, an enormous wealth of genome-wide information has become easily accessible in comprehensive databases (Table 1), and pave the way for *S. cerevisiae* to become the eukaryotic model of choice for the development of new genomic technologies.

As an experimental model system, *S. cerevisiae* offers a wide variety of tools and technological approaches [6]. Thus, the available deletion and overexpression libraries make experiments designed to uncover genetic interactions easily accessible, as well as studies examining the involvement of certain genes in a variety of processes and physiological responses. Large scale studies especially benefit from these collections. On the proteomics level, protein interactions can be discovered using techniques such as yeast-two-hybrid, TAP-TAG and co-immunoprecipitation experiments, as well as protein microarrays using e.g. nickel- or nitrocellulose-coated slides covered with tagged proteins. These can be used to distinguish protein-protein, protein-nucleic acid and protein-lipid interactions. Protein localization studies can be performed using GFP- or dsRed fusion proteins which can be expressed in yeast, as well as by endogenous tagging, where fluorescent tags can be inserted into the genome using homologous recombination. The yeast *S. cerevisiae* is especially well suited for high throughput screening studies. The availability of deletion [7] and overexpression libraries [8, 9] in combination with highly automated research methods allows to quickly assess phenotypes and effects of thousands of genes. A typical experimental design including the discovery of genes capable of modifying a toxic phenotype (e.g. due to a toxic Aβ peptide) is shown in Fig. 1. *S. cerevisiae* was also the first organism where genome-wide transcriptional profiling was performed, using cDNA microarrays, thus allowing systematic large scale profiling of mRNA levels in a cell population [10]. The biological effects of small molecules can also be easily investigated using yeast high throughput assays. A large amount of molecules can be tested simultaneously by utilizing automatized systems, and the yeast deletion collection can help to clarify the role of certain genes in the effect of a given molecule [11, 12].

For a long time, the use of yeast in human disease-related research was constricted to disease genes with a direct homologue in yeast. Recently however, 'humanized yeast' models, where human genes without a yeast homologue are introduced into yeast cells are becoming implemented as a valuable research tool for deciphering molecular causes underlying human disease [13–15]. In a research community where the use of a multitude of model

Table 1
Websites and bioinformatics tools of interest in yeast research

Databases and bioinformatics tools	URL
General yeast genome and proteome databases	
Saccharomyces Genome Database (SGD, Stanford)	http://www.yeastgenome.org/
Comprehensive Yeast Genome Database (CYGD-MIPS)	http://mips.gsf.de/genre/proj/yeast/index.jsp
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/
Yeast mutant collections	
Saccharomyces Genome Deletion Project	http://www-sequence.stanford.edu/group/yeast_deletion_project/
EUROpean Saccharomyces Cerevisiae ARchive for Functional analysis (EUROSCARF)	http://web.uni-frankfurt.de/fb15/mikro/euroscarf/
Yeast-mammalian and yeast-human homology search tools	
Mammalian homology to yeast (SGD)	http://www.yeastgenome.org/mammal/
Clusters of orthologous groups of proteins (COGs)	http://www.ncbi.nlm.nih.gov/COG/
Discover homologs (Homologene)	http://www.ncbi.nlm.nih.gov/homologene
Yeast tools for studies on human diseases	
Yeast homologs of human disease-associated genes	http://mips.gsf.de/proj/yeast/reviews/human_diseases.html
Mitochondria-related proteins, genes and diseases (MitoP)	http://www.mitop.de:8080/mitop2/
Yeast proteome analysis	
Yeast Protein Localisation database (YPL.db)	http://ypl.uni-graz.at/pages/home.html
Yeast GFP Fusion Localization database (yeastgfp)	http://yeastgfp.yeastgenome.org/
Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu
Molecular Interactions Database (MINT)	http://160.80.34.4/mint/
Information Hyperlinked Over Proteins	http://www.ihop-net.org/UniPub/iHOP/
Yeast expression analysis	
Princeton Microarray database	http://puma.princeton.edu/
Yeast Microarray Global Viewer (yMGV)	http://www.transcriptome.ens.fr/ymgv/
Yeast phenotypic analysis	
Saccharomyces cerevisiae Morphological Database (SCMD)	http://yeast.gi.k.u-tokyo.ac.jp/
PROfiling of PHEnotypic Characteristics in Yeast (PROPHECY)	http://prophecy.lundberg.gu.se/

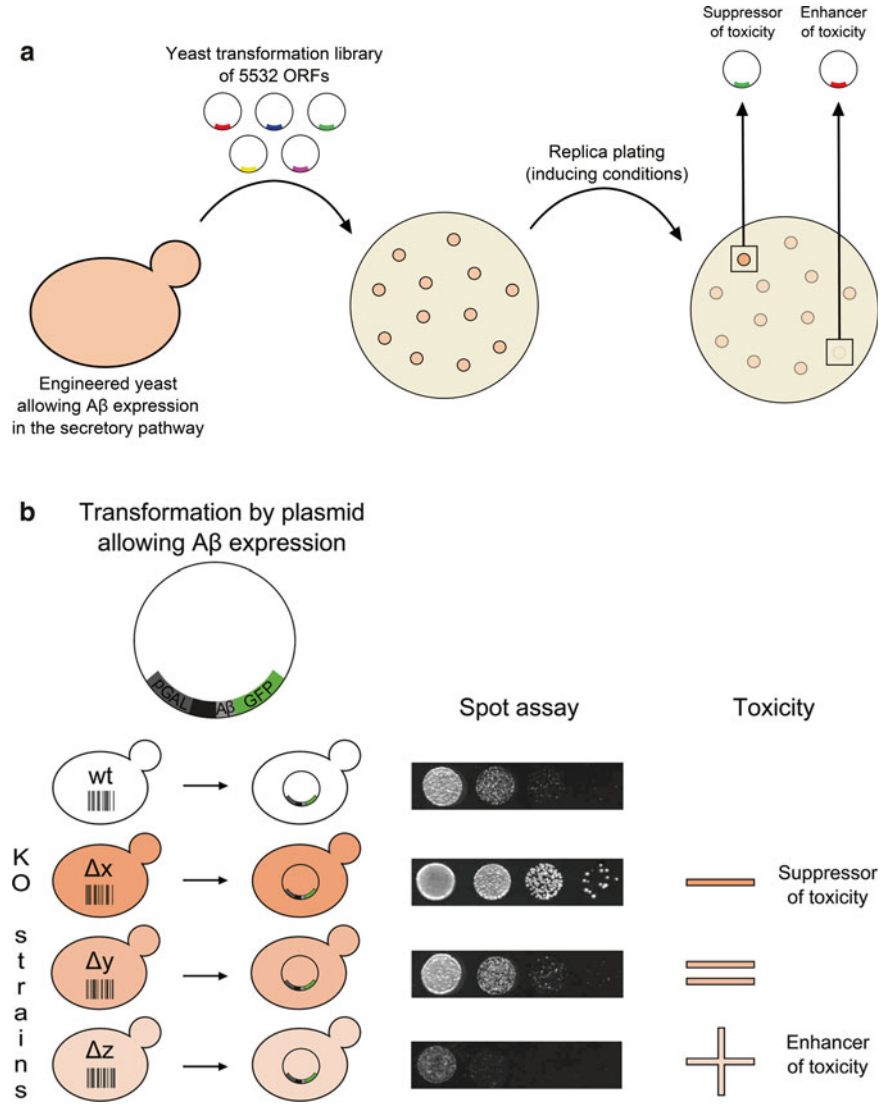


Fig. 1 High-throughput experiments can easily be performed in *Saccharomyces cerevisiae*. (a) A yeast strain expressing an inducible toxic Aβ construct is transformed by a library of 5,532 ORFs. After replica-plating from non-inducing to inducing conditions, plasmids from transformants displaying additional toxicity or an increase in growth can be isolated and sequenced, in order to identify the responsible ORF. (b) In order to further analyse the effect of deletion or overexpression of selected genes, the Aβ construct can be transformed into the proper deletion or overexpression strains, after which the necessary experiments can be performed

systems is becoming more prevalent and accepted, the use of yeast to model complex diseases is often met with surprise or disbelief, even when the advantages of using yeast as an early screening tool are clear. While it is true that *S. cerevisiae* is a simple unicellular eukaryote, the reality of conservation of central mechanisms, essential pathways and networks in all eukaryotes, and the genetic

homology between yeast to human have to be considered. So far, yeast is contributing to the understanding of mechanisms underlying Alzheimer's disease, Parkinson's disease [16–18], Huntington's disease [19–21], amyotrophic lateral sclerosis (ALS) [22], proteopathies like cystic fibrosis [23] and several tauopathies [24].

2 Approaches to Modeling Aβ Toxicity in Yeast

2.1 Amyloid Precursor Protein (APP) Processing

Senile or amyloid plaques are the pathological hallmark of AD which have received the most attention. They consist mainly out of deposits of Aβ-peptides. These peptides are generated through the subsequent cleavage of the amyloid precursor protein (APP) by β- and γ-secretase, in the amyloidogenic pathway. Alternatively, APP can be processed via the non-amyloidogenic pathway, in which α-secretase cleaves APP inside the Aβ region (Fig. 2) [25]. APP is a type I transmembrane protein with putative functions related to signaling, cell adhesion and neuronal maturation and migration [26]. An iron-export ferroxidase function has been reported as well [27]. Processing by β-secretase releases the N-terminal ectodomain and leaves a membrane-bound C-terminal fragment called C99. This fragment can then be cleaved intramembraneously by γ-secretase (Fig. 3a) [28]. This cleavage is somewhat more promiscuous, and Aβ peptides ranging from 38 to 43 amino acids can be produced [29]. In healthy individuals, mainly $A\beta_{40}$ is generated, but in AD, the balance between $A\beta_{40}$ and $A\beta_{42}$ is shifted in favor of the longer form, which is more hydrophobic and more prone to aggregation [30]. In some familial AD, $A\beta_{42}$ is present in greater amounts due to mutations in either APP or presenilin, the catalytic subunit of γ-secretase. Mutations in the Aβ part of APP can change its aggregation properties, another cause of familial AD [31].

Earlier research focused mainly on the extracellular amyloid plaques, but more recently attention has shifted towards the intracellular effects of $A\beta_{42}$ [32]. APP is translocated into the ER and is transported through maturation in the Golgi complex to the plasma membrane. A significant fraction of Aβ can be reinternalized, and evidence suggests APP processing can take place in the secretory and endosomal pathway as well [33, 34]. There still are a lot of questions regarding what the toxic species is, but toxicity has been reported from a range of soluble Aβ oligomers. The large extracellular plaques are mainly inert and considered to be less toxic, although it might represent a reservoir for soluble Aβ. Exactly how the toxic oligomers cause cellular distress remains unclear, but it is suggested that the cytotoxicity might arise through inhibition of the proteasome [35], oxidative stress due to ROS production and damaged mitochondria [36, 37], changes in endocytic efficiency [38], disruption of Ca^{2+}-signaling [39] and alterations to synaptic receptor levels and activity [40].

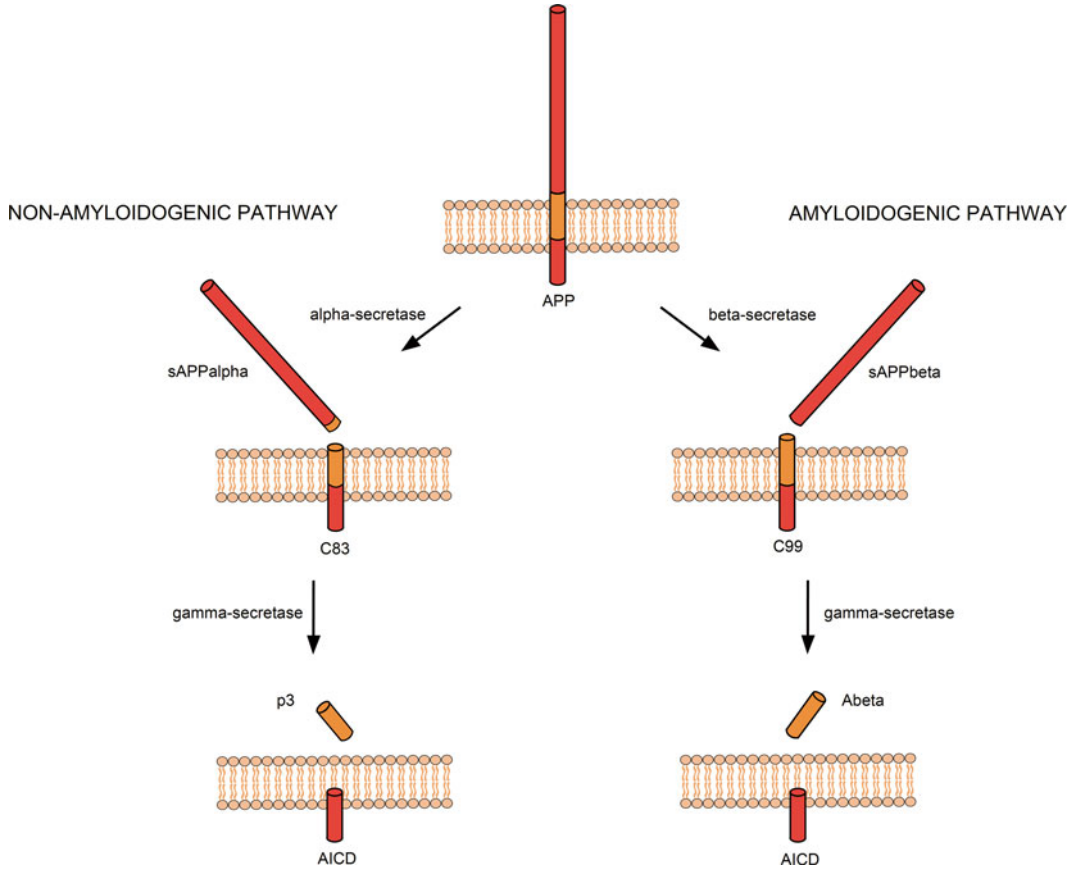


Fig. 2 Amyloid precursor protein APP processing. Non-amyloidogenic and amyloidogenic pathways. The amyloid precursor protein APP is targeted to the plasma membrane, where it can be cleaved, in the non-amyloidogenic pathway, by α-secretase, leading to the production of a soluble fragment sAPPalpha, which is released into the extracellular space. The C-terminal part of 83 amino acids length (C83) remains embedded in membrane and can subsequently be cleaved by γ-secretase, releasing the p3 fragment. In the amyloidogenic pathway, β-secretase cleavage produces a 99 amino acids fragment (C99) retained within the membrane. C99 is cleaved by γ-secretase present in the plasma membrane, but also in the ER, resulting in release of the Aβ peptide

Fig. 3 (continued) APP (b), the C99 fragment (c), Aβ alone (d) or fused to GFP (f) were targeted to the secretory pathway by the addition of the secretion signal of alpha factor (b, c and f) or KAR2 (d). These models allowed Zhang et al. [41] to identify "secretase-like" activities in yeast (b, Yap3 and Mkc7), whereas Sparvero et al. [45] put in light the essential role of the proteasome in the removal of aggregation prone species after expression of C99 (c). The toxicity models of Treusch et al. [47] (d) and D'Angelo et al. [50] (f) have highlighted the involvement of clathrin-dependent endocytosis (via Yap1802, Yap1801, *red star*) in Aβ toxicity, but also demonstrated a rapid alteration of cellular respiration following Aβ expression in the secretory pathway. Cytoplasmic models of Aβ aggregation using Aβ fused to GFP (e) or to the C-terminal part of Sup35 (lacking prion domain) (g), helped to underline the role of Hsp104 in Aβ aggregation and led to an efficient oligomerization assay

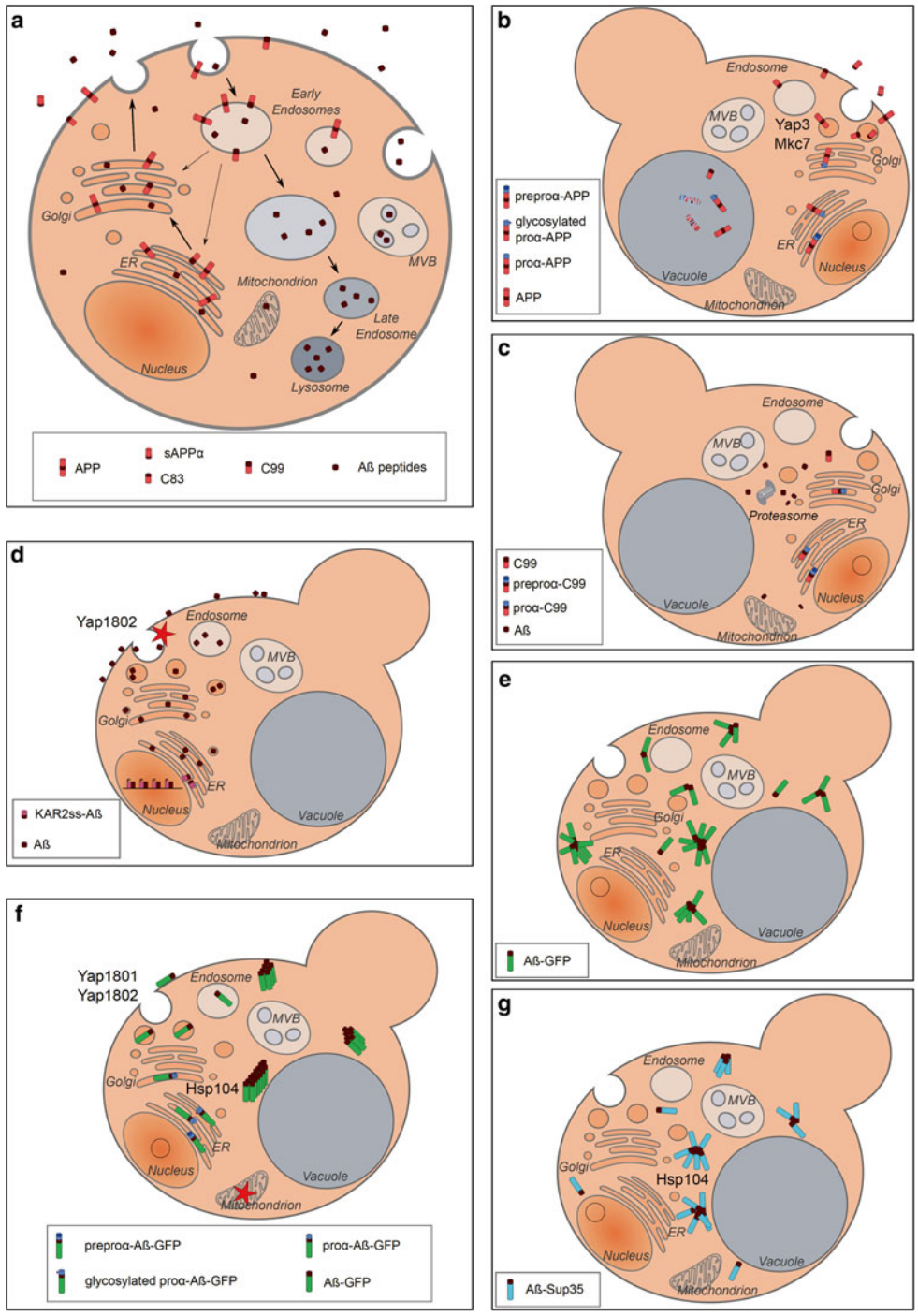


Fig. 3 APP processing and Aβ expression in neurons and yeast models. In neurons, the Aβ peptide is not only produced at the plasma membrane, but may also be generated in intracellular compartments such as the endoplasmic reticulum and the *trans*-Golgi (a). Also, extracellularly released peptides can be re-endocytosed. These different mechanisms allow to detect Aβ in multivesicular bodies (MVB), lysosomes, ER and Golgi, but also in mitochondria and the cytosol. APP processing, Aβ toxicity and aggregation were modeled in *S. cerevisiae*.

Initial yeast studies on Aβ focused on the processing of APP and C99. The first study from Zhang and colleagues expressed APP fused to the prepro-α-mating factor, which serves as a signal sequence, in a protease deficient yeast strain [41]. After Kex2-processing in the late-Golgi complex, cleaving off the α-factor, full-length APP could be detected. The researchers showed that APP underwent further processing, as an N-terminal ectodomain was released in the medium and a C-terminal fragment could be detected, the same size of the C-terminal fragment released by α-secretase cleavage in human cells. In a follow-up study, it was shown that two GPI-linked aspartyl proteases, encoded by YAP3 and MKC7, were responsible for this cleavage [42]. Cleavage did not occur in mutants where transport from the ER to the Golgi was blocked. Experimental evidence supports an α-secretase-like activity, most likely processing APP in the late Golgi (Fig. 3b). No endogenous β- or γ-secretase activity was observed in yeast cells, but strains have been engineered where APP fragments could be cleaved by human β-secretase (BACE1) or a reconstituted γ-secretase complex. In case of BACE1, an APP fragment containing the β-site, the transmembrane domain and the C-terminal domain was fused to yeast invertase. Upon expression of β-secretase, growth could be restored on selective plates [43]. For γ-secretase, the C_{1-55} fragment of APP was fused to the GAL4 transcription factor. Upon expression of all four γ-secretase subunits, GAL4 was released and triggered the transcription and translation of β-galactosidase, which can easily be detected by assaying using ONPG (*ortho*-nitrophenyl-β-galactoside). β-galactosidase activity could only be detected when all four subunits were expressed [44].

The proteasome is able to remove aggregation-prone peptides from the cell, and as it is suggested that proteasomal activity is impaired in AD [35]. Sparvero et al. assayed the processing of the C99 fragment both in wild type yeast, and in a yeast strain where two subunits of the proteasome were mutated, resulting in a severe impairment of its activity [45]. ZipTip immunocapture and mass spectrometry were used to analyze peptides reactive to an Aβ antibody. They found that when proteasomal activity is impaired, the proteomic fragment profiles of C99 were radically different, suggesting that other protein quality control mechanisms could compensate and act upon C99 (Fig. 3c). This response led to the production of peptides that were more hydrophobic and could thus aggregate more readily. Cells with a functioning proteasome showed a smaller number of fragments, which were generally less aggregation prone. Two larger, amyloidogenic species were however detected in the wild type yeast cells as well, so a small portion of proteasomal activity might contribute to the generation of toxic species.

2.2 Aβ Toxicity

Up until recently, studies attempting to express native $A\beta_{42}$ in *S. cerevisiae* failed to produce detectable levels of the peptide, either because of extremely rapid degradation or through counterselection due to Aβ-toxicity. It had already been reported that an Aβ-GFP fusion protein was able to slightly lower growth yields and induce heat shock responses in yeast [46]. In 2011 however, two groups succeeded in making a yeast model capable of expressing the native Aβ peptide. Treusch et al. fused the KAR2 signal sequence to the N-terminus of the Aβ sequence, codon optimized for expression in yeast (Fig. 3d) [47]. The function of KAR2 was to direct the peptide into the secretory pathway. As the yeast cell wall restrains any secreted peptides, Aβ remains in the periplasm and can interact with the plasma membrane and undergo endocytosis, trafficking through pathologically relevant cellular compartments. When expressed from a galactose-inducible multi-copy plasmid, Aβ caused a slight decrease in growth rate. In order to allow genetic screens, tandem copies of the construct were integrated in the genome, resulting in a robust cytotoxic effect, leading to a strain with impaired growth on galactose, without a major increase in lethality. An $A\beta_{40}$ construct was created as control, this peptide was less toxic for yeast cells. The native peptide could be detected using western blotting, and unboiled samples revealed oligomers of $A\beta_{42}$, and significantly less oligomers of $A\beta_{40}$, indicating that as in neurons, oligomeric species contribute to cytotoxicity. An overexpression library of 5,532 open reading frames (ORFs) was transformed into the $A\beta_{42}$ screening strain, displaying an intermediate Aβ-toxicity, so that toxicity enhancers or suppressors could be identified easily. Seventeen enhancers and 23 suppressors were discovered, many of which showed sequence similarity to human genes. Twelve hits had clear human orthologues, with three results involved in clathrin-mediated endocytosis and seven that were functionally associated with the cytoskeleton. Several of these genes had human homologues with a connection to AD risk factors, most notably the human homologue of YAP1802, PICALM, which is one of the most highly confirmed risk factors for sporadic AD [48, 49]. For selected hits, experiments have been carried out using transgenic *C. elegans* and rat cortical neurons. The outcome of these experiments supported results obtained from yeast, thus validating the model. Finally, the effect of $A\beta_{42}$ on clathrin-mediated endocytosis was looked into, by analyzing the localization of the Ste3-YFP fusion protein. In control cells, this protein localizes to the lumen of the vacuole, while endocytosis in $A\beta_{42}$-expressing cells was severely perturbed, as Ste3-YFP could be found in several foci. Interestingly, overexpression of three of the toxicity suppressor hits was in each case able to partially rescue the endocytic defect.

D'Angelo and colleagues were also able to validate a yeast model for $A\beta_{42}$-induced cytotoxicity [50]. A fusion between the prepro-α-mating factor, $A\beta_{42}$ and GFP caused clear and reproducible cytotoxicity (Fig. 3f). The α-mating factor-$A\beta_{42}$ fusion without GFP was cytotoxic as well, but western blot analysis revealed the expression of this construct was significantly lower. GFP appears to stabilize Aβ, which is otherwise quickly broken down, but is not responsible for any observed cytotoxic effect. A fusion protein containing the Arctic mutant of $A\beta_{42}$ showed an increase in toxicity when compared to wild type Aβ. The importance of trafficking through the secretory pathway was exemplified, as Aβ without prepro signal sequence did not cause any clear cytotoxicity. Interestingly, aggregates formed by Aβ with and without signal sequence appeared to be different, after filter-trap and fluorescent microscopic analysis. When expressed in the cytoplasm, without signal sequence, Aβ-GFP was distributed homogenously, and some foci could be seen (Fig. 3e). When expressed through the secretory pathway, there was no fluorescence at all, suggesting that aggregation of the protein somehow obstructed GFP to fold into its native state. A linker between Aβ and GFP was able to restore fluorescence. Respiration rates were monitored in aerobic conditions, and it was shown that the presence of Aβ caused a decrease in oxygen consumption. As the mitochondrial content was unchanged, it is plausible that this decrease might arise from an inhibition of the electron transport chain. As Hsp104 has been reported to play an important role in handling Huntingtin aggregates in yeast [21], the researchers tested whether it would affect Aβ toxicity. Somewhat surprisingly, deletion of Hsp104 partially restored the growth defect caused by expression of Aβ. As Hsp104 is a cytoplasmic protein, this observation suggests that at least part of the Aβ peptides are able to escape the secretory pathway and end up in the cytosol. No fluorescent foci were observed in the *hsp104Δ* strain, indicating that Hsp104 has an effect on the aggregation properties of Aβ. Deletion of YAP1801 and YAP1802, the yeast homologues of human PICALM, caused a slight decrease in toxicity, and expression of mouse PICALM was able to partially restore the toxic phenotype observed after expression of Aβ in wild type yeast. This result is somewhat at odds with results obtained by Treusch et al., where overexpression of YAP1802 rescued the cell from toxicity, due to an upregulation of endocytosis. This discrepancy might be due to a fundamental difference in both models: the model by Treusch et al. uses codon optimized tandem constructs integrated in the genome, which causes a very high production rate. D'Angelo et al. express their Aβ construct using a multi-copy plasmid, creating intermediate levels of Aβ. Another possibility is that PICALM might decrease toxicity when Aβ levels are very high, while increasing endocytosis leads to an inverse effect when the amount of Aβ is below a certain threshold. The fact that two different signal sequences are used might provide an explanation for some different

outcomes as well. Overall, both models clearly show how yeast can be used as a relevant tool for human disease research, replicating important pathological events as cytotoxicity, oligomerization and the involvement of genes with relevant human counterparts. In both cases, correct processing through the secretory pathway proved instrumental for Aβ to cause cytotoxicity in yeast.

2.3 Aβ Aggregation

Yeast models have been used as well to specifically study Aβ oligomerization. Aβ was fused to the C-terminal part of Sup35p, an essential translation termination factor (Fig. 3g) [51]. Sup35p is a known yeast prion protein, and its aggregation state can easily be assayed using the nonsense allele *ade1-14*, in which a premature stopcodon is introduced. Sup35p in its normal state will result in a truncated enzyme, rendering the cells unable to grow on synthetic medium without adenine. Additionally, they will accumulate a red intermediate of adenine biosynthesis when grown on complete medium. In its prion form [PSI+], the efficiency of translation termination at the premature stop codon is impaired, so the cells gain the ability to grow on medium without adenine and stop producing the red intermediate pigment. The N-terminal part of Sup35p is responsible for its prion properties, but is not necessary for the essential function of the protein. By fusing Aβ to the essential C-terminal part of Sup35p, without the prion domain, an easy oligomerization assay was created, which has been used to discover specific point mutations which inhibited Aβ oligomerization. Furthermore, consistent with the results of D'Angelo et al. [50], the assay demonstrated that Hsp104 could interact with Aβ. Deletion of Hsp104 inhibited the oligomerization process [51]. More recently, this assay has been used in a high-throughput screen in an effort to discover anti-oligomeric compounds [52]. A library of 12,800 small molecules was tested, and two relevant hits were identified. These two compounds were subjected to further biochemical analysis, which confirmed the anti-oligomeric effect, validating this screening method as a reliable and cost-effective approach to address these types of questions [52].

3 Yeast Approaches to Study Tau Biology

The tau protein is a microtubule-associated protein involved in the stabilization and spacing of microtubules (MT), which makes it an important factor in the regulation of axonal transport [53]. Apart from binding to microtubules, tau is also known to interact with other cellular components and enzymes, for example the plasma membrane [54, 55], actin filaments [56] and src tyrosine kinases like FYN [57–59]. Tau can exist in six different isoforms, arising from differential mRNA splicing. The isoforms differ in the presence or absence of two N-terminal inserts of which the function is still undetermined, and an additional repeat of a tubulin-binding motif in the

microtubule-binding domain, affecting the ability of tau to stabilize microtubules [53]. The binding of tau to microtubules is largely regulated by phosphorylation, and the interplay of various kinases and phosphatases creates a dynamic scenario where the stability of microtubules can be easily altered [60, 61]. On the longest tau isoform, 79 putative Ser/Thr phosphorylation sites can be found, and phosphorylation of 30 of these sites has already been reported.

In Alzheimer's disease, tau can be detected in a variety of aggregated forms, including intracellular paired-helical fragments (PHFs) and neurofibrillary tangles (NFTs) [62]. Tau protein in these aggregates is hyperphosphorylated, which has resulted in increasing interest in tau phosphorylation studies [24]. The hyperphosphorylation changes the conformation of tau, facilitating oligomerization and aggregation (gain of function), and leading to the release of tau from microtubules, thus destabilizing them (loss of function) (Fig. 4a). In vitro studies support the notion that hyperphosphorylation of tau facilitates aggregation, but the exact phosphorylation sites important for the process remain elusive, as well as the mechanisms by which tau eventually causes cytotoxicity and cell death.

A few studies have used yeast to study the biology of tau, but their results show the potential and reliability of yeast models, as key aspects of tau pathophysiology such as phosphorylation, conformational change and aggregation can be faithfully recapitulated [63]. When human isoforms containing three or four microtubule-binding repeats and tau (3R-tau and 4R-tau) are expressed in *S. cerevisiae*, phosphorylation on pathologically relevant tau epitopes can be detected using phosphospecific antibodies, proving the existence of yeast kinases that are able to recognize and phosphorylate tau [64]. In addition, tau could also be detected using the conformation-dependent antibody MC-1, a marker for pathological tau filaments and their precursors [65–67]. In part, tau could be detected in the sarkosyl-insoluble fraction (SinT—*sarkosyl-insoluble t*au), indicating tau aggregated species [64]. Importantly, these characteristics were found to be modulated by yeast tau kinases Mds1 and Pho85, orthologues of human kinases GSK-3β and cdk5 respectively. Both are able to phosphorylate tau directly, and evidence suggests that cdk5 is also able to affect tau phosphorylation

Fig. 4 (continued) determined by immunodetection with phosphospecific antibodies. Sarkosyl-insoluble tau (SinT) assays also determined that a minor fraction of tau is insoluble in sarkosyl, a measurement for aggregated species. In *mds1Δ* and *pho85Δ* deletion strains, respectively the orthologues of human kinases GSK-3β and cdk5, these characteristics were affected, with less phosphorylation and aggregated species in *mds1Δ* and a significant increase in phosphorylation and aggregation levels in *pho85Δ*. Interestingly, when applying oxidative stress, phosphorylation levels were slightly lower than compared to the wild type, while aggregation was increased, suggesting a second, alternative pathway leading to tau aggregation

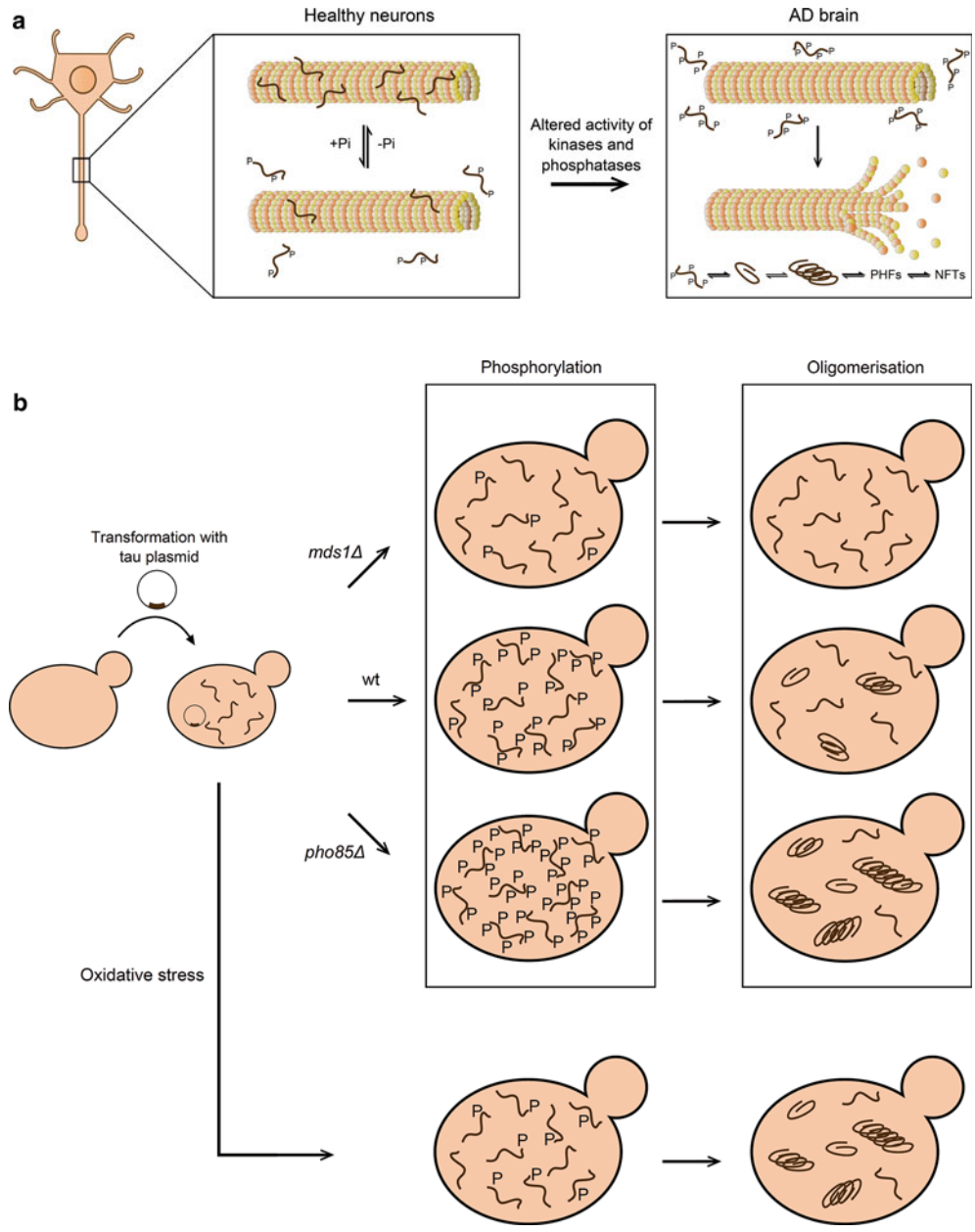


Fig. 4 Tau pathology in neurons and yeast models. (a) In healthy neurons, tau binds to and stabilizes microtubules. This binding can be regulated by differential levels of tau phosphorylation, creating a very dynamic scenario where the cell can easily alter the stability of microtubules, important for efficient transmission of signals. In neurons affected by AD however, tau is hyperphosphorylated due to an altered activity of diverse kinases and phosphatases, impairing its ability to bind microtubules. Hyperphosphorylation of tau changes its conformation and causes tau to oligomerize and eventually to aggregate into paired helical fragments (PHFs) and neurofibrillary tangles (NFTs). The microtubules on the other hand are destabilized and will begin to depolymerize, leading to a disruption of the cytoskeleton and severe problems in signal transmission. (b) In *Saccharomyces cerevisiae*, the tau protein can be expressed after transformation with a vector containing a tau construct. In a wild type strain, a fraction of tau is phosphorylated on relevant pathological epitopes, as

indirectly, by an inhibition of GSK-3β [68]. When Mds1 was deleted in *S. cerevisiae*, phosphorylation on the AD2 (P-S396-P-S404) and PG-5 (P-S409) epitopes was decreased significantly. This was expected for the AD2 epitope, as this is a direct target of GSK-3β in mammals [69]. PG-5, however, is not a typical target of GSK-3β, but of PKA, which might indicate that Mds1 is able to affect phosphorylation on the PG-5 epitope indirectly [70, 71]. Interestingly, deleting Pho85 resulted in a significant increase in immunoreactivity of the phosphospecific antibodies. This hyperphosphorylation was accompanied by an increase of MC-1 immunoreactivity and the presence of higher tau levels in the sarkosyl-insoluble fraction [64] (Fig. 4b). This observation links phosphorylation on these two epitopes to tau aggregation, and supports the idea that, like cdk5 in mammals [72], Pho85 is able to affect tau phosphorylation indirectly, by acting as a negative regulator of phosphorylation, and thus conformational changes and aggregation.

In vitro studies revealed further characteristics of hyperphosphorylated tau using two different techniques [64, 73]. Thus, soluble tau isolated from wild type (WT), *mds1Δ* and *pho85Δ* strains retained its phosphorylation state, and it was shown that filament formation happens significantly faster with tau isolated from the *pho85Δ* strain, consisted with its hyperphosphorylated state. Moreover, upon further fractionation, an MC-1 positive fraction could be obtained, and addition of these species to soluble tau vastly accelerated tau aggregation. This observation is consistent with a seeding capacity of hyperphosphorylated tau. In a second experiment, the in vitro binding capacity of yeast-isolated tau to mammalian microtubules was tested, using taxol-stabilized microtubules consisting of pig tubulin. An inverse correlation between phosphorylation state and the ability to bind and stabilize microtubules could be demonstrated, as tau isolated from the *pho85Δ* strain showed the poorest MT binding. Tau expressed in the wild type strain performed better in this assay, followed by *mds1Δ* tau, which showed an impaired phosphorylation state. Binding of tau to yeast tubulin has not been shown, most likely due to differences between yeast and mammalian tubulin.

Several clinical FTDP-17 (frontotemporal dementia and Parkinsonism linked to chromosome 17) tau mutants were expressed in WT, *mds1Δ* and *pho85Δ* yeast strains, and their phosphorylation patterns and SinT levels were analyzed [74]. The effect of these FTDP mutations has been speculated to range from altering the ratio between 3R and 4R-tau isoforms by influencing splicing efficiency, affecting the microtubule binding capacities of tau, to changing the conformation of tau [24, 62, 75]. Interestingly, both P301L and R406W tau mutants showed a decrease in phosphorylation at the PG5 epitope compared with the wild type tau, and SinT levels were reduced. These findings again suggested the

importance of phosphorylation at the PG5 site in triggering tau phosphorylation. To confirm this, the PG5 epitope was mutated and the results of expression of the synthetic S409A- and pseudo-phosphorylated S409E-mutants were analyzed. The S409A mutant showed a reduction in aggregation, while the SinT levels of S409E tau were increased or comparable to the wild type. Furthermore, phosphorylation of the PG5 epitope was also revealed to be detrimental for tau-microtubule interactions. Phosphoepitope mapping of tau bound to microtubules and soluble tau revealed the presence of several phosphoepitopes in the microtubule-bound fraction, while the PG5 epitope was almost completely absent, revealing an inverse relation between microtubule binding abilities and the propensity to aggregate of tau. In all, these results point towards S409 phosphorylation as a relevant factor determining physiological and pathological tau function. Interestingly, the S409A tau mutant showed a decrease in immunoreactivity at the AD2 epitope, and pseudophosphorylated S409E tau conversely showed an increase of phosphorylation at this site. Hence, it appears that phosphorylation at S409 and S396/S404 sites is interdependent, and that S409 phosphorylation might prime subsequent phosphorylation at the AD2 site [74].

As increasing evidence suggests that oxidative damage plays a role in the development of neurodegenerative disorders [76–79]. Thus, the effect of oxidative stress and mitochondrial dysfunction on SinT formation of wild type and mutant tau was assessed [74]. Oxidative stress was applied by adding ferrous sulfate to the medium, which increases free radical formation. SinT levels in cells that had undergone this treatment were markedly increased, in particular for FTDP mutants, demonstrating that these mutants render tau more prone to aggregation when subjected to oxidative stress. Interestingly, tau phosphorylation was decreased following ferrous sulfate treatment, especially at the AD2 and PG5 epitopes, suggesting that oxidative stress triggers an aggregation mechanism parallel to hyperphosphorylation (Fig. 4b). Moreover, oxidative stress treatment increased SinT levels of S409A tau to levels comparable with wild type and S409E tau in untreated cells. This again indicates that PG5 site phosphorylation is not a prerequisite for aggregate formation. As oxidative stress and mitochondrial dysfunction are closely linked, two deletion strains, i.e. *sod2Δ*, lacking mitochondrial manganese-dependent superoxidase dismutase [80] and *rim1Δ*, lacking a single-stranded DNA-binding protein essential for mitochondrial genome maintenance [81] were analyzed. Similar to the results obtained in the oxidative stress experiments, SinT levels for tau 2N/4R were significantly elevated. The increase in SinT for S409A and S409E tau was in the same range, once again indicating that oxidative stress and hyperphosphorylation might be working synergistically on tau aggregation.

At this point, it is important to note that tau exhibits important features like hyperphosphorylation and aggregation in *S. cerevisiae*, with no toxic effect observed in exponentially growing cells, implying that aggregation is not necessarily linked to cytotoxicity. However, because experiments were performed on cells in the exponential phase, important factors like mitochondrial dynamics and oxidative stress were somewhat excluded. It is possible that tau does have a toxic effect in aging cell populations, affecting the longevity of the yeast culture.

4 Concluding Remarks and Future Perspectives

The excellent studies performed in recent years presented here show the potential of yeast to decipher essential molecular mechanisms underlying neurodegeneration. It is clear that relevant elements of neurodegenerative diseases cannot be faithfully recapitulated in a simple unicellular eukaryote such as *S. cerevisiae* (e.g. supra-cellular features, synapses; spreading of the disease). However, the high level of homology with the human genome and the conservation of central essential pathways relevant to human disease makes yeast a powerful starting point, reference model with its simple techniques of cultivation and molecular manipulation. More specifically, advanced yeast models open the way to the study of potential links between the production and toxicity of Aβ and the interplay with tau modifications and tau biology. New robust and reliable models for Aβ and tau are being developed, with exciting developments regarding the underlying molecular mechanisms of cytotoxicity ahead. The involvement of several groups of genes and networks is already proved, and the results can be confirmed in higher organisms and finally in human. We strongly believe that the inclusion and integration of discoveries from several model systems, from yeast to human is the way forward.

References

1. Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 genes. Science 274(546): 563–567
2. Suter B, Auerbach D, Stagljar I (2006) Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. Biotechniques 40:625–644
3. Christie KR, Hong EL, Cherry JM (2009) Functional annotations for the Saccharomyces cerevisiae genome: the knowns and the known unknowns. Trends Microbiol 17:286–294
4. Botstein D, Chervitz SA, Cherry JM (1997) Yeast as a model organism. Science 277: 1259–1260
5. Foury F (1997) Human genetic diseases: a cross-talk between man and yeast. Gene 195:1–10
6. Sherman F (2002) Getting started with yeast. Methods Enzymol 350:3–41
7. Winzeler EA, Shoemaker DD, Astromoff A et al (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285:901–906
8. Hu Y, Rolfs A, Bhullar B et al (2007) Approaching a complete repository of sequence-verified protein-encoding clones for Saccharomyces cerevisiae. Genome Res 17: 536–543

9. Jones GM, Stalker J, Humphray S et al (2008) A systematic library for comprehensive overexpression screens in Saccharomyces cerevisiae. Nat Methods 5:239–241
10. DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686
11. Hoon S, St Onge RP, Giaever G et al (2008) Yeast chemical genomics and drug discovery: an update. Trends Pharmacol Sci 29:499–504
12. Parsons AB, Lopez A, Givoni IE et al (2006) Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell 126:611–625
13. Mager WH, Winderickx J (2005) Yeast as a model for medical and medicinal research. Trends Pharmacol Sci 26:265–273
14. Khurana V, Lindquist S (2010) Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? Nat Rev Neurosci 11:436–449
15. Winderickx J, Delay C, De Vos A et al (2008) Protein folding diseases and neurodegeneration: lessons learned from yeast. Biochim Biophys Acta 1783:1381–1395
16. Franssens V, Bynens T, Van den Brande J et al (2013) The benefits of humanized yeast models to study Parkinson's disease. Oxid Med Cell Longev 2013:760629
17. Zabrocki P, Pellens K, Vanhelmont T et al (2005) Characterization of alpha-synuclein aggregation and synergistic toxicity with protein tau in yeast. FEBS J 272:1386–1400
18. Swinnen E, Buttner S, Outeiro TF et al (2011) Aggresome formation and segregation of inclusions influence toxicity of alpha-synuclein and synphilin-1 in yeast. Biochem Soc Trans 39:1476–1481
19. Krobitsch S, Lindquist S (2000) Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. Proc Natl Acad Sci U S A 97:1589–1594
20. Muchowski PJ, Schaffar G, Sittler A et al (2000) Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. Proc Natl Acad Sci U S A 97:7841–7846
21. Meriin AB, Zhang X, He X et al (2002) Huntington toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1. J Cell Biol 157:997–1004
22. Bastow EL, Gourlay CW, Tuite MF (2011) Using yeast models to probe the molecular basis of amyotrophic lateral sclerosis. Biochem Soc Trans 39:1482–1487
23. Fu L, Sztul E (2009) ER-associated complexes (ERACs) containing aggregated cystic fibrosis transmembrane conductance regulator (CFTR) are degraded by autophagy. Eur J Cell Biol 88:215–226
24. Sergeant N, Delacourte A, Buee L (2005) Tau protein as a differential biomarker of tauopathies. Biochim Biophys Acta 1739:179–197
25. Esler WP, Wolfe MS (2001) A portrait of Alzheimer secretases—new features and familiar faces. Science 293:1449–1454
26. De Strooper B, Annaert W (2000) Proteolytic processing and cell biological functions of the amyloid precursor protein. J Cell Sci 113: 1857–1870
27. Duce JA, Tsatsanis A, Cater MA et al (2010) Iron-export ferroxidase activity of beta-amyloid precursor protein is inhibited by zinc in Alzheimer's disease. Cell 142:857–867
28. Kopan R, Ilagan MX (2004) Gamma-secretase: proteasome of the membrane? Nat Rev Mol Cell Biol 5:499–504
29. Steiner H, Fluhrer R, Haass C (2008) Intramembrane proteolysis by gamma-secretase. J Biol Chem 283:29627–29631
30. Turner PR, O'Connor K, Tate WP et al (2003) Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog Neurobiol 70:1–32
31. Ertekin-Taner N (2007) Genetics of Alzheimer's disease: a centennial review. Neurol Clin 25:611–667
32. LaFerla FM, Green KN, Oddo S (2007) Intracellular amyloid-beta in Alzheimer's disease. Nat Rev Neurosci 8:499–509
33. Suzuki T, Araki Y, Yamamoto T et al (2006) Trafficking of Alzheimer's disease-related membrane proteins and its participation in disease pathogenesis. J Biochem 139:949–955
34. Vetrivel KS, Thinakaran G (2006) Amyloidogenic processing of beta-amyloid precursor protein in intracellular compartments. Neurology 66(2 Suppl 1):S69–S73
35. Almeida CG, Takahashi RH, Gouras GK (2006) Beta-amyloid accumulation impairs multivesicular body sorting by inhibiting the ubiquitin-proteasome system. J Neurosci 26:4277–4288
36. Chen X, Yan SD (2006) Mitochondrial Abeta: a potential cause of metabolic dysfunction in Alzheimer's disease. IUBMB Life 58: 686–694
37. Reddy PH (2006) Amyloid precursor protein-mediated free radicals and oxidative damage: implications for the development and progression of Alzheimer's disease. J Neurochem 96:1–13

38. Keating DJ, Chen C, Pritchard MA (2006) Alzheimer's disease and endocytic dysfunction: clues from the Down syndrome-related proteins, DSCR1 and ITSN1. Ageing Res Rev 5:388–401
39. Scragg JL, Fearon IM, Boyle JP et al (2005) Alzheimer's amyloid peptides mediate hypoxic up-regulation of L-type Ca^{2+} channels. FASEB J 19:150–152
40. Andersen OM, Schmidt V, Spoelgen R et al (2006) Molecular dissection of the interaction between amyloid precursor protein and its neuronal trafficking receptor SorLA/LR11. Biochemistry 45:2618–2628
41. Zhang H, Komano H, Fuller RS et al (1994) Proteolytic processing and secretion of human beta-amyloid precursor protein in yeast. Evidence for a yeast secretase activity. J Biol Chem 269:27799–27802
42. Zhang W, Espinoza D, Hines V et al (1997) Characterization of beta-amyloid peptide precursor processing by the yeast Yap3 and Mkc7 proteases. Biochim Biophys Acta 1359: 110–122
43. Luthi U, Schaerer-Brodbeck C, Tanner S et al (2003) Human beta-secretase activity in yeast detected by a novel cellular growth selection system. Biochim Biophys Acta 1620:167–178
44. Edbauer D, Winkler E, Regula JT et al (2003) Reconstitution of gamma-secretase activity. Nat Cell Biol 5:486–488
45. Sparvero LJ, Patz S, Brodsky JL et al (2007) Proteomic analysis of the amyloid precursor protein fragment C99: expression in yeast. Anal Biochem 370:162–170
46. Caine J, Sankovich S, Antony H et al (2007) Alzheimer's Abeta fused to green fluorescent protein induces growth stress and a heat shock response. FEMS Yeast Res 7:1230–1236
47. Treusch S, Hamamichi S, Goodman JL et al (2011) Functional links between Abeta toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. Science 334:1241–1245
48. Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41: 1088–1093
49. Lambert JC, Heath S, Even G et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094–1099
50. D'Angelo F, Vignaud H, Di Martino J et al (2013) A yeast model for amyloid-beta aggregation exemplifies the role of membrane trafficking and PICALM in cytotoxicity. Dis Model Mech 6:206–216
51. Bagriantsev S, Liebman S (2006) Modulation of Abeta42 low-n oligomerization using a novel yeast reporter system. BMC Biol 4:32
52. Park SK, Pegan SD, Mesecar AD et al (2011) Development and validation of a yeast high-throughput screen for inhibitors of Abeta(4)(2) oligomerization. Dis Model Mech 4: 822–831
53. Buee L, Bussiere T, Buee-Scherrer V et al (2000) Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. Brain Res Brain Res Rev 33:95–130
54. Maas T, Eidenmuller J, Brandt R (2000) Interaction of tau with the neural membrane cortex is regulated by phosphorylation at sites that are modified in paired helical filaments. J Biol Chem 275:15733–15740
55. Brandt R, Leger J, Lee G (1995) Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain. J Cell Biol 131:1327–1340
56. Fulga TA, Elson-Schwab I, Khurana V et al (2007) Abnormal bundling and accumulation of F-actin mediates tau-induced neuronal degeneration in vivo. Nat Cell Biol 9: 139–148
57. Scales TM, Derkinderen P, Leung KY et al (2011) Tyrosine phosphorylation of tau by the SRC family kinases lck and fyn. Mol Neurodegener 6:12
58. Lee G (2005) Tau and src family tyrosine kinases. Biochim Biophys Acta 1739:323–330
59. Ittner LM, Gotz J (2011) Amyloid-beta and tau—a toxic pas de deux in Alzheimer's disease. Nat Rev Neurosci 12:65–72
60. Sergeant N, Bretteville A, Hamdane M et al (2008) Biochemistry of Tau in Alzheimer's disease and related neurological disorders. Expert Rev Proteomics 5:207–224
61. Gendron TF, Petrucelli L (2009) The role of tau in neurodegeneration. Mol Neurodegener 4:13
62. Lee VM, Goedert M, Trojanowski JQ (2001) Neurodegenerative tauopathies. Annu Rev Neurosci 24:1121–1159
63. De Vos A, Anandhakumar J, Van den Brande J et al (2011) Yeast as a model system to study tau biology. Int J Alzheimers Dis 2011:428970
64. Vandebroek T, Vanhelmont T, Terwel D et al (2005) Identification and isolation of a hyperphosphorylated, conformationally changed intermediate of human protein tau expressed in yeast. Biochemistry 44:11466–11475
65. Carmel G, Mager EM, Binder LI et al (1996) The structural basis of monoclonal antibody Alz50's selectivity for Alzheimer's disease pathology. J Biol Chem 271:32789–32795

66. Uboga NV, Price JL (2000) Formation of diffuse and fibrillar tangles in aging and early Alzheimer's disease. Neurobiol Aging 21:1–10
67. Weaver CL, Espinoza M, Kress Y et al (2000) Conformational change as one of the earliest alterations of tau in Alzheimer's disease. Neurobiol Aging 21:719–727
68. Hallows JL, Chen K, DePinho RA et al (2003) Decreased cyclin-dependent kinase 5 (cdk5) activity is accompanied by redistribution of cdk5 and cytoskeletal proteins and increased cytoskeletal protein phosphorylation in p35 null mice. J Neurosci 23:10633–10644
69. Spittaels K, Van den Haute C, Van Dorpe J et al (1999) Prominent axonopathy in the brain and spinal cord of transgenic mice overexpressing four-repeat human tau protein. Am J Pathol 155:2153–2165
70. Jicha GA, Weaver C, Lane E et al (1999) cAMP-dependent protein kinase phosphorylations on tau in Alzheimer's disease. J Neurosci 19:7486–7494
71. Zheng-Fischhofer Q, Biernat J, Mandelkow EM et al (1998) Sequential phosphorylation of Tau by glycogen synthase kinase-3beta and protein kinase A at Thr212 and Ser214 generates the Alzheimer-specific epitope of antibody AT100 and requires a paired-helical-filament-like conformation. Eur J Biochem 252: 542–552
72. Wen Y, Planel E, Herman M et al (2008) Interplay between cyclin-dependent kinase 5 and glycogen synthase kinase 3 beta mediated by neuregulin signaling leads to differential effects on tau phosphorylation and amyloid precursor protein processing. J Neurosci 28:2624–2632
73. Vandebroek T, Terwel D, Vanhelmont T et al (2006) Microtubule binding and clustering of human Tau-4R and Tau-P301L proteins isolated from yeast deficient in orthologues of glycogen synthase kinase-3beta or cdk5. J Biol Chem 281:25388–25397
74. Vanhelmont T, Vandebroek T, De Vos A et al (2010) Serine-409 phosphorylation and oxidative damage define aggregation of human protein tau in yeast. FEMS Yeast Res 10:992–1005
75. Gamblin TC, Berry RW, Binder LI (2003) Tau polymerization: role of the amino terminus. Biochemistry 42:2252–2257
76. Bobba A, Petragallo VA, Marra E et al (2010) Alzheimer's proteins, oxidative stress, and mitochondrial dysfunction interplay in a neuronal model of Alzheimer's disease. Int J Alzheimers Dis 621870, 11 pg
77. Martinez A, Portero-Otin M, Pamplona R et al (2010) Protein targets of oxidative damage in human neurodegenerative diseases with abnormal protein aggregates. Brain Pathol 20:281–297
78. Melov S, Adlard PA, Morten K et al (2007) Mitochondrial oxidative stress causes hyperphosphorylation of tau. PLoS One 2:e536
79. Moreira PI, Santos MS, Oliveira CR et al (2008) Alzheimer disease and the role of free radicals in the pathogenesis of the disease. CNS Neurol Disord Drug Targets 7:3–10
80. Van Loon AP, Pesold-Hurt B, Schatz G (1986) A yeast mutant lacking mitochondrial manganese-superoxide dismutase is hypersensitive to oxygen. Proc Natl Acad Sci U S A 83:3820–3824
81. Van Dyck E, Foury F, Stillman B et al (1992) A single-stranded DNA binding protein required for mitochondrial DNA replication in S. cerevisiae is homologous to E. coli SSB. EMBO J 11:3421–3430

Chapter 12

Yeast as a Model for Studies on Aβ Aggregation Toxicity in Alzheimer's Disease, Autophagic Responses, and Drug Screening

Afsaneh Porzoor and Ian Macreadie

Abstract

The Aβ peptide is widely considered a major cause of Alzheimer's disease since it causes neuronal death in an oligomerisation-dependent manner. In order to identify new inhibitors of Aβ that may be chemo preventative for Alzheimer's disease, a yeast assay that qualitatively determines the amounts and state of the human Aβ42 peptide has been developed. Yeast assays such as this can be applied to studies on aggregation toxicity, autophagic responses and drug screening in Alzheimer's disease.

Key words Alzheimer's disease, Amyloid beta, Autophagy, Bioassays, Flow cytometry, Oligomerisation, Protein aggregation, Protein misfolding, Yeast

1 Introduction

There is current widespread recognition that Alzheimer's disease (AD) is caused by the amyloid beta peptide (also known as Aβ). Aβ generally includes a number of species, usually in the range of 40–43 amino acids in length. However, in this chapter our focus is the 42 amino acid form (Aβ42) whose sequence is shown below:

DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMV GGVVIA.

For convenience, we refer to this form here as 'Aβ' (*see* **Note 1**). There is intense interest in this form because it is associated with the oligomerisation-dependent killing of neural cells in vitro. The Aβ induced killing of neuronal cells is a likely cause of the development of Alzheimer's disease, so there is widespread research to inhibit such activity.

Strategies to reduce Aβ-induced killing include inactivation and/or removal of Aβ through antibodies, and efforts to block the formation of the toxic oligomeric species. Outcomes so far has

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_12,

been very limited, largely due to the limitations of animal models that have deficiencies in their assays, including the assaying of AD chemo preventatives.

Our focus in this report is the use of yeast to screen for compounds targeting Aβ that may have utility in people for the prevention of Alzheimer's disease. Our approaches are "open minded" in that the assays could identify compounds that affect Aβ in a number of ways. We consider that compounds we are screening for include:

- Compounds that bind Aβ directly and block its oligomerisation.
- Compounds that bind Aβ and target it for degradation.
- Compounds that enhance cellular processes leading to Aβ degradation.
- Compounds that enhance cellular processes to keep Aβ in a monomeric form.

The assay that we use employs *Saccharomyces cerevisiae* transformed with a plasmid that enables constitutive expression of a green fluorescent protein (GFP) with Aβ fused to it at either the N- or C-terminus. The construct uses the strong, constitutive glyceraldehyde-3-phosphate dehydrogenase (GPD) or phosphoglycerate kinase (PGK) promoters so all cells produce the fusion protein. When expressing the unfused GFP this can be easily detected since ~80 % of the cells in a growing population have green fluorescence, dispersed throughout the cell (Fig. 1).

In contrast, when expressing Aβ fused to GFP a minority of cells have green fluorescence. The reasons for this are that the majority of cells in a growing population of yeast are young, newly budded, non-fluorescent cells. The most fluorescent cells are older cells, as indicated by Calcofluor staining which stains bud scars (Fig. 1). Interestingly Aβ-GFP can be observed to pass from an older cell to its progeny by microscopy. Also note from Fig. 1 that the Aβ-GFP fluorescence is all localized to punctate patches.

We know that the expression is constitutive, so what is the status of Aβ-GFP in non-fluorescent cells? To answer this question a cell sorter was used to fractionate the population into four subpopulations that ranged from non-fluorescent to highly fluorescent cells. The cells in each of these populations were disrupted with glass beads breakage and the cell lysates were then fractionated by SDS-PAGE for Coomassie staining and Western blotting. The results of the analysis, shown in Fig. 2, indicates that the level of fluorescence is proportional to the amount of Aβ-GFP. All cells contain the plasmid for constitutive expression of Aβ-GFP, and the levels of Aβ-GFP range from not detectable to high levels. This leads us to the conclusion that most cells are non-fluorescent because they readily degrade the Aβ-GFP fusion protein. From the results presented in Fig. 1 and analysis of the populations (Fig. 2) we conclude that young cells, majority in a growing population, efficiently degrade the Aβ-GFP

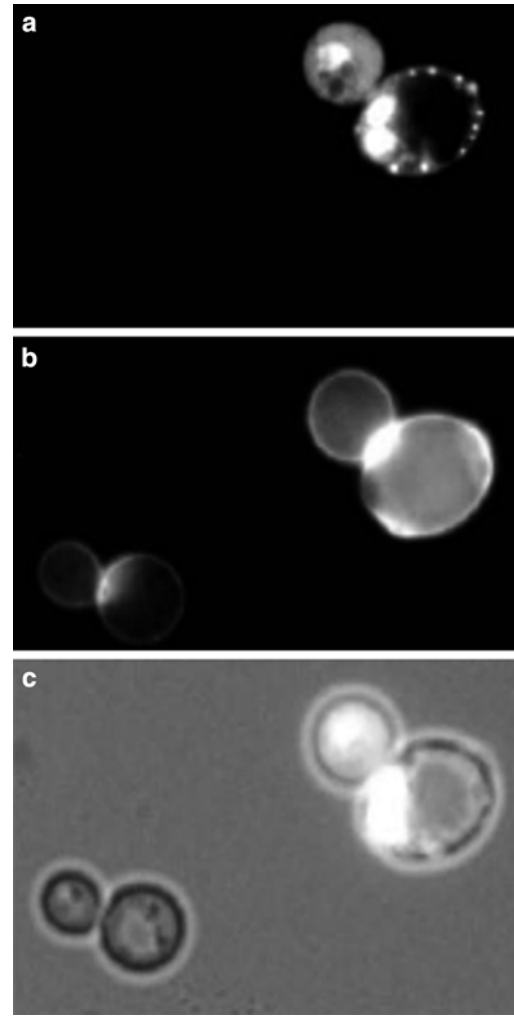


Fig. 1 *S. cerevisiae* yeast cells of different age expressing GFP fused to Aβ. (young on the *left*; old on the *right*, with higher number of bud scars). (a) GFP fluorescence. (b) Calcofluor fluorescence. (c) Light microscopy

fusion protein, and daughters from older mothers are less efficient at degradation of the Aβ-GFP fusion protein.

Escherichia coli has also been engineered by Michael Hecht and colleagues to produce GFP fused to Aβ, similar to the yeast assay described here. However, in *E. coli* the fusion protein forms insoluble aggregates with no green fluorescence [1]. Nevertheless, these *E. coli* transformants offer an alternate means to examine Aβ aggregation mutants [1] and to screen compounds that affect Aβ aggregation [2]. Because the expression levels are much higher in *E. coli*, when the protein is correctly folded the levels of green fluorescence can be observed in a plate reader. Some of the most profound differences between the yeast and *E. coli* assays are outlined in Table 1. Although simpler to use, there are some shortcomings

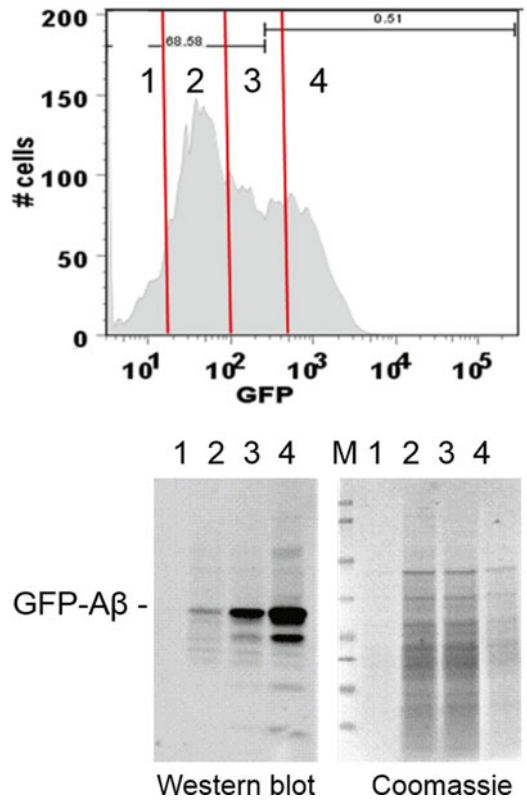


Fig. 2 Analysis of yeast transformed with pAS1N-GFP.Aβ 2 h after inoculation into fresh medium. The population was sorted with a FACS Aria flow cytometer into four populations with levels of fluorescence ranging from lowest (fraction 1) to highest (fraction 4). Cells were disrupted and whole lysates were electrophoresed on SDS-PAGE for staining with Coomassie or for Western blotting with WO2 antibody to Aβ [4]

Table 1
Comparison of GFP-Aβ fusion protein assays in *E. coli* and *S. cerevisiae*

GFP fused to Aβ in *E. coli*	GFP fused to Aβ in *S. cerevisiae*
No fluorescent cells	3–20 % fluorescent cells
All fusion protein is misfolded, aggregated or oligomeric	Only fluorescent cells have fusion protein. Such cells are older
Useful for screening compounds that inhibit misfolding, aggregation or oligomerisation	Useful for screening compounds that inhibit misfolding, aggregation or oligomerisation
Fluorescence assays in microtiter trays	Fluorescence assays requires flow cytometry or microscopy
Doesn't screen for autophagy	Can screen for autophagy

with the *E. coli* assays. In particular, the fact that *E. coli* is not a eukaryote. Yeast on the other hand, as a reference model eukaryote, is expected more likely to deal with Aβ in a closer-to-human proteome networks and organelle environment. In fact, yeast continues to hold a high place as an appropriate reference organism for research on Alzheimer's disease, as discussed in our recent review [3]. There are additional relevant yeast assays to study tau and Aβ effects. Here, we have focused on aggregation assays, autophagy and drug screening relating to the effects of Aβ.

2 Materials

2.1 Yeast Media

1. YEPD medium: 1 % (w/v) yeast extract, 2 % (w/v) Bacto™ peptone, 2 % (w/v) D-glucose.
2. Minimal medium: 0.67 % (w/v) yeast nitrogen base (YNB) without amino acids, 2 % (w/v) glucose, as required 20 mg/L leucine, 20 mg/L tryptophan, 20 mg/L histidine, 20 mg/L uracil, 20 mg/L adenine (*see* **Note 2**).

2.2 Yeast Strains

Freshly grown yeast strains. A variety of *Saccharomyces cerevisiae* strains, wild type or mutant can be used, provided selection for plasmids can be maintained. Strains used in our experiments include the wild-type strains:

- W303-1a (*MATa*, *ura3*-52, *leu2*-3,112, *ade2*-1, *his3*-11, *trp1*-1).
- BY4743 (*MATa*/α, *his3*Δ1/*his3*Δ1, *leu2*Δ0/*leu2*Δ0, *lys2*Δ0/*LYS2*, *MET15*/*met15*Δ0, *ura3*Δ0/*ura3*Δ0) (*see* **Note 3**).

2.3 Transformation of Yeast and Cultivation

1. The yeast plasmids used in these assays include the following: pAS1N.Aβ-GFP and pAS1N.GFP-Aβ [4, 5], as well as p416GPD, p416GPD.GFP and p416GPD.GFP-Aβ [10] (*see* **Note 4**).
2. EZ-Yeast™ transformation Kit (containing transformation solution and carrier DNA) which can be purchased or transformation reagents (*see* **Note 5**) [6].
3. 500 mL flasks for growth of transformants.
4. 24 well tissue culture plates or 15 mL culture tubes.
5. Shaker incubator set at 30 °C with shaking at 200 rpm.
6. Incubator at 30 °C.

2.4 Chemicals to Screen for Compounds Targeting Aβ

1. Chemical compounds can be added to yeast cultures producing human Aβ to monitor their effects, screening for compounds targeting Aβ. First, prepare stocks dissolving chemicals in appropriate solvent (water, ethanol or DMSO) according to manufacturer's instructions.

2. Compounds of known molecular weight can be added to result in 30 and 100 μM final concentration. For mixtures, such as food fractions, a final concentration of 2 mg/mL can be tested (*see* Subheading 3).

2.5 Flow Cytometry

1. Flow cytometry tubes or 96 well microtiter plate.
2. Propidium iodide (PI) dye stock solution made at a concentration of 1 mg/mL in filtered sterile water. Working solutions can be made from this (*see* **Notes 6** and **7**).
3. Disposable 0.2 μm syringe filters and syringe.
4. Flow cytometer.
5. Data analysis software.

3 Methods

The steps involved in the cell culture and analysis by flow cytometry are relatively straightforward and are performed over 3 days after the transformants have been obtained. The first day only requires inoculation of the transformants and incubation. The second day treatments are usually performed in culture tubes, however, cultivation in 96-well flat bottom microtiter trays has also been used. The volume of media can be adjusted according to requirements. All steps can be performed at room temperature unless otherwise stated.

All solvents used in the experiment should be filtered by 0.2 μm pore size disposable filters. Propidium iodide (PI) can be used if the level of cell death by the chemicals needs to be estimated (see below).

3.1 Transformation of Yeast Cells and Cultivation

1. After cultivation at 30 °C [6], transform yeast cells with the appropriate plasmid (*see* **Note 5**) to obtain transformants directing production of GFP and GFP fused to Aβ. Include a transformation with the control vector (e.g. p416GPD; empty vector) that produces no GFP. Transformants usually appear after few days. These are checked and validated by standard molecular biology procedures [6].
2. Prepare three flasks (500 mL volume) containing 100 mL of YNB selective medium and sterilize in autoclave.
3. Inoculate each flask with one of the yeast transformants: (a) expressing GFP alone, (b) GFP fused to Aβ, and (c) empty vector.
4. Incubate these flasks overnight in an incubator at 30 °C with shaking at 200 rpm to allow air transfer into the culture.
5. The next day, take 1 mL of the overnight culture to be inoculated into 100 mL of fresh minimal selective medium. Incubate at 30 °C, 200 rpm for 2–4 h to get cells in exponential phase (*see* **Note 8**).

3.2 Addition of Chemicals to Screen for Compounds Targeting Aβ

1. Add compounds to the cultures at a final concentration of 30 and 100 μM (for compounds of known molecular weight). For mixtures such as food fractions, a concentration of 2 mg/mL can be tested. Add each compound to both cultures, growing GFP and GFP-Aβ transformants. Continue incubation for at least 4 h at 30 °C and 200 rpm.
2. Control samples for GFP, GFP-Aβ and those with empty vector should be included in triplicate to aid the set up of the flow cytometer and the analysis of the results.
3. Control samples for any solvent that has been used to dissolve the chemicals (i.e. adding only the solvent) should also be included.

3.3 Flow Cytometry and Data Analysis

1. We use a BD Fluorescence Activated Cell Sorting (FACS) Canto™ II flow cytometer with the following settings: FITC 530/30 filter with excitation at 494 nm and emission of 519 nm to measure GFP levels. Red fluorescence due to propidium iodide (PI) staining is also measured with PerCP 670LP filter with excitation at 488 nm and emission at 617 nm.
2. 300 μL of the suspension are transferred into a flow cytometry tube containing 5 μL of propidium iodide solution (*see* **Note 9**).
3. GFP fluorescence and percentage of cell death are estimated with 20,000 cells counted in each sample and percentage of cells exhibiting red and green fluorescence recorded.
4. The control data from samples without any fluorophore is used to set up the gating and minimise any background autofluorescence. Perform flow cytometry to measure the fluorescent population.
5. Recorded data are saved as FCS3 files for further analysis using appropriate data analysis software such as WEASEL, Flow Jo or equivalent.
6. Analyse data using an appropriate statistical analysis package. Student's *t*-tests can be performed to ascertain which compounds cause significant alterations to the number of green fluorescent cells in the population, compared to controls.

3.4 Screening for Compounds Targeting Aβ: Interpretation of Results

Compounds causing changes to the population of yeast fluorescent cells (i.e. expressing the Aβ-GFP fusion protein) are of interest because they may have an effect on Aβ. Considerable work may be required to decipher mechanisms underlying their effects though. For novel compounds with no history of human use, this means there will be a long time before they have the opportunity to be tested and approved for clinical use. Thus, instead of novel compounds, it is worth discussing some of the outcomes obtained with compounds known to be relevant in Alzheimer's disease.

1. Folic acid

   Folic acid (vitamin B9) is a compound that has long been associated with Alzheimer's disease. Many studies point to AD people having low levels of folate and a recent study shows beneficial reduction in grey matter atrophy in AD via supplementation with vitamins B6, B9 and B12 [7]. The study of effect of folate on Aβ in yeast involved the use of genetically-engineered strains lacking folate synthesis. In such yeast strains, the addition of folate (in the form of folinic acid) caused a significant increase in the fluorescent population. Folinic acid caused the population to increase from 23 to 28 % [5]. The mechanism by which folate has this effect on yeast may relate to effects on Aβ aggregation, on the cell cycle or possibly through an antioxidant mechanism.

2. Latrepirdine (Dimebon™),

   An experimental drug in AD therapy is latrepirdine (Dimebon™), a compound, like rapamycin, that appears to promote autophagic clearance of protein aggregates. The addition of latrepirdine reduced the punctate green fluorescent aggregates and led to more cells with diffuse green fluorescence [10]. Latrepirdine stimulated autophagy, thereby reducing Aβ levels [10].

3. Problematic compounds. Fluorescent and toxic compounds

   Compounds that are fluorescent themselves as well as compounds that give cells green fluorescence are a major problem. These are often found and may be readily identified because they can stain the entire cell population, unlike real leads that affect a sub-population. Such compounds are entirely unsuitable in this assay.

Together with these, some compounds may have an adverse effect on yeast cell growth leading to significantly reduced fluorescence. However, it is worth persevering with these compounds, perhaps at lower concentrations. As an example, it is worth noting that one of the best current AD chemo preventatives is simvastatin [8], which has a high inhibitory effect on yeast growth [9]. Such compounds may cause considerable PI staining. We have observed this with curcumin, a food additive associated with AD chemo prevention and reducing Aβ oligomerisation. All these aspects should be carefully considered when analysing and interpreting results screening different compounds, together with the latest reported information on their direct effects on e.g. aggregation toxicity and autophagic responses, among others.

4 Notes

1. Nomenclature in this chapter:
   (Aβ42), DAEFRHDSGYEVHHQKLVFFAEDVGSNKGA IIGLMVGGVVIA is 'Aβ'.
   (Aβ40), DAEFRHDSGYEVHHQKLVFFAEDVGSNKGA IIGLMVGGVV is 'Aβ40'
2. Amino acid and nucleobase requirements for auxotrophic strains. Stocks are conveniently prepared as 1 % (w/v) stock solutions. Note that adenine readily precipitates out of solution. To overcome the difficulty one can make a more diluted solution or heat it before use. Store solutions in refrigerator except for adenine and uracil which should be stored at room temperature. Tryptophan solutions should be kept in dark; use aluminium foil to wrap the bottle. All solutions can be sterilized by autoclaving [6].
3. It may be useful to use mutant strains for certain studies. For example, our studies have used strains EHY1 (*MATa ura3*-52 *leu2*-3,112 *trp1 tup1 DHPS::LEU2*) and LCY1 (*MATa ura3-52 leu2*-3,112 *trp1 tup1 FOL3::URA3*) [11] that do not produce folate, to study the effect of folate levels on Aβ.
4. The plasmids are DNA shuttle vectors that can be amplified in an *E. coli* strain through selection on media containing ampicillin. The pAS1N plasmids are designed to exist in yeast in multiple copies, due the presence of a yeast 2 μ origin of replication. The selectable marker in yeast is the *LEU2* gene and expression of the GFP fusion protein is under control of the strong constitutive *PGK* promoter. The p416.GPD vector has a yeast centromere and a *URA3* gene for low copy maintenance in yeast. Expression of the GFP fusion protein utilises the strong constitutive *GPD* promoter.
   Sequences encoding GFP fused to Aβ were placed in the multiple cloning site. We have an Aβ-GFP linker in our sequence: DLNMSRAQASNSAVDGTAGPGSIATM [4].
5. Transformation can be done using electroporation, commercial kits or refer to "Protocols and techniques #1", page 109 in [6].
6. Working stock solutions can be made at 100 μg/mL using phosphate buffered saline (PBS) and filtered to remove any particles that might interfere with the flow cytometry. The final concentration of the PI in samples should be 1 μg/mL. After addition, incubate samples in dark for 30 min at room temperature, or for 20 min at 37 °C.
7. If using PI, a positive control sample for this dye should be prepared by incubating 1 mL of sample containing yeast transformed with p416GPD (plasmid without insert) for 15 min at 70 °C to kill the cells.

8. After overnight cultivation many cells enter stationary phase and fluorescence levels are at a minimum. After return to exponential phase up to 20 % of the population are fluorescent cells.
9. To screen for effect of compounds on green fluorescence, it is most desirable to achieve effects with the smallest amount of compound. It is not difficult to identify compounds causing altered green fluorescence with hit rates in some screens around 5 %.

Acknowledgements

This work is supported in part by a grant from the Medical Advances Without Animals Trust (MAWA). We thank Deborah Shapira (CSIRO) for cell sorting and Andrew Tsatsanis (Mental Health Research Institute) for developing the Western blot.

References

1. Wurth C, Guimard NK, Hecht MH (2002) Mutations that reduce aggregation of the Alzheimer's Aβ42 peptide: an unbiased search for the sequence determinants of Aβ amyloidogenesis. J Mol Biol 319:1279–1290
2. Kim W, Kim Y, Min J et al (2006) A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. ACS Chem Biol 1:461–469
3. Porzoor A, Macreadie I (2013) Application of yeast to study the tau and amyloid-beta abnormalities of Alzheimer's disease. J Alzheimers Dis 35:217–235
4. Caine J, Sankovich S, Antony H et al (2007) Alzheimer's Aβ fused to green fluorescent protein induces growth stress and a heat shock response. FEMS Yeast Res 7:1230–1236
5. Macreadie I, Lotfi-Miri M, Mohotti S et al (2008) Validation of folate in a convenient yeast assay suited for identification of inhibitors of Alzheimer's Aβ aggregation. J Alzheimers Dis 15:391–396
6. Amberg DC, Burke D, Strathern JN (2005) Laboratory CSH. Methods in yeast genetics: a Cold Spring Harbor laboratory course manual. CSHL Press, New York
7. Douaud G, Refsum H, de Jager CA et al (2013) Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. Proc Natl Acad Sci U S A 110: 9523–9528
8. Wolozin B, Wang SW, Li NC et al (2007) Simvastatin is associated with a reduced incidence of dementia and Parkinson's disease. BMC Med 5:20
9. Macreadie IG, Johnson G, Schlosser T et al (2006) Growth inhibition of *Candida* species and *Aspergillus fumigatus* by statins. FEMS Microbiol Lett 262:9–13
10. Mumberg D, Mailer R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156:119–122
11. Bharadwaj PR, Verdile G, Barr RK et al (2012) Latrepirdine (Dimebon™) enhances autophagy and reduces intracellular GFP-Aβ42 levels in yeast. J Alzheimers Dis 32: 949–967

Chapter 13

Drosophila melanogaster as a Model for Studies on the Early Stages of Alzheimer's Disease

Jung Yeon Lim, Stanislav Ott, and Damian C. Crowther

Abstract

Fruit flies (*Drosophila melanogaster*) have been widely used to study the cellular and molecular basis of human neurodegenerative disease. The biological similarities between the human and the fly have been explored successfully to further investigate the pathological basis of Alzheimer's disease (AD). Here, we discuss transgenic *Drosophila* models systems and the methodologies that have been employed in the study of AD.

Key words Alzheimer's disease, *Drosophila melanogaster*, Amyloid β peptide, Protein aggregation, Invertebrate animal model

1 Introduction

Alzheimer disease is one of the most prevalent causes of dementia in the elderly, with an estimated 35.6 million people affected worldwide in 2009. The AD population is expected to rise to 66 million by 2030 [1, 2]. Pathologically, AD is characterized by the accumulation of two very different proteins, each with a distinct distribution. Amyloid β (Aβ) peptide accumulates extracellularly in amyloid plaques while hyperphosphorylated tau, a microtubule binding protein, accumulates intracellularly in neurofibrillary tangles [3–5]. In particular, the Aβ peptide is formed by the sequential cleavage of the amyloid precursor protein (APP) when β-secretase cuts extracellularly followed by intramembraneous cleavage by γ-secretase [6]. The precise location of γ-secretase cutting determines the nature of the Aβ generated; predominant production of the shorter, less aggregation prone isoform, $A\beta_{1\text{-}40}$, has been linked to health, while the generation of the longer peptide, $A\beta_{1\text{-}42}$, is a risk for AD [7, 8]. Before Aβ peptides form plaques they appear as oligomeric aggregates and it is these forms that are thought to be the neurotoxic intermediate products that cause neuronal dysfunction and death. The biological similarities between human and *Drosophila* have been exploited with great success in the modeling of Aβ

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_13,

toxicity [9]. The fly has a brain, containing approximately 200,000 neurons, and like the vertebrate central nervous system, it is composed of a series of functionally specialized substructures [10–12]. Efforts at modeling AD in *Drosophila* have been predicated in large part on the amyloid cascade hypothesis which states that Aβ aggregation is the first step in a chain of pathological events [13, 14]. Human Aβ peptides can be expressed in the fly in two main ways. Firstly a partially humanized system has been created by Greeve and colleagues by engineering triple transgenic flies that express human wild type APP, human β-secretase (BACE) and the catalytically active subunit of fly γ-secretase (dPsn) [15]. These lines exhibit modest elevations in Aβ40 and Aβ42 peptides, thioflavin S positive amyloid plaques in the retina and an age-dependent degeneration of the photoreceptors. Retinal pathology is particularly marked in flies expressing the FAD related dPsn variant. A second model consists of flies expressing a truncated APP coupled to a secretion signal peptide but still requiring endogenous fly γ-secretase to generate Aβ [16]. Both these approaches result in the secretion of Aβ peptides from neurons and generate similar phenotypes. In this review we will focus on the simpler approach of expressing only Aβ peptides, coupled to a secretion signal peptide. In these fly models of Aβ toxicity, we have taken advantage of the widely-used gal4-UAS system that allows one line of flies that carries a gal4-responsive transgene, to be crossed to any number of tissue-specific gal4-expressing driver lines (*see* Fig. 1 and Table 1) [17].

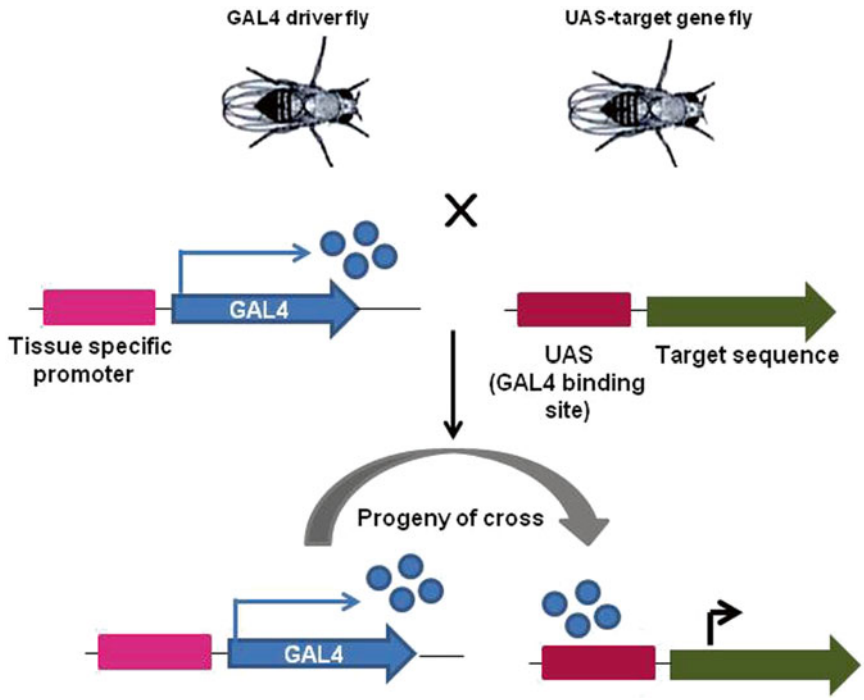


Fig. 1 Schematic representation of the GAL4-based systems for transgene expression

Table 1
Gal4 drivers for tissue specific expression in *Drosophila*

Tissue specificity	Driver line
Brain	elav-gal4
Retina	GMR-gal4
Ubiquitous	act5C-gal4

Once flies expressing Aβ peptides have been generated we then undertake molecular and biochemical analyses, along with phenotype assessment at both the micro- and macroscopic levels.

2 Materials

2.1 *Drosophila* Husbandry

1. Fly stocks: the transgenic lines carrying the Aβ transgene have been generated by a number of groups and are described elsewhere (Crowther et al. ref. 18). Other fly stocks (e.g. driver lines such as *elav^{C155}-GAL4*) are obtained from the Bloomington *Drosophila* Stock Center (Indiana University; http://flystocks.bio.indiana.edu/).
2. *Drosophila* culture medium: a mixture of 1.25 % (w/v) agar, 10.5 % (w/v) dextrose, 10.5 % (w/v) maize, 2.1 % (w/v) yeast to 80 °C is heated and then 5 mL are dispensed into tubes (12 cm × 2 cm diameter) or 30 mL into bottles (10 cm diameter), allowing the mix to cool and solidify.
3. The flies are then incubated at an appropriate temperature (range 18–29 °C) with a 12:12 h light:dark cycle at constant humidity (60 %) (*see* **Note 1**).

2.2 Protein Extraction

1. Control *w^{1118}* flies or flies expressing the Aβ transgene under the control of neuron specific driver *elav-GAL4* (*elav-GAL4*; *w^{1118}* vs *elav-GAL4*; *UAS-Aβ$_{1-42}$*).
2. Disposable pestle.
3. Electric homogenizer.
4. Bath sonicator.
5. Soluble fraction buffer: 2 % (w/v) sodium dodecyl sulfate (SDS) solution in distilled water (dH_2O) with protease inhibitors (Complete, Roche).
6. Insoluble fraction buffer: 80 % (w/v) dimethyl sulfoxide (DMSO), 50 mM Tris–HCl, pH 8.8.
7. Bench top centrifuge.

2.3 Immunoblotting

1. SDS-PAGE gel: Bis-Tris 4–12 % (w/v) gels (Invitrogen).
2. 1 × LDS sample buffer: glycerol, lithium dodecyl sulfate (LDS) sample buffer.
3. Heat block at 70 °C.
4. 2-[morpholino]ethanesulfonic acid (MES) SDS running buffer (Invitrogen).
5. Transfer membrane: Nitrocellulose membrane 0.11 μm pore (Whatman, GE Healthcare).
6. Semi-dry transfer kit (Bio-Rad).
7. Transfer buffer (Tris-glycine buffer): 0.025 M Tris, 0.192 M glycine, 20 % methanol.
8. Washing buffer: 0.05 % (v/v) Triton X-100 solution in phosphate buffer saline (PBS).
9. Blocking buffer: 5 % (w/v) non-fat milk in 0.05 % (v/v) Triton-PBS.
10. Shaker.
11. Primary antibody: anti-$A\beta_{1\text{-}16}$ (6E10) monoclonal antibody (Covance) in blocking buffer.
12. Secondary antibody: HRP-conjugated goat anti-mouse antibody (DAKO) in blocking buffer.
13. Detection: SuperSignal Chemiluminescent Pico and Femto substrates (Thermo Scientific).
14. Kodak X-Omat LS film.
15. Stripping buffer: 10 % (w/v) SDS, 0.5 M Tris–HCl, pH 6.8, 0.8 % ß-mercaptoethanol.
16. Fume hood.

2.4 Brain Dissection

1. Control w^{1118} flies or flies expressing the Aβ transgene under the control of neuron specific driver *elav-GAL4* (*elav-GAL4*; w^{1118} vs *elav-GAL4*; *UAS-*$A\beta_{1\text{-}42}$).
2. Petri dish.
3. 0.5 mL microcentrifuge tubes.
4. Fixation solution: 0.1 % (v/v) Triton X-100 in phosphate buffer (PB), pH 7.4. Prepare 4 % (w/v) paraformaldehyde (PFA) in Triton-PB.
5. Washing buffer: 0.1 % (v/v) Triton-PB.
6. Dissecting microscope.
7. Two pairs of sharp forceps.

2.5 Immunohistochemistry

1. Washing buffer: 0.1 % (v/v) Triton-PB.
2. Blocking buffer: 5 % (w/v) normal goat serum in 0.5 % (v/v) Triton-PB.

3. Primary antibody: anti-$A\beta_{1\text{-}16}$ (6E10) monoclonal antibody (Signet) in blocking buffer.
4. Secondary antibody: Alexa Fluor 488-conjugated goat anti-mouse (Invitrogen) in blocking buffer.
5. TOTO-3 iodide stain (Invitrogen) diluted in 0.5 % (v/v) Triton-PB.
6. Glass slide.
7. Cover slip.
8. Mounting solution Vectashield (Vectors Lab).
9. Nikon Eclipse C1si confocal microscope on Nikon E90i upright stand and imaging software.

2.6 Longevity Assay

1. Control w^{1118} flies or flies expressing the Aβ transgene under the control of neuron specific driver *elav-GAL4* (*elav-GAL4*; w^{1118} vs *elav-GAL4*; *UAS-*$A\beta_{1\text{-}42}$) (*see* **Notes 2** and **3**).
2. 12 cm × 2 cm diameter glass vials containing standard fly food [18].
3. Barcode that can be transferred from tube to tube (*see* **Note 4**).
4. Barcode.
5. Scanner.
6. Database software such as Flytracker2 software designed and programmed by Damian C. Crowther (www.flytracker.gen.cam.ac.uk) (*see* ref. 18).

2.7 Climbing Assay

1. Control w^{1118} flies or flies expressing the Aβ transgene under the control of neuron specific driver *elav-GAL4* (*elav-GAL4*; w^{1118} vs *elav-GAL4*; *UAS-*$A\beta_{1\text{-}42}$).
2. 25 cm × 1.5 cm diameter sterile plastic column.
3. Timer (*see* ref. 19).

2.8 Locomotor Assay

1. Control w^{1118} flies or flies expressing the Aβ transgene under the control of neuron specific driver *elav-GAL4* (*elav-GAL4*; w^{1118} vs *elav-GAL4*; *UAS-*$A\beta_{1\text{-}42}$).
2. 10 cm × 2 cm diameter glass tube.
3. iFly apparatus (iFly chamber, camera, mirror) (*see* ref. 20).

3 Methods

3.1 *Drosophila* Crossing

Figure 2 shows the crossings set up to obtain flies expressing an Aβ transgene under the control of the neuron specific driver, *elav-Gal4*, in the X chromosome. In this instance male flies, homozygous for UAS-Aβ transgenes, are crossed with *elav-Gal4* virgin females (*see* **Note 5**). Flies are then reared at an appropriate temperature (range 18–29 °C) on 12:12 h light:dark cycle at constant humidity (60 %) until they eclose (Fig. 2).

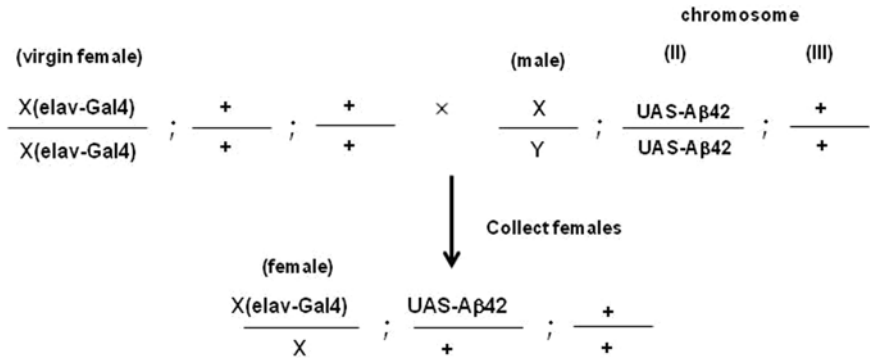


Fig. 2 Crossing scheme for generation of transgene flies. Crossing for generating flies expressing Aβ protein under the control of neuron specific driver *elav-GAL4* (*see* **Note 6**)

3.2 Soluble and Insoluble Fraction of Total Aβ Peptide in the Fly Brain

1. Male flies carrying *UAS-Aβ* transgenes are crossed with *elav-Gal4* virgin females. Developing flies are then reared at 18–29 °C. Progeny are collected 24 h after eclosion and aged for up to 30 days before being collected for the assay of soluble vs insoluble Aβ.
2. At least 30–50 flies are decapitated and the heads homogenised in 50 μL 2 % (w/v) SDS/water solution supplemented with protease inhibitors.
3. The samples are then sonicated in an ice water bath for 8 min, followed by centrifugation at 4 °C for 20 min, at 18,000×*g*.
4. After the centrifugation step, the supernatant is collected and labelled as the 'soluble fraction'. The remaining pellet is washed in PBS which is then removed after a further round of centrifugation for 20 min at 18,000×*g*. The pellet is then resuspended in 5 μL of a solution containing 80 % (w/v) DMSO and left at 55 °C in a sealed tube for 1 h.
5. 15 μL of 50 mM Tris–HCl are added, and a brief centrifugation step (18,000×*g*) is performed to eliminate debris. The supernatant is retained and labelled as the 'insoluble fraction'.
6. Soluble and insoluble samples are denatured in lithium dodecyl sulfate (LDS) sample buffer for 10 min at 70 °C. Proteins are then separated by polyacrylamide gel electrophoresis on 4–12 % (w/v) Bis-Tris gels and transferred onto a 0.1 μm pore nitrocellulose membrane using semi-dry transfer, at 15 V for 35 min.

3.3 Immunoblotting

1. The membranes are boiled for 5 min in PBS and then blocked for 1 h in 5 % (w/v) dried milk in PBS (*see* **Note 7**). The primary antibody, 6E10 is diluted 1:2,500 in 5 % (w/v) dried milk in 0.05 % (v/v) Triton-PBS and incubated with the blot overnight at 4 °C, followed by washing five times with PBS, 5 min each time.

2. The secondary antibody, HRP-labelled anti-mouse IgG, diluted 1:2,500 in 5 % (w/v) dried milk in 0.05 % (v/v) Triton-PBS is incubated with the blot for 2 h at room temperature (RT).
3. The membrane is then washed five times with PBS, 5 min each time, before developing the blot using Super Signal West Pico (for insoluble fraction) or Femto (for soluble fraction).
4. After developing, the membrane is washed with PBS and then incubated with the stripping solution at 50 °C for up to 45 min with some agitation (*see* **Note 8**).
5. Dispose of the solution and the membrane is washed five times with PBS, 5 min each time, and then blocked for 1 h in 5 % (w/v) dried milk in 0.05 % (v/v) Triton-PBS.
6. Repeat **steps 2** and **3** with new antibody. This step often includes anti-β-actin to provide a loading control for the western blot (Fig. 3).

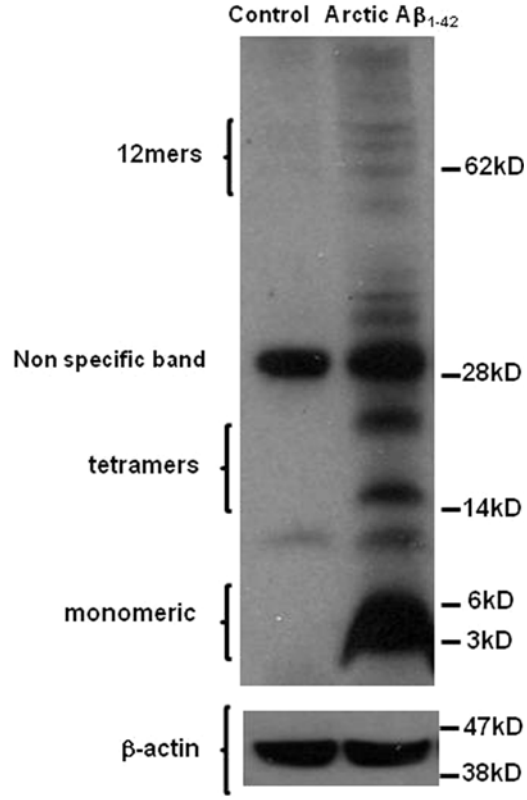


Fig. 3 Western blotting analysis of flies expressing Aβ transgenes. The western blot is performed with head extracts of *Drosophila*. The SDS soluble fraction is probed for Aβ using the monoclonal 6E10 antibody. The E22G variant of $A\beta_{1\text{-}42}$ (*Arctic* $A\beta_{1\text{-}42}$) exhibits a variety of higher molecular weight (25–65 kDa) immunoreactive species that are not present in the negative control (*Drosophila* brain extracts without Aβ expression). Loading of equal amounts total protein are confirmed using anti-β-actin antibody

3.4 Brain Dissection

1. Anaesthetize flies with ice and place them in a petri dish.
2. Prepare a 500 µL tube containing 500 µL of fresh 4 % (w/v) paraformaldehyde and place the flies into the tube and fix them for 2.5 h at room temperature (RT).
3. The samples are washed three times (10 min) with 0.1 % (v/v) Triton-PB at RT.
4. Place the samples into the petri dish filled with PB buffer and put it under a dissecting microscope.
5. The brain is removed from the head cuticle using gentle manipulations with dissection forceps. Firstly the fly should be held with forceps in the petri dish filled with PB buffer, belly up. With the second pair of forceps, gently insert one side into the cavity just below the eye to obtain a grip on the eye. Be careful to avoid internal head structures such as the brain. Gently pull the head off of the fly and discard the body. With the free forceps, obtain a grip on the other eye from the underside. Gently pull the two pairs of forceps away from each other to open the head cuticle.
6. The tissue is now ready for immunohistochemistry.

3.5 Immunohistochemistry

1. Brains are washed in 0.1 % (v/v) Triton-PB and blocked (to prevent non-specific staining), in 5 % (w/v) normal goat serum in 0.5 % (v/v) Triton-PB for 2 h at room temperature (RT). Brains are stained for Aβ using a dilution 1:1,000 of the anti Aβ monoclonal antibody 6E10 in blocking buffer for 48 h at 4 °C.
2. Brains are then washed three times for 10 min with 0.1 % (v/v) Triton-PB and left in a solution containing the secondary antibody, goat anti-mouse Alexa Fluor 488 diluted 1:1,000 in blocking buffer at 4 °C overnight.
3. After removal of the secondary antibody, brains are washed in 0.1 % (v/v) Triton-PB and stained for DNA using TOTO-3 iodide diluted in 0.1 % (v/v) Triton-PB 1:5,000 for 10 min.
4. After an additional wash in 0.1 % (v/v) Triton-PB, brains are orientated on a glass slide under the light microscope and covered with a drop of the Vectashield mounting solution.
5. Confocal laser scanning images are collected at intervals of 5 µm using a Nikon Eclipse C1si on Nikon E90i upright stand, 20× objective. Laser intensities are set at the beginning of each image acquisition session and kept constant to allow comparison of the fluorescence intensity between different samples. Images are processed using ImageJ software (Fig. 4).

3.6 Longevity Assays

1. Male flies carrying *UAS-Aβ* transgene are crossed with *elav-Gal4* virgin females. Flies are then reared at an appropriate temperature (range 18–29 °C) for survival analysis.

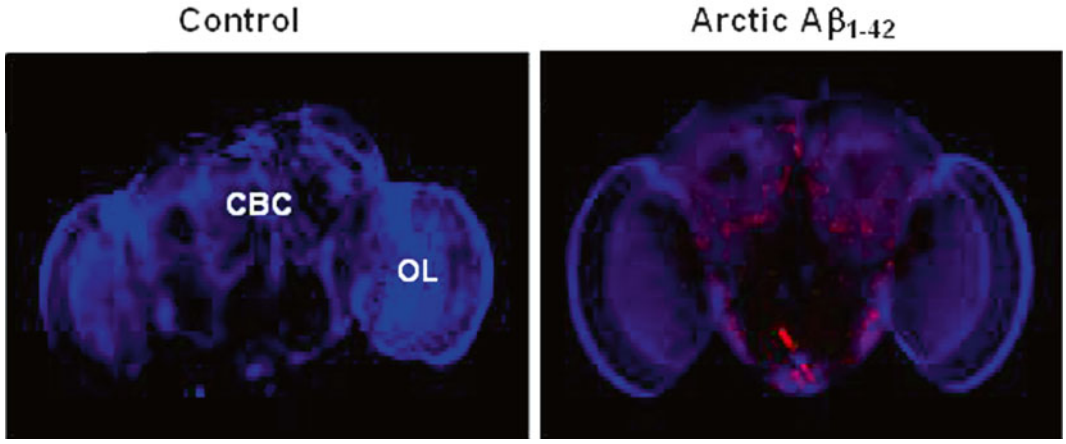


Fig. 4 Confocal micrographs of control or flies expressing Aβ transgenes under the control of the *elav-Gal4* driver. Fly brains are counter-stained with TOTO-3 (DNA staining, *blue*) and probed for amyloid with the 6E10 antibody (*red*). The *left panel* illustrates a typical *Drosophila* brain without beta-amyloid expression. Roughly, the brain is divided into the central brain complex (CBC) and the optic lobes (OL) area to each side. Flies that overexpress the E22G variant of $A\beta_{1\text{-}42}$ (*Arctic* $A\beta_{1\text{-}42}$) display more brain plaques (*right panel*). High resolution color images appear in the online version of this work

2. From the progeny, mated females flies are collected 24 h post-eclosion and divided between ten tubes, each containing ten flies. The flies are then incubated on standard fly food and yeast at an appropriate temperature [18].
3. The number of flies surviving is documented with a frequency appropriate to the lifespan of the particular flies: For short-lived flies, daily observations are made; however, for most strains, the flies are counted on days 1, 3, and 5 of a 7 days cycle. At each time point, the number of flies that are observed to die and the number of flies that are lost to follow up (for example, flies that escape or are accidentally killed) are noted. A computer database may be required to maintain large amounts of data.
4. When all the test and control flies are either dead or lost, the data can be visualized using Kaplan–Meier survival plots and statistical comparisons are performed using the log rank test (GraphPad Prism), assuming that the total population of 100 flies is homogenous. A more conservative approach is to calculate the median survival for each of the ten tubes and use a non-parametric test to assess differences in these survival estimates as described by Crowther et al. [18]. This analysis provides median and mean survival times for a population and determines the significance of any difference in survival times (Fig. 5). Typically, reliable data are derived from the assessment of at least 50 flies from three or more independent crosses.

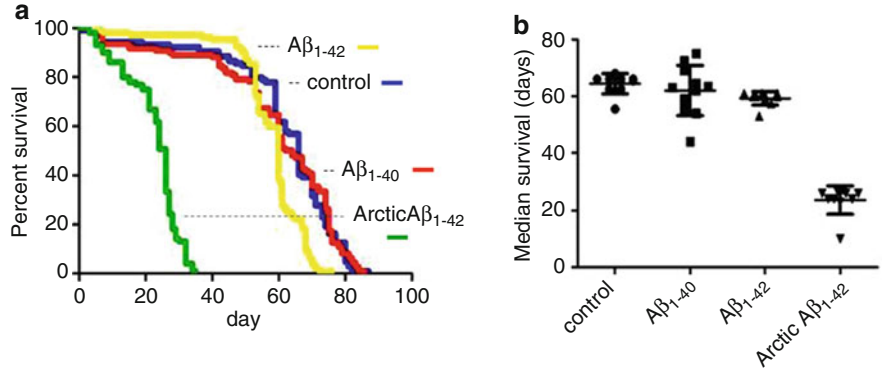


Fig. 5 Longevity assay of flies expressing Aβ transgenes. The longevity of control w^{1118} flies is compared with flies expressing *Arctic* $A\beta_{1\text{-}42}$, $A\beta_{1\text{-}42}$ or $A\beta_{1\text{-}40}$ peptide under the control of *elav-GAL4* driver. The survival curves are drawn using GraphPad Prism (**a**). Expression of the $A\beta_{1\text{-}42}$ peptide in the fly's nervous tissue reduces the longevity of the flies and this is further accelerated by expression of the transgene containing the E22G variant of $A\beta_{1\text{-}42}$ (*Arctic* $A\beta_{1\text{-}42}$). The analysis of the survival data shows that expression of the $A\beta_{1\text{-}42}$ peptide in the fly's nervous tissue results in reduction in the median survival (**b**). In this example there are no remarkable differences in the survival profiles for flies expressing $A\beta_{1\text{-}40}$ compared to the control flies indicating no toxic effect on adult neurons; however $A\beta_{1\text{-}42}$, and more potently *Arctic* $A\beta_{1\text{-}42}$, reduce fly longevity. Color figure appears in the online version of this work

3.7 Climbing Assays

1. Male flies carrying *UAS-Aβ* transgenes are crossed with *elav-Gal4* virgin females. Flies are then reared at an appropriate temperature (range 18–29 °C).
2. From the progeny, mated females flies are collected 24 h post-eclosion, and sets of 15 flies are placed at the bottom of a clean plastic column of 1.5 cm diameter and 25 cm height.
3. The flies are bought down to the bottom of the column by firmly tapping the tube on the bench and start the timer.
4. After 45 s the flies at the top of the column (N_{top}) and the flies remaining at the bottom (N_{bot}) are counted.
5. The flies are bought down again by firmly tapping the tube on the bench and repeat this for a total of three climbing opportunities at 1 min intervals.
6. Another test run is performed with the same parameters.
7. A performance index is calculated for each group of flies and repeated tests are performed to allow statistical comparison of different fly populations. The performance index (PI) is calculated as $PI = (15 + N_{top} - N_{bot})/30$. Statistical analysis are performed using the two-tailed Student's *t*-test (*see* ref. 19).

3.8 Locomotor Assay

1. Male flies carrying *UAS-Aβ* transgenes are crossed with *elav-Gal4* virgin females. Flies are then reared at an appropriate temperature (range 18–29 °C).
2. From the progeny, mated females flies are collected 24 h post-eclosion and, for each condition, five flies are placed in each of four different tubes of 2 cm diameter and 10 cm height. Videos are taken on days 1, 3, 5, 8, 10, 12, 15 and 18.
3. Test tubes are tapped down into the iFly apparatus at time zero and videos recorded for 90 s, with the tube tapped down again into the apparatus after 30 and 60 s to give a total of three climbing opportunities. All videos are processed with the iFly software (*see* ref. 20). Tapping of the test tube results in flies dropping to the bottom of the tube on impact and immediate initiation of upwards movement is triggered by innate negative geotaxis reflexes in the flies. The trajectories of all flies are recorded by the computer for each of the three 30 s movie chips, each in form of time-stamped Cartesian coordinates of the derived fly positions in 3D space.
4. The Cartesian coordinates for the locomotion behavior are analysed to extract statistical descriptors of the fly populations. The statistical properties of these parameters are useful to discriminate flies at various stages of their lives with a high level of confidence (Fig. 6).

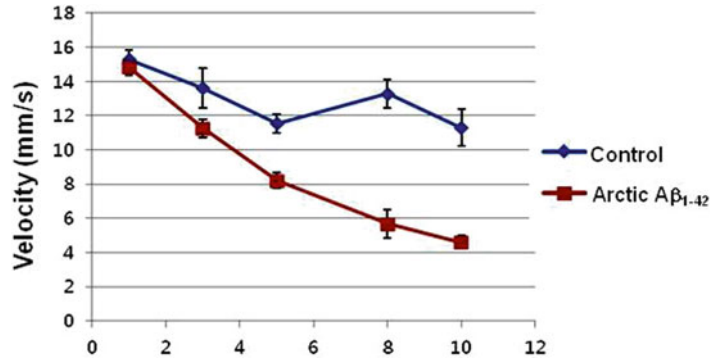


Fig. 6 Locomotor assay of flies expressing Aβ transgenes. The locomotor result for flies of control w^{1118} is compared with flies expressing *Arctic* $A\beta_{1\text{-}42}$, peptide under the control of the *elav-Gal4*. Velocities are assigned to bins of width 3 mm per second, with the first bin starting at zero millimetres per second. Error bars indicate the bin-wise standard deviation over three replicates of 20 flies each. The analysis of the locomotor data shows that expression of the $A\beta_{1\text{-}42}$ peptide in the fly's nervous tissue resulted in reduction in the locomotor activity

4 Notes

1. Higher temperature induces stronger Aβ expression, which reduces the life span of flies expressing Aβ transgenes under the control of neuron-specific driver *elav-GAL4*.
2. Ensure that flies are entered into longevity assays within a few days to avoid differences in food and environmental conditions affecting the outcome.
3. Once virgin females have been collected, add the males. Females will then begin laying fertile eggs soon after. Check the flies in 2–3 days to see if larvae are present and then remove the parent flies.
4. Anonymize tubes by using barcode labels and also randomize tubes within trays so that the operator does not know the identity of the flies being followed.
5. Virgin females are required for each crossing. In a mixed culture females remain virgin for only 8 h at 25 °C after eclosion and must be collected within this time frame. To confirm that females are virgin they can be cultured on standard food for 2–3 days at 25 °C. The presence of larvae on the food indicates that at least one female is non-virgin. In this case all the females in the particular tube will be discarded. If only eggs are present, the females are all likely to be virgin.
6. There is a choice of collecting females or virgin females. We choose to allow 24 h mating so they are all females.
7. Boiling of the nitrocellulose is required for western blot detection of Aβ from fly brains. Bring the blot to boil in the microwave and allow it to stand for 5 min before placing it in the blocking buffer.
8. Put the membrane with the buffer into a small plastic box with a tight lid. Use a volume of buffer that covers the whole membrane.

Acknowledgement

This work was supported by the Korea-UK Alzheimer's disease research consortium program from the Ministry of Health and Welfare, Republic of Korea (J.Y.L.) and the Medical Research Council (UK) (G0700990; D.C.C.) and the Wellcome Trust (082604/2/07/Z; D.C.C.). D.C.C. is an Alzheimer's Research UK Senior Research Fellow (ART-SRF2010-2).

References

1. Citron M (2004) Strategies for disease modification in Alzheimer's disease. Nat Rev Neurosci 5:677–685
2. Brunden KR, Trojanowski JQ, Lee VM (2009) Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies. Nat Rev Drug Discov 8:783–793
3. Mandelkow E, von Bergen M, Biernat J, Mandelkow EM (2007) Structural principles of tau and the paired helical filaments of Alzheimer's disease. Brain Pathol 17:83–90
4. Alonso AC, Li B, Grundke-Iqbal I, Iqbal K (2008) Mechanism of tau-induced neurodegeneration in Alzheimer disease and related tauopathies. Curr Alzheimer Res 5:375–384
5. Querfurth HW, LaFeria FM (2010) Alzheimer's disease. N Engl J Med 362: 329–344
6. Wolfe MS (2006) The gamma-secretase complex: membrane-embedded proteolytic ensemble. Biochemistry 45:7931–7939
7. Mehta PD, Pirttila T (2002) Biological markers of Alzheimer's disease. Drug Dev Res 56:74–84
8. Haass C, Selkoe DJ (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nat Rev Mol Cell Biol 8:101–112
9. Jeibmann A, Paulus W (2009) Drosophila melanogaster as a model organism of brain diseases. Int J Mod Sci 10:407–440
10. Vosshall LB, Stocker RF (2007) Molecular architecture of smell and taste in *Drosophila*. Annu Rev Neurosci 30:505–533
11. Fahrbach SE (2006) Structure of the mushroom bodies of the insect brain. Annu Rev Entomol 51:209–232
12. Strauss R (2002) The central complex and the genetic dissection of locomotor behavior. Curr Opin Neurobiol 12:633–638
13. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81: 741–766
14. Selkoe DJ (2011) Alzheimer's disease. Cold Spring Harb Perspect Biol 3. doi: 10.1101/cshperspect.a004457. http://cshperspectives.cshlp.org/content/3/7/a004457
15. Greeve I, Kretzschmar D, Tschape JA et al (2004) Age-dependent neurodegeneration and Alzheimer-amyloid plaque formation in transgenic *Drosophila*. J Neurosci 24: 3899–3906
16. Fossgreen A, Bruckner B, Czech C et al (1998) Transgenic *Drosophila* expressing human amyloid precursor protein show gamma-secretase activity and a blistered-wing phenotype. Proc Natl Acad Sci U S A 95:13703–13708
17. Fischer JA, Giniger E, Maniatis T, Ptashne M (1988) GAL4 activates transcription in *Drosophila*. Nature 332:853–856
18. Crowther DC, Kinghorn KJ, Miranda E et al (2005) Intraneuronal Abeta, non-amyloid aggregates and neurodegeneration in a *Drosophila* model of Alzheimer's disease. Neuroscience 132:123–135
19. Rival T, Page RM, Chandraratna DS et al (2009) Fenton chemistry and oxidative stress mediate the toxicity of the b-amyloid peptide in a *Drosophila* model of Alzheimer's disease. Eur J Neurosci 29:1335–1347
20. Jahn TR, Kohlhoff KJ, Scotta M et al (2011) Detection of early locomotor abnormalities in a *Drosophila* model of Alzheimer's disease. J Neurosci Methods 197:186–189

Chapter 14

Chronic Mild Stress Assay Leading to Early Onset and Propagation of Alzheimer's Disease Phenotype in Mouse Models

Mar Cuadrado-Tejedor and Ana García-Osta

Abstract

A comprehensive chronic mild stress (CMS) procedure is presented, which consists in the application of unpredictable mild stressors to animal models in a random order for several weeks. This assay can be applied to Alzheimer's disease (AD) mouse models, leading to accelerated onset and increased severity of AD phenotypes and signs, including memory deficits and the accumulation of amyloid-β and phospho-tau. These assays open the way towards advanced studies on the influence of sustained mild stress, stress responses and pathways on the onset and propagation of Alzheimer's disease.

Key words Chronic mild stress, Stressors, Alzheimer's disease, Mouse models, Tg2576 mice

1 Introduction

The chronic mild stress (CMS) model was originally developed by Paul Willner and colleagues in the late 1980s as a model of depression in rodents. It consists of repeated exposure to a variety of mild and unpredictable stressors over a sustained period of time (from 10 days to 8 weeks) [1, 2]. This experimental procedure which is considered more similar to the forms of stress experienced by humans in everyday life [3], was first designed to mimic anhedonia [2], a core symptom of clinical depression, i.e., loss of interest in normally rewarding stimuli.

On the basis of different epidemiologic studies considering depression as a risk or prodromal factor of cognitive decline [4], several studies have demonstrated that chronic stress may accelerate the onset and progression of Alzheimer's disease (AD)-like phenotype in different murine models of AD [5–7]. However, most of these studies involved the use of strong stressors. More remarkably, recent studies in our lab have demonstrated that the application of chronic mild stress, which is thought to represent a

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_14,

good model of everyday stress in humans, may also influence the onset and development of the main signs of AD, as this approach, induced cognitive impairment and increased amyloid and phospho-tau pathology in young-AD transgenic mice (Tg2576) (ref. 8).

2 Materials

The implementation of the chronic mild stress (CMS) procedure leading to early onset and accelerated progression of the AD-phenotype in transgenic Tg2576 mice requires:

2.1 AD Animal Models. Transgenic Mice

Tg2576 transgenic mice that overexpress the human 695-aa isoform of the human amyloid precursor protein (hAPP) containing the Swedish double mutation (hAβPP^{Swe}: (AβPP695) Lys670-Asn, Met671-Leu) driven by the hamster prion promoter.

In these mice, the brain Aβ peptide content increases exponentially between 6 and 12 months of age, and memory impairment measured in the Morris water maze (MWM) test is evident by 12–15 months [9–11]. At the age of 4 months or earlier these animals do not display any features of AD. For the assay, 4-month-old female Tg2576 mice are used ($n = 10$) (*see* **Note 1**).

2.2 Materials, Mild Stressors for the CMS Procedure

1. Intermittent bell (10 dB 1/10 s).
2. A novel object, e.g. a wood piece similar to the ones used in the novel object recognition (NOR) procedure [12].
3. A low intensity stroboscopic lamp.
4. Radio (to apply white noise, un-tuned radio noise).
5. Morris water maze (also know as Morris water navigation task) [11].

3 Methods

3.1 Animal Selection

1. Select 4-month female Tg2576 mice, $n = 10$ per group (*see* **Note 1**).
2. Weight the animals and divided them in two groups in a random order:
   (a) Control group: Tg2576.
   (b) Chronic mild stress group: Tg2576-CMS.
3. House the Tg2576-CMS mice in individual cages in the room where the chronic mild stress procedure will be performed, and leave them in that environment for 2 weeks, to habituate (*see* **Note 2**). As part of the stress protocol, during the last week of the CMS procedure the animals must be changed to the "behavioral room" where the MWM will be performed.

4. House the control animals (Tg2576 mice) in normal conditions (see below) until the last week, when they must be changed to the "behavioral room" (where the water maze task, MWM, is located) (*see* **Note 3**).

Unless otherwise specified, food and water should be available ad libitum for the duration of the experiments. The animals should be maintained in a temperature- and humidity-controlled environment with a 12 h light-dark cycle.

3.2 Chronic Mild Stress (CMS) Procedure

The following stressors should be scheduled over a 1-week period and repeated throughout the 6-week experiment. The 1-week schedule is summarized in Table 1. However, we compile here the complete procedure including the period of application of each stressor and references to notes related to each event.

3.2.1 First Day

1. Retire water and food from each cage during 8 h, from 9.00 (9 am) to 17.00 h (5 pm).
2. From 9:00 to 12:00 h, intermittent bell ringing (10 dB, 1/10 s).
3. From 12:00 to 15:00 h (3 pm) introduce a novel object in the cage (*see* **Note 4**).

Table 1
One-week schedule of stressors for the CMS procedure

Day	Stressors
First	Intermittent bell (10 dB, 1/10 s) Placement of a novel object in the home cage (3 h) Water and food deprivation (8 h)
Second	Low intensity stroboscopic illumination (in the dark 2 h) Illumination and removal of nesting material overnight (12 h)
Third	Swimming in cold water (18 °C) for 5 min Turn off the light during the day (3 h) Move the rack to another room (6 h)
Fourth	Soiled bedding (200 ml of water per cage; 6 h) Removal of nesting material overnight (12 h) Placement of a novel object in the home cage (3 h)
Fifth	Rat odor (saw dust from rat cages; 8 h) Cage tilted (45°, 8 h) White noise (an un-tuned radio 4 h) Water and food deprivation (overnight)
Sixth	Intermittent bell (10 dB, 1/10 s) Illumination overnight (12 h)
Seventh	Low intensity stroboscopic illumination (in the dark 10 h) Cage tilted 45° and removal of nesting material (overnight)

Total duration of the whole CMS procedure, 6 weeks

3.2.2 Second Day

1. At 9:00 h turn the light off, and turn on a low stroboscopic illumination during 2 h.
2. At 20:00 h remove the nesting material from each cage and, to change the light-dark cycle, turn the light on during 12 h.

3.2.3 Third Day

1. At 9:00 h using the water maze, introduce each CMS-mouse in cold-water (18 °C) during 5 min.
2. From 10:00 to 13:00 h turn off the light.
3. From 13:00 to 19:00 h move the rack to the corridor of the animal facility.

3.2.4 Fourth Day

1. From 9:00 to 12:00 h introduce a novel object in the home cage (*see* **Note 4**).
2. At 12:00 h add 200 ml of water in each cage (soiled bedding) and leave it for 6 h.
3. At 20:00 h remove the nesting material from each cage during 12 h.

3.2.5 Fifth Day

1. From 12:00 to 20:00 h add sawdust material from rat cages during 8 h in each CMS-mouse cage and tilt them over the wall, tilted 45° (*see* **Note 5**).
2. From 12:00 to 16:00 h use an un-tuned radio as a white noise source.
3. At 20:00 h change the nesting material, adding new and clean sawdust and retire water and food from each cage during 12 h.

3.2.6 Sixth Day

1. From 9:00 to 12:00 h, intermittent bell ringing (10 dB, 1/10 s).
2. To change the light-dark cycle leave the light on for 12 h.

3.2.7 Seventh Day

1. At 12:00 h turn the light off, and turn on a low stroboscopic illumination during 8 h.
2. At 20:00 h remove the nesting material from each cage and tilt them over the wall, tilted 45°.

3.3 CMS Validation

As a final step, the CMS model is validated. To validate the model, since CMS has been reported to impair animal spatial cognitive function [13], the spatial learning and memory capacity can be tested by using the Morris water maze task (MWM). Together with this, to determine long-lasting biochemical changes, the animals can be sacrificed long time after completing the CMS procedure (*see* **Note 6**).

A schematic diagram with the whole duration of the experimental protocol, CMS procedure, MWM validation, and study of long-lasting effects reported revealing acceleration of the onset and progression of AD phenotype in Tg2576 mice [8], is shown in Fig. 1. The application of CMS procedure to 4-month Tg2576

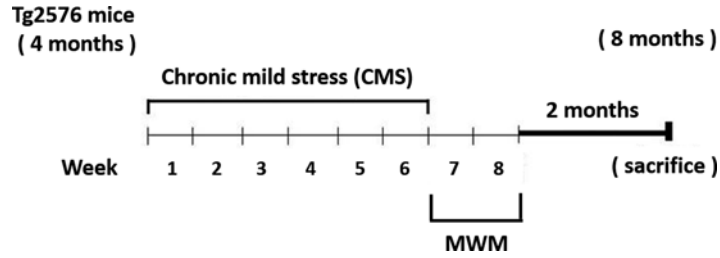


Fig. 1 Schematic diagram with the whole duration of the experimental protocol, including CMS procedure, MWM validation and study of long-lasting effects. The 4-month-old Tg2576 mice were exposed to the chronic mild stress procedure which lasts 6 weeks. Immediately after the CMS completion, the Morris water maze (MWM) test was conducted and 2 months later the animals were sacrificed

mice significantly affected the behavior of Tg2576 mice in the MWM test accelerating the onset of memory deficits observed in these transgenic animals. By using monoclonal anti-Aβ and/or anti-phospho-tau antibodies, biochemical analysis also demonstrated an increase in both amyloid and phospho-tau-pathology in the brain of Tg2576 mice subjected to CMS procedure [8]. The combined CMS protocol and advanced biochemical techniques can be applied to other AD mouse models towards characterization of AD impaired pathways, stress responses and networks, for advanced studies on the influence of chronic mild stress on the onset and propagation of Alzheimer's disease.

4 Notes

1. The Aβ pathology is influenced by mouse gender [14], particularly in the Tg2576 line [15]. Thus, in the design of studies it is important to take this variable into account. In our particular case, females were selected since the aggressive behavior of males might influence the stress environment.
2. Any change in the environment may affect the animals, thus, it is important to habituate the animals to the new conditions, and leave them in that environment for at least 2 weeks.
3. Habituate the animals to the new environment housing them in the "behavioral room" (where the water maze task is located) 1 week before initiating the test.
4. Introduce in each cage a wood piece similar to the ones used in novel recognition tests [12].
5. Spread some rat sawdust material into each cage and maintain them tilted at 45° over the wall to maintain this sawdust in continuous contact with the animal.

6. To validate the CMS model, memory function was evaluated by using the MWM test in the last week of the CMS procedure. To analyze long-lasting changes on the brain of the CMS-Tg2576 mice, the animals were sacrificed 2 months later.

Acknowledgements

This work was supported by the Fund for Health Research-Fondo de Investigación Sanitaria (FIS project 11/02861).

References

1. Willner P (1984) The validity of animal models of depression. Psychopharmacology (Berl) 83:1–16
2. Willner P, Towell A, Sampson D et al (1987) Reduction of sucrose preference by chronic unpredictable mild stress, and its restoration by a tricyclic antidepressant. Psychopharmacology (Berl) 93:358–364
3. Harkin A, Houlihan DD, Kelly JP (2002) Reduction in preference for saccharin by repeated unpredictable stress in mice and its prevention by imipramine. J Psychopharmacol 16:115–123
4. Jorm AF (2001) History of depression as a risk factor for dementia: an updated review. Aust N Z J Psychiatry 35:776–781
5. Dong H, Goico B, Martin M et al (2004) Modulation of hippocampal cell proliferation, memory, and amyloid plaque deposition in APPsw (Tg2576) mutant mice by isolation stress. Neuroscience 127:601–609
6. Jeong YH, Park CH, Yoo J et al (2006) Chronic stress accelerates learning and memory impairments and increases amyloid deposition in APPV717I-CT100 transgenic mice, an Alzheimer's disease model. FASEB J 20: 729–731
7. Lee KW, Kim JB, Seo JS et al (2009) Behavioral stress accelerates plaque pathogenesis in the brain of Tg2576 mice via generation of metabolic oxidative stress. J Neurochem 108: 165–175
8. Cuadrado-Tejedor M, Ricobaraza A, Frechilla D et al (2012) Chronic mild stress accelerates the onset and progression of the Alzheimer's disease phenotype in Tg2576 mice. J Alzheimers Dis 28:567–578
9. Hsiao K, Chapman P, Nilsen S et al (1996) Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. Science 274:99–102
10. Westerman MA, Cooper-Blacketer D, Mariash A et al (2002) The relationship between Abeta and memory in the Tg2576 mouse model of Alzheimer's disease. J Neurosci 22: 1858–1867
11. Reed MN, Liu P, Kotilinek LA, Ashe KH (2010) Effect size of reference memory deficits in the Morris water maze in Tg2576 mice. Behav Brain Res 212:115–120
12. Ennaceur A, Delacour J (1988) A new one-trial test for neurobiological studies of memory in rats. 1: behavioral data. Behav Brain Res 31:47–59
13. Song L, Che W, Min-Wei W et al (2006) Impairment of the spatial learning and memory induced by learned helplessness and chronic mild stress. Pharmacol Biochem Behav 83:186–193
14. Ribé EM, Pérez M, Puig B et al (2005) Accelerated amyloid deposition, neurofibrillary degeneration and neuronal loss in double mutant APP/tau transgenic mice. Neurobiol Dis 20:814–822
15. Callahan MJ, Lipinski WJ, Bian F et al (2001) Augmented senile plaque load in aged female beta-amyloid precursor protein-transgenic mice. Am J Pathol 158:1173–1177

Chapter 15

Gene Expression Studies on Human Trisomy 21 iPSCs and Neurons: Towards Mechanisms Underlying Down's Syndrome and Early Alzheimer's Disease-Like Pathologies

Jason P. Weick, Huining Kang, George F. Bonadurer III, and Anita Bhattacharyya

Abstract

The cause of Alzheimer disease (AD) is not well understood and there is no cure. Our ability to understand the early events in the course of AD is severely limited by the difficulty of identifying individuals who are in the early, preclinical stage of this disease. Most individuals with Down's syndrome (DS, trisomy 21) will predictably develop AD and that they will do so at a young age makes them an ideal population in which to study the early stages of AD. Several recent studies have exploited induced pluripotent stem cells (iPSCs) generated from individuals with familial AD, spontaneous AD and DS to attempt to identify early events and discover novel biomarkers of disease progression in AD. Here, we summarize the progress and limitations of these iPSC studies with a focus on iPSC-derived neurons. Further, we outline the methodology and results for comparing gene expression between AD and DS iPSC-derived neurons. We highlight differences and commonalities in these data that may implicate underlying genes and pathways that are causative for AD.

Key words Pluripotent stem cells, Microarray, Expression analysis, Neurons, Down's syndrome, Alzheimer's disease

1 Introduction

Alzheimer's disease (AD) is characterized by progressive dementia associated with amyloid plaque formation, neurofibrillary tangles (NFTs) and cortical neuron degeneration. AD typically begins with subtle memory failure that becomes more severe and is eventually incapacitating. Amyloid plaques and NFTs are considered late events in AD pathology, so defining the initial events in AD pathology is key to understanding its progression. A number of hypotheses have been proposed as to the earliest changes that underlie AD symptoms. For instance, synapse loss is believed to be one of the earliest events in neurodegeneration associated with AD [1] and has been substantiated by the decreased number of

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_15,

synapses in post-mortem AD brain [2–5]. Oxidative stress (OS) has also emerged as a potential early systemic trigger of AD pathology, where post-mortem AD brains and those of AD animal models show indications of oxidative damage [6–11].

Defining the root cause of AD, whether it is triggered by oxidative stress, synapse loss, Aβ deposits, tau phosphorylation or other mechanisms relies on the ability to analyze the earliest events in AD. Yet, the study of disease progression in AD has been hindered by the fact that diagnosis is confirmed by post-mortem evidence of amyloid plaques in the brain [12, 13]. Furthermore, most individuals who are diagnosed are no longer in early stages of the disease. Because the age of onset of AD is generally over age 65, the first symptoms are often mistakenly attributed to aging or stress. This makes it difficult to identify individuals who are in early stages of AD and to define early events in the disease. Individuals with many familial forms usually develop AD symptoms between 50 and 65 years of age, providing an early window into AD. Yet, early-onset familial AD is relatively uncommon, accounting for about 5 % of total Alzheimer's disease [14, 15] or about 250,000 cases in the United States.

In addition to familial AD caused by single gene mutations or duplications, individuals with Down's syndrome (DS) develop AD before age 65 and provide a good model for studying AD pathology progression. The incidence of DS in the US is approximately 1 in 1,200 births resulting in a total of approximately 250,700 individuals with DS in the U.S. [16]. Adults with DS are at an extremely high risk for developing AD, with most individuals over age 40 showing amyloid deposits (based upon autopsy findings) and over half of DS adults older than 60 years of age diagnosed with AD [17–26].

DS is caused by triplication of human chromosome 21 (Hsa21) and many AD candidate genes are located on Hsa21. These genes include *APP*, *DYRK1A*, and *SOD1*. It is believed that the presence of an extra copy of APP on chromosome 21 provides more substrate for production of Aβ peptide and puts individuals with DS at considerably greater risk than the general population for early Aβ plaque deposition and the appearance of AD symptoms. In support of this notion, mice that overexpress only human APP develop early biochemical and cognitive hallmarks of AD [27]. Dual specificity tyrosine-phosphorylation-regulated kinase 1A (*DYRK1A*) can phosphorylate Tau [28] and may therefore be involved in its hyperphosphorylation and subsequent aggregation. Superoxide dismutase 1 (*SOD1*) is responsible for destroying free superoxide radicals and its imbalance may affect levels of oxidative stress. In addition, with up to 500 genes located on Hsa21, it is possible that other genes contribute to the progression of AD in DS individuals. Elucidation of the mechanisms of these genes can inform both disease progression and potential therapeutic strategies.

Taken together, there are several advantages to studying AD progression in individuals with DS: (1) DS is diagnosed at birth (or prenatally), (2) DS is a predictor of individuals who will likely develop AD, (3) DS affects more people than early onset familial AD, (4) DS individuals develop AD symptoms before age 40, and (5) many AD candidate genes are encoded by chromosome 21. These traits make DS individuals a unique population in which to examine early stages of AD progression and identify early biomarkers.

As described, it is crucial to define the earliest events in AD so as to study disease cause and progression. DS individuals provide a unique population that will reliably develop AD at an early age that can be used to study early neuropsychological and biochemical events in AD. Yet, there remains the problem of developing a system in which to identify early cellular and molecular events in AD pathology. Induced pluripotent stem cell (iPSC) technology allows the creation of disorder-specific human cells to define errors in human neurodevelopmental and neurodegenerative disease [29–31]. The application of iPSCs to AD has been demonstrated by recent studies that identified cellular pathologies in AD neurons as well as altered gene expression patterns [32–36] (see below). In addition, iPSCs and their neuronal derivatives have been used to identify early cellular abnormalities, and in one case, directly related to known alteration of AD neurons. Therefore, comparisons of AD and DS iPSCs can now be used to identify underlying genes and pathways that are causative for AD.

2 Methodological Considerations for iPSC Studies

While comparisons between iPSC-derived neurons/glia and DS/AD patient tissue samples may provide the most promising avenue for uncovering mechanisms of disease, iPSC technology is in its infancy and it is necessary to understand the sources of variability we can expect from in vitro studies prior to moving to in vivo/ex vivo exploration. There are multiple factors in iPSC studies that introduce variability and affect the ability to compare data from multiple studies including patient differences, iPSC reprogramming methods, and neuronal differentiation paradigms. Inherent genetic variation among individuals due to genetic diversity as well as disease presentation presents a major challenge to iPSC disease modeling [37, 38]. Further, epigenetic and copy number diversity add another layer of complexity [39] which may plague iPSCs to a greater extent than other samples. Many of these problems can be overcome by using cells from enough different individuals to enable statistically meaningful results. Alternatively, either engineered or spontaneously-generated isogenic cell lines can provide a more practical alternative to limit genetic diversity. For investigations of

single gene mutations, genetic modifications to repair these defects have been sufficient to reverse cellular phenotypes [40]. New genetic technologies such as TALENs and CRISPR/Cas9 will likely play a major role moving forward with in vitro detection of disease phenotypes [41–44]. In addition, new methods allow the silencing of entire chromosomes to correct aneuploidy [45], paving the way for complex gene regulatory analyses.

A potential confounding factor for iPSC research in general, and for comparing data across multiple cells lines from different laboratories, is the method of choice for reprogramming. While most published studies have used integrating retroviruses [30], recent studies have also utilized non-integrating viruses such as Sendai virus [46] as well as episomal vectors [47] to deliver the reprogramming factors Oct4, Klf4, Sox2, cMyc (OKSM). However, while integration of retroviruses has been postulated to cause genomic instability and transcriptional alterations, no evidence of this has been reported to significantly alter neuronal differentiation or identification of disease phenotypes. Additionally, exogenous retroviruses are quickly suppressed (within weeks) in newly-generated iPSCs, leading to activation of the endogenous OKSM factors [30], but silencing does not appear to be required for directed differentiation [48]. Thus, while newer methods are quickly adopted by iPSC researchers, little data exist to suggest retrovirus use is detrimental to the study of iPSCs and their differentiated progeny, and use of "original" cell lines should continue. Additionally, the somatic cell source has been a point of debate for iPSC researchers. While skin fibroblasts as the somatic cell source still dominate in the published literature due to ease of procurement and reliable reprogramming, successful creation of iPSCs has been demonstrated from other cells including lymphocytes [49].

Perhaps the largest determinant of variability for comparison of iPSC studies with each other, and across model systems, is that of differentiation method and resulting neuronal populations. Directed differentiation using exogenous factors has been established to generate numerous transmitter- and region-specific neuronal subtypes including midbrain dopamine (DA), spinal motoneurons (MNs), medium spiny neurons (MSNs), basal forebrain cholinergic neurons (BFCNs), and forebrain (FB) cortical-like glutamatergic/GABAergic neurons. It is generally agreed that for cell replacement therapies, it will be critical to match the transplanted neuronal subtype with that of the primary degenerative phenotype (e.g. DA neurons for Parkinson's disease, MSNs for Huntington's, etc.). This is likely true for mechanistic studies as well. While many other lineages have been generated [50–52], we have outlined the most well-characterized protocols for generation of dorsal and ventral FB (including cholinergic neurons), DA, BFCNs, and MNs in Fig. 1, as these are of primary importance for a number of neurological disorders, including DS and AD.

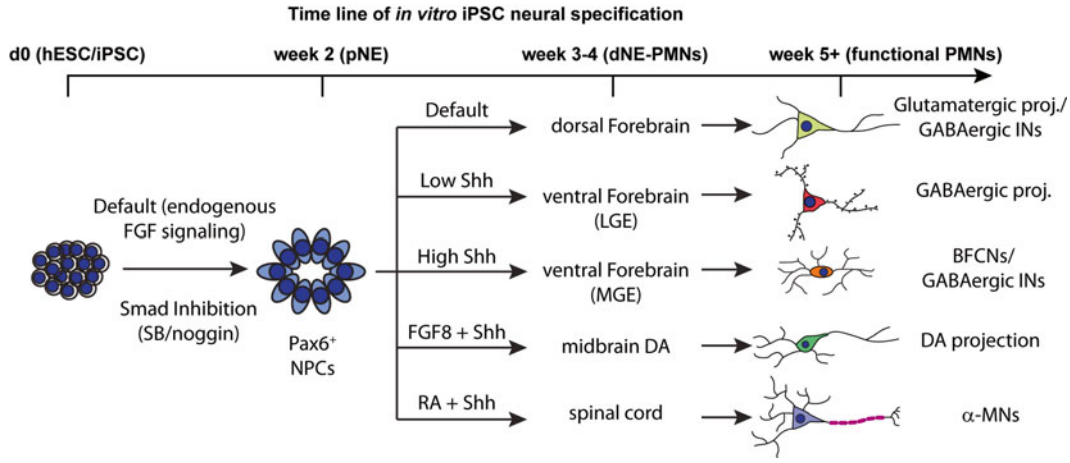


Fig. 1 Directed differentiation of human pluripotent stem cells to transmitter- and region-specific neuronal subtypes. Human pluripotent stem cells (hESC or iPSC) can be differentiated into PAX6+ primitive neuroectoderm (pNE) within 2 weeks in culture either by relying on endogenous FGF signaling or by dual SMAD inhibition. Without additional morphogens (Default), these pNE will differentiate into dorsal forebrain neurons (Glutamatergic and GABAergic). pNE can also be patterned to various lineages via exposure to exogenous patterning factors. Addition of a ventralizing factor such as sonic hedgehog (Shh) allows cells to retain a forebrain phenotype, but will induce ventralization to GABAergic interneurons (INs) and basal forebrain cholinergic neurons (BFCNs). Treatment of pNE with FGF8 and Shh results in differentiation of midbrain dopaminergic (DA) neurons. Treatment of pNE with caudalizing factors such as retinoic acid with Shh can lead to motor neuron (MN) specification

Most recent studies use one of two primary methods to initiate differentiation to an ectodermal lineage. For human embryonic stem cells (hESCs) and iPSCs that are inherently primed toward the neural lineage, exposure to minimally supportive media (e.g. DMEM/F-12 + N2 supplement) is sufficient to allow differentiation along a "default" program that includes a primitive neuroectodermal fate defined by robust expression of the paired homeobox 6 (*PAX6*) gene [53, 54]. This method, pioneered by Su-Chun Zhang at the University of Wisconsin-Madison, has been shown to rely on endogenous FGF acting via FGF receptors to activate the MAPK pathway [55, 56]. A second method, introduced by the Studer Laboratory at Sloan Kettering, demonstrates that hESCs and iPSCs can be directed to ectoderm by inhibition of the Smad pathway using inhibitors of the transforming growth factor beta (TGF-β) and activin/nodal signaling [57]. Both of these methods produce robust *PAX6* expression in early neuroectodermal cells within the first 10 days of differentiation. If no other factors are present during subsequent stages of differentiation the primitive *PAX6*$^+$ neural progenitor cells (NPCs) will proceed through a definitive neuroectodermal stage and go on to become dorsal forebrain (dFB), cortical-like neurons, and then astrocytes with prolonged culturing periods [58]. These dFB neuronal populations typically include robust

numbers of excitatory glutamatergic and inhibitory GABAergic cells (Fig. 1, upper pathway). Interestingly, recent reports suggest that a combination of retinoic acid and dual Smad inhibition can enhance glutamatergic projection neuron differentiation [59]. As retinoic acid is a potent neural inducer, it may enhance early-born neurons to differentiate, mimicking endogenous retinoid signaling from meningeal cells in the developing cortex [60].

It is now generally agreed that *PAX6*$^{+}$ NPCs can also be patterned to various lineages via exposure to exogenous patterning factors [61, 62]. For example, treatment of primitive NPCs with caudalizing factors such as retinoic acid can induce Hox gene expression, which is required for the establishment of spinal cord fates. Additional treatment using a ventralizing factor such as sonic hedgehog (Shh) can lead to the induction genes required for MN specification such as *HB9*, acetylcholine (ACh; the MN neurotransmitter), and *ISL1* (Fig. 1, lower pathway). Similarly, omission of the caudalizing factors during Shh treatment allows cells to retain a FB phenotype, but will induce ventralization to GABAergic interneurons (INs) and BFCNs (Fig. 1, pathways 3–4). This is thought to involve a gradient of Shh treatment both in terms of concentration and timing. Early, high Shh levels will bias the cells toward more ventral fates, leading to increased proportions of BFCNs [63], while moderate Shh treatment allows for increased production of GABAergic interneurons that derive from the medial and lateral ganglionic eminences [64].

It is critical to note that gene expression patterns in cells of various regional and transmitter phenotypes differ substantially from one another, as indicated by changes in differential expression of unique markers. This underscores the fact that these are truly distinct populations, highlighting the need to compare similar populations between control and disease conditions, as well as across studies, to reveal relevant disease phenotypes. While AD affects many neuronal populations as disease progression enters later stages, neocortical and cholinergic neurons are the primary affected population during early phases in AD patients and animal models [65]. Thus, to understand how early deficits in DS neurons may underlie later problems in AD it is important to generate appropriate neuronal populations, such as cortical-like neurons.

3 Defining Mechanisms Underlying Early Alzheimer's Disease-Like Pathologies in Down's Syndrome

3.1 Pathology and Gene Expression of AD iPSC-Derived Neurons

Three recently published studies examined hallmarks of AD pathology as well as transcriptome analysis [32–34] using iPSC-derived neurons generated from fibroblasts taken from AD patients, as well as non-demented controls (NDCs). Interestingly, these reports took advantage of very different patient populations of familial AD

in an effort to reveal molecular deficits in AD neurons. The first report from the Suzuki laboratory created AD iPSCs from patients with autosomal-dominant mutations of the presenilin genes [34] while Inoue and colleagues [33] chose patients who expressed autosomal-recessive mutations of the *APP* gene. In contrast, Goldstein and colleagues created AD iPSCs from patients overexpressing *APP* (APP^{Dp}), which may have the most relevance to studies of DS [118]. In addition, two reports generated iPSCs from sporadic AD patients, although gene expression was not assessed from these samples [32, 33]. Because this chapter is focused on gene expression changes across relevant cellular populations affected in both DS and AD (neurons), we will target our discussion to those studies that performed global transcriptome analyses related to the familial cases as they relate to general AD phenotypes and how gene expression from DS studies may help inform AD pathophysiology.

As mentioned, the magnitude of gene expression changes between two iPSC-derived cell populations will primarily reflect the differentiation state of the cells, while disease phenotypes can be expected to be smaller in magnitude. Goldstein and colleagues [118] found that neurons from APP^{Dp} and sporadic AD patient-derived iPSCs showed significantly elevated Aβ, p-Tau, and GSK-3β, a key kinase involved with post-translational modifications of the amyloid and Tau proteins, as well as defects in the formation of early endosomes, all hallmarks of AD patients and animal models [66, 67]. However, reported gene expression patterns from neurons that are directly relevant to AD are likely to be subtle due to the choice of differentiation paradigm. Israel and colleagues co-cultured neural progenitor cells with PA6 cells, a method that generates a large proportion of midbrain DA neurons (Fig. 1; [68, 69]). In addition, their data are compared with fetal brain samples that are not described in detail but likely represent a mixed population with a majority of cells from the cerebral cortex. Examination of the gene expression data reveal significant expression of midbrain and DA neuronal markers such as Iroquois homeobox factors [70] and tyrosine hydroxylase in the iPSC-derived neurons compared with brain tissue. In contrast, fetal brain samples showed enriched expression of cortical markers such as *SATB2*, *LHX2*, *TBR1*, and *EMX2* [32]. It is not surprising then, that the neuronal population analyzed contains relatively minimal proportions of glutamatergic/GABAergic neurons as measured by immunocytochemistry and physiology [32]. Thus, it is difficult to assess the changes that are specific to AD cells rather than those that are due to neuronal specification when comparing this dataset and other DS/AD cortical neuronal populations (see below). However, as the iPSCs and their neuronal derivatives carried a duplication of the *APP* locus (APP^{Dp}), APP transcript levels were significantly upregulated compared with NDCs and fetal brain samples that carried no know genetic abnormalities.

In contrast to the methods used by Israel and colleagues, Kondo et al. [33] use differentiation methods similar to those developed by the Studer laboratory [57], using inhibitors of TGF-β and activin/nodal signaling to initiate differentiation and generate cortical-like cells [71]. In their study, patient fibroblasts contained two different mutations of the APP gene (APP-E693Δ and APP-V717L), but possessed only two copies. Therefore, no significant increase in APP transcript was observed in their AD iPSC samples. However, both mutations demonstrated robust, neuron-specific effects on Aβ levels and the production of reactive oxygen species, which were blocked by inhibitors of GSK3β. Kondo et al. [33] also performed transcriptome analysis on samples of neurons carrying the APP-E693Δ, a rare, autosomal-recessive mutation that causes early-onset AD but without extracellular Aβ plaque deposition [72]. Interestingly, only 50 identified genetic loci were differentially regulated (>1.5-fold) in AD neurons, some previously implicated in AD but many novel transcripts as well. Significant increases were observed in oxidative stress (OS) genes such as peroxiredoxin, oxidoreductase and peroxidase activities for neurons carrying the APP-E693Δ mutation, pathways that have previously been implicated in AD [11]. In contrast, a number of β-glucuronidase isoforms are down-regulated, suggesting that multiple metabolic pathways are disrupted that may involve mitochondrial, ER, and golgi functions. In addition, Kondo et al. [33] observed a number of synaptic/cell adhesion markers, zinc finger proteins, and regulators of apoptosis were altered as well. Thus, these AD iPSC studies identified multiple deficits that are hallmarks of AD in neurons as well as novel changes, supporting the use of the iPSC platform even for a disease that takes decades to manifest in human patients.

3.2 DS iPSCs as a Model of Early AD

As mentioned, DS patients represent a unique population of individuals that may help to uncover early deficits of AD pathology. To date, three reports have explored various aspects of DS from iPSC-derived neurons [73–75], and each has focused on cortical-like neurons, with subtle differences in methodology. For instance, while Weick and colleagues used "default" methods to generate mixed populations of excitatory and inhibitory neurons [75], Shi and colleagues used methods to enrich for excitatory glutamatergic neurons [74], while Briggs and colleagues used the dual SMAD inhibition method [73]. All three studies used different methods of reprogramming, from retroviral and sendai virus transduction, as well as episomal methods. Interestingly, despite evidence for aberrant neuronal differentiation in DS brain [76–84], none found that DS iPSCs were deficient in their ability to differentiate to neuroepithelia and post-mitotic neurons compared with control cells. Thus, the results from these studies are can be directly compared, with a relatively high degree of confidence that methodological differences play a minimal role in the differences observed (see below).

DS iPSCs and their neuronal derivatives display aberrant phenotypes consistent with previous human and animal studies of DS neurodevelopment as well as phenotypes consistent with early AD pathology [73–75]. All studies found increased expression of APP in DS neurons, which would suggest that increased APP is available for proteolytic processing in these cells [73–75]. In fact, Shi et al. [74] found that DS neurons from both iPSCs and hESCs demonstrated elevated Aβ species (both 40 and 42), a decreased Aβ40:Aβ42 ratio, amyloid aggregate formation, as well as hyperphosphorylated tau in neurons cultured for extended periods (>60 days). Therefore, DS iPSC-neurons display similar phenotypes to AD iPSC-neurons. Similarly, all studies found that DS iPSCs and neurons exhibited significant increases in OS markers and/or increased sensitivity to reactive oxygen species (ROS) challenge. Metabolism and oxidative stress have been consistently reported as an underlying target of dysfunction in both DS and AD [11, 85].

Loss of synapses represents a major clinical feature of AD progression and synaptic abnormalities are correlated with both human and animal models of DS [86–89]. As for human DS iPSC-derived neurons, it was shown that both excitatory and inhibitory synaptic activity was diminished in DS neurons compared with controls [75]. This was paralleled by a decrease in synapsin-1 punctae on DS neurites, indicating either a failure to form similar numbers of synapses, or instability of synaptic junctions, which are then subsequently lost. In contrast, no deficits in synaptic punctae were found by Shi et al. [74], who analyzed the proportion of synaptophysin and PSD-95 doubled labeled punctae along iPSC-derived neurites. While Israel et al. performed some quantification on the physiological properties of AD neurons they also did not find differences between control and AD cells and concluded that "extended culture periods may be required to study Alzheimer's disease-associated loss of synaptic proteins" [118]. Because all investigations examined relatively early timepoints (<100 days in vitro), more functional data is needed on the iPSC-derived neurons from individuals with AD and DS to determine whether this is a repeatable phenotype in cultured cells.

3.3 Gene Expression Changes in DS Related to AD

Due to trisomy of Hsa21 in DS, in terms of number of genes altered, transcriptome changes have been found to be primarily a function of gene duplication of the genes on Hsa21 [90]. However, the largest magnitude of transcript changes primarily occur on genes located on autosomes and sex chromosomes other than Hsa21 [90–93]. Previous Gene Ontology (GO) analyses have pointed to alterations in gene families involved with the usual suspects (APP, Aβ and Tau), as well as oxidative stress, neuronal differentiation, a variety of second messenger cascades, and synaptic development/loss [91, 94, 95].

These findings are largely recapitulated in DS iPSCs. Interestingly, both Briggs et al. [73] and Weick et al. [75] developed isogenic control lines that were disomic for Hsa21. This fact allowed expression arrays of DS cells with reduced background "noise" due to genetic variability. Both studies found significant gene dosage effects of Hsa21 genes, with similar numbers of genes overexpressed (63 and 125, respectively), with a small number underrepresented (7 and 14, respectively). The differences in number primarily reflect the cutoff value of the analysis. Increasing the cutoff of the Weick et al. dataset [75] from 1.2 to 1.5-fold reduces the number of overexpressed genes to 60, nearly identical to the number reported in Briggs et al. [73]. Interestingly, greater than 60 % of the altered genes were identical in the two datasets. However, some interesting differences are noteworthy. While APP and *DYRK1A* were both upregulated in iPSCs in the study by Weick et al. [75], APP was not changed and *DYRK1A* was significantly *downregulated* in Briggs et al. [73]. Furthermore, 40 % of upregulated genes differed between the populations and none of the downregulated genes were shared between the datasets, suggesting significant differences in genetic regulation between iPSC lines with the same underlying genetic defect.

With regard to common expression changes in DS and AD neuronal populations, we will focus our discussion by comparing data from Weick et al. [75] with the data obtained from Kondo et al. [33] as these two studies produced neurons of similar phenotypes (i.e. forebrain). To more accurately assess the utility of DS iPSC-derived neurons to inform early AD pathology we directly compared the expression arrays from both datasets. The microarray gene expression data sets from the two studies were generated using different Affymetrix platforms. Weick et al. [75] using GeneChip Human Genome U133 plus 2.0 Array while Kondo et al. [33] used GeneChip the Human Gene 1.0 ST Array. The major difference between the two platforms lies in the fact that former interrogates a few hundred bases proximal to the 3′ end of each mRNA species to approximate expression of the entire gene whereas the latter queries the entire transcript of each gene. Despite this difference, gene expression measured in both platforms are highly concordant [96] which makes it possible to compare the results from both platforms and also integrate the data sets into one analysis.

In this analysis we considered a number of given Gene Ontology (GO) terms such as response to reactive oxygen species (GO:0000302) and cellular response to oxidative stress (GO:0034599), processes known to be affected in both DS and AD. We tried to identify genes annotated at each of these GO terms that are differentially expressed between diseased samples and controls in both data sets. The analysis can be divided into low-level analysis (data preprocessing) and high-level analysis (statistical differential expression analysis).

Affymetrix arrays (in both platforms) use multiple probes to measure the same transcript. One low-level analysis step is to summarize these repeated measurements into a single value for each probeset while removing undesired sources of variation so that the resulting single values (estimates for gene expression level) of all the probesets reflect the true changes in mRNA abundance as accurately as possible. This was achieved in our analysis through the use of the robust multi-array average (RMA) [97] algorithm implemented in Bioconductor package oligo. The RMA was performed on the two data sets separately because of the difference in the platform.

The second step of the data preprocessing is to match the probesets between the two data sets [98]. Affymetrix provides a matching file available on the company's website which has 29,129 mappings corresponding to unique U133 plus 2 probeset IDs; for each of the 29,129 probesets in the DS data set we could find one and only one corresponding probeset in AD data set. Each pair of the matched probesets represents the same gene. We restricted our analysis to these 29,129 probesets of the first data set and corresponding probesets in the second data set and called them as matched data sets. We next performed the high-level analysis (i.e. differential expression analysis) to identify the genes that are differentially expressed in both data sets.

For each of the given GO terms we obtained all Affymetrix probesets that are annotated at that node, either directly or by inheritance, using function 'lookUp' of the Bioconductor package 'annotate'. The differential expression analyses were performed on the overlap of these probesets and those available in the matched data sets. In order to identify the genes that are differentially expressed in both data sets, we first performed student t-test for each probeset on both data sets, separately. Then we rank-ordered the maximum of the two p-values and considered the genes with smallest maximum p-values as significant if $p < 0.1$. This value was chosen due to the stringency of the comparisons in this high-level analysis and small number of samples in each group.

The data is illustrated by Gene Ontology (GO) results in Table 1, which indicates significant overlap in the pathways that are disrupted in both cell types. We performed GO analysis on terms with previously-indicated relevance to each disorder, and noticed significant overlap in the number of genes shared between the two datasets (column 5). For instance, for all GO terms examined, the number of genes shared between DS and AD (when present) had an average of 22.6 % overlap. Overall, 115 genes were found to be significantly altered in both datasets. Notable genes include *CAT*, *ITSN1*, *MAP2*, *MAPK1*, *PRKAR2A*, *PSEN2*, *RAB4A*, and *STX7*. These genes are involved with cell cycle regulation, oxidative stress, synaptic transmission, endosomal trafficking [99], and signal transduction from plasma membrane to the nucleus. Remarkably, significant similarities were found between the two datasets despite

Table 1
Comparison of microarray data from DS iPSC-derived neurons with AD iPSC-derived neurons

GO pathway	Total genes	Matched (analyzed)	Changed (AD)	Changed (DS)	Overlap (same genes)
Synaptic transmission	1,529	1,143	231	90	22
MAPK cascade	1,373	987	164	100	15
Glycosylation	591	413	66	45	6
Aging	506	379	79	34	6
Neuron apoptotic process	436	335	60	38	5
Oxidative stress	296	205	41	15	6
ATP metabolic processes	261	186	23	20	2
Synapse Assembly	188	142	27	15	3
Mitochondrial permeability	99	80	10	8	0
Beta-amyloid metabolism	36	25	4	2	1

Direct comparison of the expression arrays from Weick et al. [75] with the data obtained from Kondo et al. [33]. Gene Ontology (GO) terms of processes known to be affected in both DS and AD were analyzed in the datasets. Table shows a significant overlap in the pathways that are disrupted in both cell types

the lack of APP duplication in the AD iPSC-derived neurons. This result points to common pathways engaged by very different underlying mechanisms of AD pathogenesis.

Dysfunction of endosomes has been proposed to represent one of the earliest shared phenotypes of DS and AD, when Aβ levels are relatively low [66]. The retrograde signaling of neurotrophins through endosomal trafficking, specifically nerve growth factor (NGF), has been implicated in the neuronal cell death in AD and DS [100–103]. Proper NGF signaling requires endocytosis and retrograde transport, which is associated with activated components of the Ras-MAPK pathway located on endosomes. Additionally, members of the Rab family of GTPases play an integral role in the local processing of proteins during synaptic vesicle release and recycling. The syntaxins are a family of proteins involved with diverse vesicular docking and fusion events between various targets including the plasma membrane and other intracellular compartments. Interestingly, both *STX7* and *ITSN1* have been specifically associated with endosomal and lysosomal compartments [104–106], and *ITSN1* is known to interact with several proteins involved with synaptic vesicle recycling [107–109]. Thus, the simultaneous dysregulation of *MAPK*, *ITSN1*, *PSEN2*, *RAB4A*, and *STX7* supports the idea that endosomal signaling may be perturbed in both DS and AD neurons at early timepoints.

The presence of altered *PSEN2* is particularly interesting not only because it is expressed in intracellular vesicles, but because mutations in the Presenilin genes (*PSEN1* and 2) are strongly associated with familial AD [110] Multiple mutations in *PSEN1* and *PSEN2* are known to cause early-onset AD between the ages of 60–65 [111]. *PSEN1* and PSEN2 are part of the γ-secretase complex, which cleaves a number of membrane proteins, including APP. This proteolytic cleavage of the C-terminal end of APP, along with a second, N-terminal cut is required for production of the Aβ peptide [119]. While normal β-secretase activity primarily leads to the Aβ40 form, a small amount of Aβ42 can also be produced, which is more prone to aggregation and can cause neuronal damage. Mutations in *PSEN* lead to significant overproduction of the Aβ42, and an increase in the Aβ42/Aβ40 ratio, resulting in AD. Thus, it is curious to observe overproduction of a presumably normal *PSEN2* in these populations of DS and AD neurons. In the AD cells, a clear increase was observed in the Aβ42/Aβ40 ratio [33], while this was not studied in Weick et al. [75]. However, the DS neurons in Shi et al. [74] demonstrated significant increases in both Aβ species and alterations in the Aβ42/Aβ40 ratio. Thus, it is likely that overall increased processing of APP cleavage by *PSEN2*/β-secretase can lead to toxic levels of Aβ42.

In addition, a number of transcriptional regulators were found to be altered in both datasets, including *SP3*, *TLX2*, and multiple zinc finger proteins (*ZNF*s *22*, *248*, *37A*, *439*, *510*, and *675*). Interestingly, both *SP3* and *TLX2* have both been previously implicated in AD pathology. Using Baysian network analysis of six different datasets consisting of a total of 110 patients (62 AD and 48 controls), Yoo and Yoo identified altered expression of four genes, including *TLX2*, that showed the highest association with disease incidence [112]. Further, a study by Boutillier and colleagues showed that both *SP3* and *SP4* were both upregulated at the protein level, and associated with NFTs, in postmortem AD brains [113]. The ZNFs identified are part of a family that represents one of the most abundant proteins in eukaryotic genomes and have incredibly diverse functions [114]. However, many require the binding of zinc (Zn^{2+}) or other metals in their finger-like protrusions to regulate DNA transcription of a host of genes. The fact that most of the transcription factors identified here belong to the ZNF family, including *SP3*, suggest a general pattern of altered transcriptional response which is correlated with altered metal metabolism. While the metal hypotheses of AD suggest that direct interactions of copper (Cu^{2+}) and zinc (Zn^{2+}) with extracellular Aβ increase aggregation at the synapse [115], it is possible that changes in the availability of metal ions result from alterations in transcription factors that require these ions for activity within the nucleus.

Lastly, it is noteworthy to point out the absence of both *DYRK1A* and APP in this dataset. While there is an approximately

1.5-fold increase in APP in DS samples [75], the APP-E693Δ mutation does not lead to increased transcript expression. Thus, the lack of significance for APP expression change is not surprising when looking at the intersection of the two different data sets. However, the absence of *DYRK1A* is not as easily explained. As mentioned the *DYRK1A* gene is located on Hsa21, and is a serine/threonine protein kinase capable of phosphorylating tau protein at 11 serine and threonine residues, as well as threonine 212, a site that may prime it for further phosphorylation events by GSK-3β [28]. Moreover, DYRK1A protein levels have been found to be overexpressed in multiple AD patient tissue samples [116]. It has been hypothesized that inhibition of DYRK1A may be a potential treatment of the developmental defects of DS as well as the progression of AD pathology. Due to the lack of *DYRK1A* expression alterations in AD iPSC-derived neuronal samples may indicate that while it is important for DS-related AD pathologies it is not an early marker for all AD patients.

4 Future Perspectives

While the current iPSC studies provide proof of concept that both DS and AD can be modeled in a dish, additional gene expression analyses are needed to uncover underlying genes and pathways that are causative for AD but that are not confounded by the issues of sample variability stated previously. For example, it will be interesting to examine gene expression data from cortical-like neurons derived from familial AD patients that carry the APP^{Dp} duplication [32]. This single gene duplication is sufficient to induce early-onset AD symptoms in patients as well as both NFT and amyloidogenic pathology in iPSC-derived neurons. Thus, transcriptome analyses between these and DS neurons should provide an excellent platform to increase the signal-to-noise ratio for early expression changes that play causative role in the development of dementia. Coupling cross-comparisons of multiple datasets like the one performed here, along with analyses of the APP-V717L mutation [33], should strengthen identification of genes with an obligatory role rather than those that may be secondary to disease onset.

Whole-genome sequencing and large-scale genome-wide association studies of large populations will also assist in uncovering any underlying single nucleotide polymorphisms and gene mutations that link DS and AD at a single gene level. It may be that these types of studies, will both identify shared features of AD-like pathology as well as accelerate the segregation of various types of AD into categories based on molecular dysfunction.

As recently as 20 years ago, the life expectancy for individuals with DS was only 25 years. Since then, the life expectancy has risen to greater than 50, due in large part to the reduced institutionalization of individuals with DS and greater awareness and care for

individuals with developmental disorders in our society [16]. With the increased lifespan come additional health issues for DS individuals including premature aging and the development of AD. Yet, this situation also provides a potential resource for learning more about the development of AD. The recent implementation of the Down Syndrome Consortium Registry (DS-Connect) by the U.S. National Institutes of Health will enable researchers access to detailed information about DS individuals and provide research subjects that are likely in early stages of AD [117]. It is possible that DS may represent a single underlying cause of AD pathology that will only relate to a minority of AD patients. Nonetheless, information gleaned from studying DS will undoubtedly provide insight into early manifestations of AD neuropathology.

References

1. Selkoe DJ (2002) Alzheimer's disease is a synaptic failure. Science 298:789–791
2. Scheff SW, DeKosky ST, Price DA (1990) Quantitative assessment of cortical synaptic density in Alzheimer's disease. Neurobiol Aging 11:29–37
3. Scheff SW, Price DA (1998) Synaptic density in the inner molecular layer of the hippocampal dentate gyrus in Alzheimer disease. J Neuropathol Exp Neurol 57:1146–1153
4. Scheff SW, Sparks DL, Price DA (1996) Quantitative assessment of synaptic density in the outer molecular layer of the hippocampal dentate gyrus in Alzheimer's disease. Dementia 7:226–232
5. Scheff SW, Sparks L, Price DA (1993) Quantitative assessment of synaptic density in the entorhinal cortex in Alzheimer's disease. Ann Neurol 34:356–361
6. Subbarao KV, Richardson JS, Ang LC (1990) Autopsy samples of Alzheimer's cortex show increased peroxidation in vitro. J Neurochem 55:342–345
7. Lovell MA, Ehmann WD, Butler SM, Markesbery WR (1995) Elevated thiobarbituric acid-reactive substances and antioxidant enzyme activity in the brain in Alzheimer's disease. Neurology 45:1594–1601
8. Munch G, Thome J, Foley P et al (1997) Advanced glycation endproducts in ageing and Alzheimer's disease. Brain Res Brain Res Rev 23:134–143
9. Smith CD, Carney JM, Starke-Reed PE et al (1991) Excess brain protein oxidation and enzyme dysfunction in normal aging and in Alzheimer disease. Proc Natl Acad Sci U S A 88:10540–10543
10. Mecocci P, MacGarvey U, Beal MF (1994) Oxidative damage to mitochondrial DNA is increased in Alzheimer's disease. Ann Neurol 36:747–751
11. Zana M, Janka Z, Kalman J (2007) Oxidative stress: a bridge between Down's syndrome and Alzheimer's disease. Neurobiol Aging 28:648–676
12. Fillenbaum GG, van Belle G, Morris JC et al (2008) Consortium to Establish a Registry for Alzheimer's Disease (CERAD): the first twenty years. Alzheimers Dement 4:96–109
13. Mirra SS, Heyman A, McKeel D et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41:479–486
14. Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. Neuron 68:270–281
15. Miyoshi K (2009) What is 'early onset dementia'? Psychogeriatrics 9:67–72
16. Presson AP, Partyka G, Jensen KM et al (2013) Current estimate of Down syndrome population prevalence in the United States. J Pediatr 163:1163–1168
17. Franceschi M, Comola M, Piattoni F et al (1990) Prevalence of dementia in adult patients with trisomy 21. Am J Med Genet 7(Suppl):306–308
18. Lai F, Williams RS (1989) A prospective study of Alzheimer disease in Down syndrome. Arch Neurol 46:849–853
19. Lott IT, Head E, Doran E, Busciglio J (2006) Beta-amyloid, oxidative stress and Down syndrome. Curr Alzheimer Res 3:521–528

20. McCarron M, Gill M, McCallion P, Begley C (2005) Health co-morbidities in ageing persons with Down syndrome and Alzheimer's dementia. J Intellect Disabil Res 49:560–566
21. Mori H (1997) The biological significance of neuropathological lesions in Alzheimer's disease. Neurobiol Aging 18:379–382
22. Schupf N, Patel B, Pang D et al (2007) Elevated plasma beta-amyloid peptide Abeta(42) levels, incident dementia, and mortality in Down syndrome. Arch Neurol 64:1007–1013
23. Temple V, Jozsvai E, Konstantareas MM, Hewitt TA (2001) Alzheimer dementia in Down's syndrome: the relevance of cognitive ability. J Intellect Disabil Res 45:47–55
24. Urv TK, Zigman WB, Silverman W (2010) Psychiatric symptoms in adults with Down syndrome and Alzheimer's disease. Am J Intellect Dev Disabil 115:265–276
25. Zigman WB, Lott IT (2007) Alzheimer's disease in Down syndrome: neurobiology and risk. Ment Retard Dev Disabil Res Rev 13: 237–246
26. Zigman WB, Schupf N, Sersen E, Silverman W (1996) Prevalence of dementia in adults with and without Down syndrome. Am J Ment Retard 100:403–412
27. Moechars D, Dewachter I, Lorent K et al (1999) Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. J Biol Chem 274:6483–6492
28. Wegiel J, Gong CX, Hwang YW (2011) The role of DYRK1A in neurodegenerative diseases. FEBS J 278:236–245
29. Park IH, Arora N, Huo H et al (2008) Disease-specific induced pluripotent stem cells. Cell 134:877–886
30. Takahashi K, Tanabe K, Ohnuki M et al (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861–872
31. Yu J, Vodyanik MA, Smuga-Otto K et al (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917–1920
32. Israel MA, Goldstein LS (2011) Capturing Alzheimer's disease genomes with induced pluripotent stem cells: prospects and challenges. Genome Med 3:49
33. Kondo T, Asai M, Tsukita K et al (2013) Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Aβ and differential drug responsiveness. Cell Stem Cell 12:487–496
34. Yagi T, Ito D, Okada Y et al (2011) Modeling familial Alzheimer's disease with induced pluripotent stem cells. Hum Mol Genet 20: 4530–4539
35. Yahata N, Asai M, Kitaoka S et al (2011) Anti-Abeta drug screening platform using human iPS cell-derived neurons for the treatment of Alzheimer's disease. PLoS One 6:e25788
36. Qiang L, Fujita R, Yamashita T et al (2011) Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. Cell 146:359–371
37. Martins-Taylor K, Nisler BS, Taapken SM et al (2011) Recurrent copy number variations in human induced pluripotent stem cells. Nat Biotechnol 29:488–491
38. Gore A, Li Z, Fung HL et al (2011) Somatic coding mutations in human induced pluripotent stem cells. Nature 471:63–67
39. McConnell MJ, Lindberg MR, Brennand KJ et al (2013) Mosaic copy number variation in human neurons. Science 342:632–637
40. Chung CY, Khurana V, Auluck PK et al (2013) Identification and rescue of alpha-synuclein toxicity in Parkinson patient-derived neurons. Science 342:983–987
41. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826
42. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232
43. Sanjana NE, Cong L, Zhou Y et al (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7:171–192
44. McMahon MA, Rahdar M, Porteus M (2012) Gene editing: not just for translation anymore. Nat Methods 9:28–31
45. Jiang J, Jing Y, Cost GJ et al (2013) Translating dosage compensation to trisomy 21. Nature 500:296–300
46. Fusaki N, Ban H, Nishiyama A et al (2009) Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci 85:348–362
47. Yu J, Hu K, Smuga-Otto K et al (2009) Human induced pluripotent stem cells free of vector and transgene sequences. Science 324:797–801
48. Papapetrou EP, Tomishima MJ, Chambers SM et al (2009) Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction

and differentiation. Proc Natl Acad Sci U S A 106:12759–12764

49. Seki T, Yuasa S, Oda M et al (2010) Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. Cell Stem Cell 7:11–14
50. Meyer JS, Shearer RL, Capowski EE et al (2009) Modeling early retinal development with human embryonic and induced pluripotent stem cells. Proc Natl Acad Sci U S A 106:16698–16703
51. Oshima K, Shin K, Diensthuber M et al (2010) Mechanosensitive hair cell-like cells from embryonic and induced pluripotent stem cells. Cell 141:704–716
52. Muguruma K, Nishiyama A, Ono Y et al (2010) Ontogeny-recapitulating generation and tissue integration of ES cell-derived Purkinje cells. Nat Neurosci 13:1171–1180
53. Zhang X, Huang CT, Chen J et al (2010) Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell 7:90–100
54. Pankratz MT, Li XJ, Lavaute TM et al (2007) Directed neural differentiation of hESCs via an obligated primitive anterior stage. Stem Cells 25:511–1520
55. Lavaute TM, Yoo YD, Pankratz MT et al (2009) Regulation of neural specification from human embryonic stem cells by BMP and FGF. Stem Cells 27:1741–1749
56. Yoo YD, Huang CT, Zhang X et al (2011) Fibroblast growth factor regulates human neuroectoderm specification through ERK1/2-PARP-1 pathway. Stem Cells 29:1975–1982
57. Chambers SM, Fasano CA, Papapetrou EP et al (2009) Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol 27:275–280
58. Johnson MA, Weick JP, Pearce RA, Zhang SC (2007) Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. J Neurosci 27:3069–3077
59. Shi Y, Kirwan P, Livesey FJ (2012) Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat Protoc 7:1836–1846
60. Siegenthaler JA, Pleasure SJ (2010) There's no place like home for a neural stem cell. Cell Stem Cell 7:141–143
61. Fasano CA, Chambers SM, Lee G et al (2010) Efficient derivation of functional floor plate tissue from human embryonic stem cells. Cell Stem Cell 6:336–347
62. Liu H, Zhang SC (2011) Specification of neuronal and glial subtypes from human pluripotent stem cells. Cell Mol Life Sci 68:3995–4008
63. Liu Y, Weick JP, Liu H et al (2013) Medial ganglionic eminence-like cells derived from human embryonic stem cells correct learning and memory deficits. Nat Biotechnol 31: 440–447
64. Liu Y, Liu H, Sauvey C et al (2013) Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. Nat Protoc 8:1670–1679
65. Wenk GL (2003) Neuropathologic changes in Alzheimer's disease. J Clin Psychiatry 64(Suppl 9):7–10
66. Cataldo AM, Peterhoff CM, Troncoso JC et al (2000) Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. Am J Pathol 157:277–286
67. Hooper C, Killick R, Lovestone S (2008) The GSK3 hypothesis of Alzheimer's disease. J Neurochem 104:1433–1439
68. Buytaert-Hoefen KA, Alvarez E, Freed CR (2004) Generation of tyrosine hydroxylase positive neurons from human embryonic stem cells after coculture with cellular substrates and exposure to GDNF. Stem Cells 22:669–674
69. Zeng X, Cai J, Chen J et al (2004) Dopaminergic differentiation of human embryonic stem cells. Stem Cells 22:925–940
70. Bellefroid EJ, Kobbe A, Gruss P et al (1998) Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. EMBO J 17:191–203
71. Morizane A, Doi D, Kikuchi T et al (2011) Small-molecule inhibitors of bone morphogenic protein and activin/nodal signals promote highly efficient neural induction from human pluripotent stem cells. J Neurosci Res 89:117–126
72. Nishitsuji K, Tomiyama T, Ishibashi K et al (2009) The E693Delta mutation in amyloid precursor protein increases intracellular accumulation of amyloid beta oligomers and causes endoplasmic reticulum stress-induced apoptosis in cultured cells. Am J Pathol 174:957–969
73. Briggs JA, Sun J, Shepherd J et al (2013) Integration-free induced pluripotent stem cells model genetic and neural developmental features of Down syndrome etiology. Stem Cells 31:467–478

74. Shi Y, Kirwan P, Smith J et al (2012) A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Sci Transl Med 4:124ra29
75. Weick JP, Held DL, Bonadurer GF 3rd et al (2013) Deficits in human trisomy 21 iPSCs and neurons. Proc Natl Acad Sci U S A 110:9962–9967
76. Becker LE, Mito T, Takashima S, Onodera K (1991) Growth and development of the brain in Down syndrome. Prog Clin Biol Res 373:133–152
77. Bhattacharyya A, McMillan E, Chen SI et al (2009) A critical period in cortical interneuron neurogenesis in Down syndrome revealed by human neural progenitor cells. Dev Neurosci 31:497–510
78. Esposito G, Imitola J, Lu J et al (2008) Genomic and functional profiling of human Down syndrome neural progenitors implicates S100B and Aquaporin 4 in cell injury. Hum Mol Genet 17:440–457
79. Golden JA, Hyman BT (1994) Development of the superior temporal neocortex is anomalous in trisomy 21. J Neuropathol Exp Neurol 53:513–520
80. Guidi S, Bonasoni P, Ceccarelli C et al (2008) Neurogenesis impairment and increased cell death reduce total neuron number in the hippocampal region of fetuses with Down syndrome. Brain Pathol 18:180–197
81. Guidi S, Ciani E, Bonasoni P et al (2011) Widespread proliferation impairment and hypocellularity in the cerebellum of fetuses with Down syndrome. Brain Pathol 21: 361–373
82. Ross MH, Galaburda AM, Kemper TL (1984) Down's syndrome: is there a decreased population of neurons? Neurology 34:909–916
83. Weitzdoerfer R, Dierssen M, Fountoulakis M, Lubec G (2001) Fetal life in Down syndrome starts with normal neuronal density but impaired dendritic spines and synaptosomal structure. J Neural Transm Suppl (60): 59–70
84. Wisniewski KE, Laure-Kamionowska M, Wisniewski HM (1984) Evidence of arrest of neurogenesis and synaptogenesis in brains of patients with Down's syndrome. N Engl J Med 311:1187–1188
85. Busciglio J, Yankner BA (1995) Apoptosis and increased generation of reactive oxygen species in Down's syndrome neurons in vitro. Nature 378:776–779
86. Becker LE (1991) Synaptic dysgenesis. Can J Neurol Sci 18:170–180
87. Takashima S, Becker LE, Armstrong DL, Chan F (1981) Abnormal neuronal development in the visual cortex of the human fetus and infant with Down's syndrome. A quantitative and qualitative Golgi study. Brain Res 225:1–21
88. Belichenko PV, Kleschevnikov AM, Salehi A et al (2007) Synaptic and cognitive abnormalities in mouse models of Down syndrome: exploring genotype-phenotype relationships. J Comp Neurol 504:329–345
89. Chakrabarti L, Galdzicki Z, Haydar TF (2007) Defects in embryonic neurogenesis and initial synapse formation in the forebrain of the Ts65Dn mouse model of Down syndrome. J Neurosci 27:11483–11495
90. Mao R, Zielke CL, Ronald ZH, Pevsner J (2003) Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. Genomics 81:457–467
91. Bahn S, Mimmack M, Ryan M et al (2002) Neuronal target genes of the neuron-restrictive silencer factor in neurospheres derived from fetuses with Down's syndrome: a gene expression study. Lancet 359:310–315
92. Cairney CJ, Sanguinetti G, Ranghini E et al (2009) A systems biology approach to Down syndrome: identification of Notch/Wnt dysregulation in a model of stem cells aging. Biochim Biophys Acta 1792:353–363
93. Antonarakis SE, Lyle R, Chrast R, Scott HS (2001) Differential gene expression studies to explore the molecular pathophysiology of Down syndrome. Brain Res Brain Res Rev 36:265–274
94. Lockstone HE, Harris LW, Swatton JE et al (2007) Gene expression profiling in the adult Down syndrome brain. Genomics 90:647–660
95. Swatton JE, Sellers LA, Faull RL et al (2004) Increased MAP kinase activity in Alzheimer's and Down syndrome but not in schizophrenia human brain. Eur J Neurosci 19: 2711–2719
96. Pradervand S, Paillusson A, Thomas J et al (2008) Affymetrix whole-transcript human gene 1.0 ST array is highly concordant with standard 3′ expression arrays. Biotechniques 44:759–762
97. Irizarry RA, Hobbs B, Collin F et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264
98. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. Bioinformatics 26:2363–2367

99. Keating DJ, Chen C, Pritchard MA (2006) Alzheimer's disease and endocytic dysfunction: clues from the Down syndrome-related proteins, DSCR1 and ITSN1. Ageing Res Rev 5:388–401
100. Mufson EJ, Conner JM, Kordower JH (1995) Nerve growth factor in Alzheimer's disease: defective retrograde transport to nucleus basalis. Neuroreport 6:1063–1066
101. Salehi A, Delcroix JD, Belichenko PV et al (2006) Increased App expression in a mouse model of Down's syndrome disrupts NGF transport and causes cholinergic neuron degeneration. Neuron 51:29–42
102. Counts SE, Mufson EJ (2005) The role of nerve growth factor receptors in cholinergic basal forebrain degeneration in prodromal Alzheimer disease. J Neuropathol Exp Neurol 64:263–272
103. Cooper JD, Salehi A, Delcroix JD et al (2001) Failed retrograde transport of NGF in a mouse model of Down's syndrome: reversal of cholinergic neurodegenerative phenotypes following NGF infusion. Proc Natl Acad Sci U S A 98:10439–10444
104. Wong SH, Xu Y, Zhang T, Hong W (1998) Syntaxin 7, a novel syntaxin member associated with the early endosomal compartment. J Biol Chem 273:375–380
105. Mullock BM, Smith CW, Ihrke G et al (2000) Syntaxin 7 is localized to late endosome compartments, associates with Vamp 8, and is required for late endosome-lysosome fusion. Mol Biol Cell 11:3137–3153
106. Chen YA, Scheller RH (2001) SNARE-mediated membrane fusion. Nat Rev Mol Cell Biol 2:98–106
107. Wang W, Bouhours M, Gracheva EO et al (2008) ITSN-1 controls vesicle recycling at the neuromuscular junction and functions in parallel with DAB-1. Traffic 9:742–754
108. Sakaba T, Kononenko NL, Bacetic J et al (2013) Fast neurotransmitter release regulated by the endocytic scaffold intersectin. Proc Natl Acad Sci U S A 110:8266–8271
109. Pechstein A, Shupliakov O, Haucke V (2010) Intersectin 1: a versatile actor in the synaptic vesicle cycle. Biochem Soc Trans 38: 181–186
110. Cruts M, van Duijn CM, Backhovens H et al (1998) Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease. Hum Mol Genet 7:43–51
111. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81:741–766
112. Yoo S, Yoo C (2011) A statistical model that calculates the life time risk of Alzheimer's disease using Bayesian Networks. Proceedings of 19th International Congress on Modeling and Simulation, Perth, Australia, 1049–1055, 2011 (http://www.mssanz.org.au/modsim2011/B4/yoo2.pdf)
113. Boutillier S, Lannes B, Buee L et al (2007) Sp3 and sp4 transcription factor levels are increased in brains of patients with Alzheimer's disease. Neurodegener Dis 4:413–423
114. Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol 11:39–46
115. Bush AI, Tanzi RE (2008) Therapeutics for Alzheimer's disease based on the metal hypothesis. Neurotherapeutics 5:421–432
116. Ferrer I, Barrachina M, Puig B et al (2005) Constitutive Dyrk1A is abnormally expressed in Alzheimer disease, Down syndrome, Pick disease, and related transgenic models. Neurobiol Dis 20:392–400
117. Becker-Barroso E (2013) Strengthening connections between Down syndrome and AD. Lancet Neurol 12:931
118. Israel MA, Yuan SH, Bardy C et al (2012) Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature 482(7384):216–20
119. Martoglio B, Golde TE (2003) Intramembrane-cleaving aspartic proteases and disease: presenilins, signal peptide peptidase and their homologs. Hum Mol Genet 12:R201–6

Chapter 16

Cortical Differentiation of Human Pluripotent Cells for In Vitro Modeling of Alzheimer's Disease

Nathalie G. Saurat, Frederick J. Livesey, and Steven Moore

Abstract

Stem cell models of Alzheimer's disease provide an opportunity to study the mechanisms underlying disease pathology at a resolution that is not possible in animal models. Furthermore, the ability to reprogram patient somatic cells to a pluripotent state ensures that the disease can be investigated in the correct genetic context. Here, we describe the directed differentiation of human pluripotent cells to cortical progenitors by recapitulating key developmental signaling events in vitro. Over a timeframe that mirrors human development, these progenitors give rise to functional lower and upper layer neurons. We also describe biochemical and imaging based methods to analyse key APP and Tau phenotypes in neurons generated from pluripotent stem cells from individuals with either monogenic familial Alzheimer's disease or Down's syndrome.

Key words Disease modeling, Alzheimer's disease, Stem cell, APP, Abeta peptides, Tau, Cortical differentiation, Neurodegeneration, Down's syndrome

1 Introduction

Human cellular models of Alzheimer's disease (AD) have the potential to complement existing animal models for carrying out mechanistic studies of AD initiation and progression. There is also considerable interest in using such models for high throughput and high content analyses, including chemical and genetic screens. Ideally, human stem cell models of AD should use the cell types affected by the disease and develop disease-relevant pathologies. For practical purposes, such models would undergo disease initiation in a reproducible manner and over a relatively short timescale.

Combining cellular reprogramming to generate patient-specific pluripotent stem cells (PSCs), with our understanding of neural development has enabled researchers to generate specific neuronal cell types in vitro, including excitatory and inhibitory cortical neurons [1–3]. When generated from PSCs derived from

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_16,

individuals affected by genetic forms of AD, including monogenic familial AD and trisomy 21/Down's syndrome, neurons develop many of the classical hallmarks of AD initiation. These include altered APP processing to generate changes in Abeta peptide production, Abeta aggregation, increases in Tau phosphorylation and the cellular localization of Tau [4–6].

In this chapter, we describe methods to generate human excitatory, glutamatergic cerebral cortex neurons from PSCs. This approach for making cerebral cortex neurons works equally well for embryonic stem cells and patient derived induced pluripotent stem cells [7]. A key aspect of this process is the need to allow time for the neurons generated to undergo functional maturation, as reflected in their firing properties: this is typically of the order of 2 months for their firing properties to resemble those of mature neurons. We also describe techniques to monitor the efficiency of neuronal production by immunofluorescence and confocal microscopy, and to analyse the development of specific AD phenotypes, with an emphasis on AD initiation, reflected in changes in APP processing.

With respect to human stem cells generated from individuals with genetic forms of AD and other dementias, a number of lines have been published and well characterized, including from individuals with Down's syndrome, familial AD (Psen1 mutations, APP duplication, APP mutations) and frontotemporal dementia, as detailed in Table 1. Recommended culture times of stem cells with specific AD-like pathologies are presented in Table 2.

Table 1
Stem cell models of Alzheimer's disease

Disease/ condition	Mutation/genetics	AD phenotypes	Reference
Down's syndrome	Trisomy 21	Increased Abeta peptide production Tau hyperphosphorylation Neuronal cell death	[5, 8]
Familial AD	APP duplication	Change in Abeta ratio Change in Tau phosphorylation	[4]
Familial AD	Psen1 and Psen2 mutations	Change in APP processing, altered Abeta ratios	[9]
Familial AD	APP mutation (deletion)	Intracellular Abeta oligomer formation, reduced extracellular Abeta peptides	[10]
Sporadic AD	Sporadic	Increased Abeta peptide production, changes in Tau phosphorylation	[4]
Sporadic AD	Sporadic	Intracellular Abeta oligomer formation	[10]

Table 2
Recommended culture times

Pathology	Detection method	Timescale
Aβ peptide release into culture medium	ELISA/MSD	From 35 days with >60 days preferred
Extracellular amyloid aggregates	Immunofluorescence	>D65
Cell biology (e.g. autophagy)	Immunofluorescence; western blot	>D60
Mislocalised/phosphorylated Tau	Immunofluorescence; western blot	>D90
Synapse formation	Immunofluorescence	>D50

2 Materials

Main media and solutions. For preparation, tissue culture grade media/solutions and distilled water are used.

1. STO medium for cultivation of mouse embryonic fibroblasts (MEFs): Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % (v/v) fetal bovine serum, 1 mM L-glutamine, 50 U/mL penicillin and 50 mg/mL streptomycin. Store at 4 °C and use within 2 weeks.
2. hPSC medium for cultivation of human pluripotent stem cells: DMEM/mixture F-12-GlutaMAX medium (DMEM/F-12 GlutaMAX) supplemented with 20 % (v/v) knockout serum replacement media (KnockOut SR, KSR), 10 ng/mL FGF2 (fibroblast growth factor basic), 100 μM non-essential amino acids, 100 μM 2-mercaptoethanol, 50 U/mL penicillin and 50 mg/mL streptomycin. Store at 4 °C and use within 2 weeks.
3. MEF conditioned medium: Collected after incubating MEFs in hPSC medium without FGF2 overnight. Pass through a 0.22 μM pore filter and store at −20 °C for up to 2 months. When required, thaw medium and add 10 ng/mL FGF2. Discard thawed aliquots after 5 days.
4. PBS. Tissue culture grade phosphate-buffered saline (PBS).
5. Dispase solution, Dispase II (neutral protease) resuspended in tissue culture grade phosphate-buffered saline (PBS) to 10 mg/mL at 37 °C before sterilisation through at 0.22 μm filter. Store at −20 °C for up to 2 months. Preventing or reducing clumping of human and mouse cells.
6. hPSC freezing medium: Defined fetal bovine serum, 10 % (v/v) dimethyl sulfoxide (DMSO) and 10 μM Y-27632 (Rho-kinase inhibitor with benefits in hPSC expansion). Make fresh and do not store.

7. N2 stem cell culture medium (N2 medium): DMEM/F-12 GlutaMAX with 1× N2 Supplement, 5 μg/mL insulin, 1 mM L-glutamine, 100 μM non-essential amino acids (NEAA), 100 μM 2-mercaptoethanol, 1 mM sodium pyruvate, 50 U/mL penicillin and 50 mg/mL streptomycin. Store at 4 °C and use within 3 weeks.
8. B27 medium: Neurobasal with 1× B27 Supplement, 200 mM L-glutamine, 50 U/mL penicillin and 50 mg/mL streptomycin. Store at 4 °C and use within 3 weeks. Basal media for neural cell culture.
9. N2B27 medium: A 1:1 mixture of N2 medium and B27 medium, store at 4 °C and use within 2 weeks.
10. Neural induction medium: N2B27 with 1 μM Dorsomorphin and 10 μM SB431542 (inhibitors of bone morphogenic protein (BMP) signaling coordinating development; stimulating proliferation and differentiation). Store at 4 °C and use within 1 week.
11. Accutase cell dissociation reagent. Detachment solution.
12. Laminin stock solution. Cell attachment, coating solution.
13. Neural freezing medium: N2B27 with 10 % (v/v) DMSO and 20 ng/mL FGF2. Make fresh and do not store.
14. Fixative: 4 % paraformaldehyde (PFA) in 1× PBS.
15. Tris buffered saline (TBS; 10×): 1.5 M NaCl, 0.1 M Tris–HCl, pH 7.4.
16. TBS-Tx: TBS (1×) with 0.3 % Triton X-100.
17. Blocking solution: 5 % normal donkey serum in TBS-Tx.

3 Methods

Ongoing cultures should be handled in a sterile laminar flow tissue culture hood. All media should be warmed to 37 °C in a water bath before use. All incubation steps should be carried out at 37 °C in 5 % CO_2 and all centrifugation steps should be performed at room temperature.

3.1 Plating Mouse Embryonic Fibroblasts

1. Add 1 mL of 0.1 % gelatin in distilled water to each well of two 6-well plates. Incubate at 37 °C for 10 min.
2. Partially thaw a vial of frozen irradiated mouse embryonic fibroblasts (MEFs) in a water bath at 37 °C and transfer to 10 mL of STO medium.
3. Centrifuge for 3 min at 180×*g* and resuspend the cells in 1.2 mL of STO medium.
4. Aspirate the gelatin solution from plates and replace with 2 mL of STO medium per well.

5. Add 100 μL of the MEF cell suspension to each well.
6. Gently rock the dishes forward, back and side-to-side to disperse the cells evenly.
7. Culture the MEFs for at least 6 h before the addition of stem cells.

3.2 Culturing Human Pluripotent Cells

1. Aspirate STO medium from MEF culture and wash each well in 2 mL of tissue culture grade phosphate buffered serum (PBS).
2. Aspirate PBS and replace it with 2 mL of hPSC medium without FGF2.
3. Partially thaw a vial of human pluripotent stem cells (PSC) in a water bath and transfer to 10 mL of hPSC medium (*see* **Note 1**).
4. Centrifuge for 2 min at 180 × *g*. Aspirate the supernatant.
5. Gently resuspend the colonies in 200 μL of hPSC medium before transferring them to one well of a 6 well plate.
6. Add 10 ng/mL FGF2 and 10 μM Y-27632 solution to the wells containing PSCs, collect MEF conditioned medium from the remaining wells. Incubate the cells overnight.
7. Maintain PSC colonies by daily replacement of media with hPSC containing 10 μM FGF2 until colonies become visible to the naked eye (~1 mm in diameter) (*see* **Note 2**).

3.3 Passaging PSCs

1. Prepare fresh mouse embryonic fibroblasts (MEFs) as described in Subheading 3.1.
2. Remove STO from MEFs, wash once with PBS and add 2 mL of hPSC medium containing 10 μM FGF2 to each well of the 6 well plate. Return MEFs in hPSC medium to incubator.
3. Add 200 μL of dispase solution to each well of PSCs and incubate for 20–40 min (*see* **Note 3**).
4. Remove STO from MEFs, wash once with PBS and add 2 mL of hPSC medium containing 10 μM FGF2 to each well of the 6 well plate. Return MEFs in hPSC medium to incubator.
5. Gently rock the plate to ensure detachment of PSC colonies from MEFs and differentiated cells.
6. Transfer colonies to 10 mL of PBS and centrifuge for 2 min at 180 × *g*. Aspirate solution and repeat PBS washes a further three times.
7. To passage cells and continue the culture, resuspend PSCs in 600 μL of hPSC medium and gently break up colonies to approximately 50–100 cells with P200 pipette (*see* **Note 4**).
8. Transfer 100 μL of PSC suspension to each well of a 6 well plate containing hPSC medium with 10 μM FGF2 (1:6 passage) and maintain as described in Subheading 3.2, **step** 7 (*see* **Note 5**).

9. Alternatively, colonies can be resuspended without being broken up in hPSC freezing medium after **step 8** and transferred to cryovials.
10. Cells for frozen storage should be immediately placed at −80 °C in a cell freezing box and transferred to liquid nitrogen after 24 h.

3.4 Neural Induction

1. Coat a 12 well plate with an appropriate substrate to allow feeder-free stem cell culture. Dependent on colony density, 5 wells of PSCs will typically yield enough cells for 3–6 wells of neural induction.
2. Follow **steps 3**, **5** and **6** of Subheading 3.3, pool together the 5 wells of PSCs to be used for induction and keep the cells to be passaged separate.
3. Passage 1 well of the 6 well plate as described in Subheading 3.3, **steps 7** and **8** to continue the PSC culture.
4. Add 500 μL of Accutase to the PSC pellet for induction, gently agitate the cells and incubate in a water bath for 5 min.
5. Gently use a P1000 pipette to create a single cell suspension of PSCs by slowly pipetting the solution 4–6 times.
6. Add 5 mL of hPSC medium to the suspension to inactivate the Accutase and centrifuge at $260 \times g$ for 3 min.
7. Resuspend the cells in 1 mL of MEF conditioned medium supplemented with 10 ng/mL FGF2 and 10 μM Y-27632.
8. Determine the number of cells using a standard haemocytometer or automated cell counter. Each well of neural induction should contain 7×10^5 viable cells per 1 mL of suspension.
9. Dilute the cell suspension to the required volume with MEF conditioned medium supplemented with 10 ng/mL FGF2 and 10 μM Y-27632.
10. Transfer 1 mL of PSC suspension per well. Incubate overnight.
11. The following day, aspirate medium from cells and replace with 1 mL of MEF conditioned medium supplemented with 10 ng/mL FGF2 and culture overnight.
12. Wash PSC wells with 1 mL of PBS and check that they are 100 % confluent (*see* **Note 6**).
13. Aspirate the PBS and replace with 1 mL of neural induction medium.
14. Monitor cells and replace the media everyday for the next 11 days to give a total of 12 days in induction medium (*see* **Note 7**).

3.5 Transferring Neuroepithelial Sheet to Laminin Substrate

1. Preparation of laminin coated well plates. Dilute laminin stock to a final concentration of 10 μg/mL in PBS and add 1 mL per well of a 6 well plate. Usually, one well of neural induction is transferred to one well of a 6 well plate. Incubate the plate for at least 4 h before use.
2. Aspirate the laminin solution and replace with 2 mL of neural induction medium, return plate to the incubator.
3. Add 100 μL of dispase to each well of neural induction and incubate until the neuroepithelial sheet detaches from the substrate, approximately 5 min.
4. Transfer the sheet as complete as possible to 10 mL of N2B27 medium and centrifuge at 180 × *g* for 2 min.
5. Perform two further 10 mL N2B27 washes to remove any residual dispase.
6. Add 200 μL of neural induction medium to each cell pellet and very gently break up the neuroepithelial sheet into aggregates of approximately 500 cells with a P100 pipette (*see* **Note 8**).
7. Transfer the cell aggregate suspension to neural induction medium in laminin coated plates and incubate overnight.
8. The next day, if the aggregates have attached to the laminin coating, change the medium to N2B27 with 20 ng/mL FGF2.
9. If the aggregates have not attached, centrifuge the cell suspension at 180 × *g* for 2 min. Resuspend in 200 μL of N2B27 and transfer into 2 mL of N2B27 with 20 ng/mL FGF2 on freshly coated laminin plates.
10. Change N2B27 with 20 ng/mL FGF2 medium every 2 days for a total of 4 days. Monitor the cells for the appearance of polarised neuroepithelial rosettes, which confirms a successful neural induction (*see* **Note 9**).
11. After 4 days of culture in N2B27 with 20 ng/mL FGF2, withdraw the FGF2 medium and maintain the neural progenitors by changing N2B27 medium every 2 days.

3.6 Passaging Neural Progenitor Cells

1. Between days 16 and 20 after neural induction, progenitors should be passaged in order to expand the culture and remove differentiated or non-neuronal cells. Typically neural progenitors are passaged 1:2 to expand the culture, depending on confluency.
2. Prepare laminin coated 6 well plates as described in Subheading 3.5, **step 1**. Aspirate the laminin solution and replace with 2 mL of N2B27 medium before returning to the incubator.
3. Add 200 μL of dispase to each well of neural progenitors and incubate for 5 min (*see* **Note 10**).

4. Transfer cells to 10 mL of tissue culture grade PBS and centrifuge at 180 × *g* for 2 min.
5. Repeat PBS washes a total of three times to remove any residual dispase.
6. Add 200 μL of N2B27 to the progenitors and gently break them into aggregates of approximately 500 cells using a P1000 pipette.
7. Transfer 100 μL of the cell suspension to N2B27 in laminin coated plates, gently disperse the aggregates by rocking the plate before returning to the incubator overnight.
8. The next day, aspirate the media and replace with N2B27. Continue to change N2B27 medium every 2 days.
9. Repeat **steps 3–8** a maximum of three times between days 16 and 26 after neural induction to enrich and expand the neural progenitor culture. Substantial neurogenesis should occur at the edge of rosettes during this period.
10. At day 26, progenitor aggregates should be passaged using Accutase to create a single cell suspension. Depending on confluency, passages can be performed 1:1 to maintain cell density or up to 1:4 to expand the culture.
11. Aspirate N2B27 medium and wash the cells once with 2 mL of tissue culture grade PBS.
12. Aspirate PBS and replace with 750 μL of Accutase per well. Incubate for 5 min.
13. Gently use a P1000 pipette to create a single cell suspension before transferring to 10 mL of N2B27.
14. Centrifuge at 180 × *g* for 3 min.
15. Resuspend the cells in 200 μL of N2B27 per well, transfer 200 μL of the suspension to 2 mL of N2B27 in a laminin coated 6 well plate.
16. Maintain the culture by replacing N2B27 medium every 2 days.
17. Repeat **steps 11–15** several times between days 26 and 38 after neural induction to expand the culture.

3.7 Freezing and Thawing Neural Progenitors

1. Neural progenitors can be frozen for long-term storage as single cells at any point after day 26 by resuspending the cell pellet generated in Subheading 3.6, **step 14** in neural freezing medium and transferring to cryovials.
2. Immediately place cryovials in a cell freezing box and store at −80 °C for 24 h before transferring to liquid nitrogen for long-term storage.
3. To resume a progenitor culture from a frozen stock, prepare laminin coated plates as described in Subheading 3.5, **step 1**.

4. Partially thaw a cryovial of neural progenitors in a water bath and transfer to 10 mL of N2B27 medium.
5. Centrifuge at 260 × *g* for 3 min.
6. Resuspend the cells in 2 mL of N2B27 with 20 ng/mL FGF2, transfer the suspension to one well of a laminin coated 6 well plate and return the plate to the incubator overnight.
7. The next day, remove the media from cells and replace with N2B27 without FGF2.
8. Continue with the protocol from Subheading 3.6, **step 16**.

3.8 Final Plating of PSC Derived Cortical Neurons

1. Allow thawed neurons to recover for at least 48 h before final plating.
2. To prepare cell culture dishes for final progenitor plating, pre-coat each 35 mm dish with 400 μL of poly-ornithine and incubate overnight. The following day remove the poly-ornithine and add 400 μL of 20 μg/mL laminin in PBS and incubate for at least 4 h before applying cells (*see* **Notes 11** and **12**).
3. Between day 33 and day 38 after neural induction progenitor cells can be passaged as previously described in Subheading 3.6, **steps 11–14** and resuspended in 1 mL of N2B27 (*see* **Note 13**).
4. Progenitors should be counted and plated at a density of 100,000 cells per cm^2 (i.e. 350,000 cells per 35 mm dish).
5. Replace N2B27 medium every 2 days to maintain the neural culture.
6. Add laminin into the medium every 10 days at a final concentration of 10 μg/mL.

3.9 Immunofluorescent Analysis of PSC Derived Cortical Neurons

1. Remove N2B27 media and wash cells once with PBS.
2. Add 4 % paraformaldehyde to each well and incubate at room temperature for 10 min (*see* **Note 14**).
3. Wash each well three times with 1× TBS.
4. To permeabilise the cells, apply TBS-Tx to each well and leave on a rocking platform for half an hour at room temperature (*see* **Note 15**).
5. Apply blocking solution to each well and incubate at room temperature for 1 h.
6. Dilute primary antibodies (Table 3) to the appropriate concentration in blocking solution (e.g. manufacturer's instructions).
7. Add antibody solution to each well and incubate on a rocking platform overnight at 4 °C.
8. Wash each well three times in 1× TBS, followed by an additional three times in TBS-Tx.

Table 3
Recommended primary antibodies for identifying Alzheimer's phenotypes

Antibody	Application
Aβ42	Extracellular aggregates of Abeta peptide
Total Tau	Both phosphorylated and non-phosphorylated forms of Tau
AT8	Tau phosphorylated at Ser202 and Thr205
Pax6	Primary progenitor cells
MAP2	Neurons/dendrites
TUJ1	Neurons
LC3B	Autophagosomes
Cleaved-caspase 3	Apoptotic cells
APP	Gene dosage validation
Munc13	Presynaptic terminals
Synaptophysin	Presynaptic terminals
PSD-95	Postsynaptic terminals
Homer 1	Postsynaptic terminals

9. Dilute secondary antibodies appropriately in blocking solution.
10. Apply secondary antibodies and incubate for 1 h at room temperature protected from light.
11. Wash six times in 1× TBS (*see* **Note 16**).
12. Mount and image on a confocal microscope.

3.10 Measuring Abeta Peptide Production by Stem Cell-Derived Neurons

1. Harvest medium from neuronal cultures every 48 h and replace with fresh N2B27 media.
2. Collect media from each well into a separate 1.5 mL tube.
3. Centrifuge at 1,200 × *g* for 3 min to pellet cellular debris.
4. Remove supernatant and store in protein low-bind 1.5 mL tubes at −80 °C until needed. Avoid freeze-thaw cycles.
5. Follow the protocol provided with your analytical method of choice (*see* **Note 17**).

4 Notes

1. PSCs should be partially thawed and immediately transferred to hPSC medium to minimize the toxic effects of DMSO in the freezing medium.

2. Colonies should be checked daily for non-specific differentiation. This is often seen as a loss of defined colony boundaries and a change in local cell density at these edges.
3. Dispase preferentially dissociates PSC colonies from MEFs and differentiated cells. However, all cells will detach from the substrate if digestion is prolonged or the plate is agitated.
4. Single cell colonies of PSCs are challenging to propagate, whereas larger colonies tend to merge and undergo some differentiation. Therefore, breaking up colonies to the correct size and their even distribution in the well is critical to maintain a PSC culture.
5. A 1:6 passage ratio is typically performed to maintain a PSC culture, however cells should be split according to their confluency. This is particularly important when thawing a frozen culture, where 1:2 or 1:3 passages are more appropriate.
6. Cells must be 100 % confluent before neural induction. Gaps in the monolayer will result in differentiation to non-cortical cell types, typically neural crest.
7. During the course of induction the PSCs should become tightly packed and reduce their nuclear volume as they are specified to become neural progenitors.
8. Neuroepithelial cells must be kept as aggregates to ensure that they retain their progenitor identity. Progenitors plated at a low local density will exit the cell cycle and become neurons.
9. FGF2 promotes the proliferation of neural progenitors, however cells should not be treated for longer than 4 days as this can posteriorise the tissue.
10. Dispase has a clear preference for dissociating neural progenitors rather than differentiated cells. This allows for purification of the culture over several passages.
11. Neurons for comparison should be plated in the same type of dishes and the identical well size, as both of these factors can heavily influence the amount of free Abeta peptide in the media.
12. Cortical cultures should be plated onto plastic dishes rather than glass cover slips in preparation for fixation and immunostaining, as this provides a better substrate for long term adhesion.
13. Cortical cultures should not be passaged beyond day 40 as neurons are particularly fragile and survival rates following passaging are low.
14. To improve the detection of surface antigens, cells can be fixed in pre-cooled 100 % methanol at –20 °C for 20 min prior to fixation in PFA. Triton can be omitted from later steps if using this method.

15. Put the rocker at very low speed for washes as neurons are easily dislodged from culture plates following extended periods in culture.
16. Add DAPI fluorescent stain (binding to A-T rich DNA regions) to one of the final wash steps if it is not present in the mounting media of choice (fluorescent labeling of nuclei).
17. A number of platforms are available to determine the concentration of peptides/proteins of interest within the cell culture medium. Due to their sensitivity, dynamic range and convenience our preferred system is the multiplex MesoScale Discovery (MSD) assays.

Acknowledgements

Work in the lab of FJL is supported by the Wellcome Trust, Alzheimer's Research UK, the Medical Research Council and the EU Innovative Medicines Initiative StemBANCC. NS is supported by the Woolf Fisher Trust. FJL is a Wellcome Trust Senior Investigator.

References

1. Shi Y, Kirwan P, Smith J et al (2012) Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat Neurosci 15:477–486, S1
2. Maroof AM, Keros S, Tyson JA et al (2013) Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. Cell Stem Cell 12:559–572
3. Nicholas CR, Chen J, Tang Y et al (2013) Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development. Cell Stem Cell 12:573–586
4. Israel MA, Yuan SH, Bardy C et al (2012) Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature 482:216–220
5. Shi Y, Kirwan P, Smith J et al (2013) A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Sci Transl Med 4:124ra29
6. Qiang L, Fujita R, Yamashita T et al (2011) Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. Cell 146:359–371
7. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676
8. Park IH, Arora N, Huo H et al (2008) Disease-specific induced pluripotent stem cells. Cell 134:877–886
9. Yagi T, Ito D, Okada Y, Akamatsu W et al (2011) Modeling familial Alzheimer's disease with induced pluripotent stem cells. Hum Mol Genet 20:4530–4539
10. Kondo T, Asai M, Tsukita K et al (2013) Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. Cell Stem Cell 12: 487–496

Part IV

Experimental Systems Biology: Next Generation Molecular and High-Throughput Methods for the Study of Disease Susceptibility and Networks Dynamics Interplay in Complex Diseases

Chapter 17

Next Generation Sequencing in Alzheimer's Disease

Lars Bertram

Abstract

For the first time in the history of human genetics research, it is now both technically feasible and economically affordable to screen individual genomes for novel disease-causing mutations at base-pair resolution using "next-generation sequencing" (NGS). One popular aim in many of today's NGS studies is genome resequencing (in part or whole) to identify DNA variants potentially accounting for the "missing heritability" problem observed in many genetically complex traits. Thus far, only relatively few projects have applied these powerful new technologies to search for novel Alzheimer's disease (AD) related sequence variants. In this review, I summarize the findings from the first NGS-based resequencing studies in AD and discuss their potential implications and limitations. Notable recent discoveries using NGS include the identification of rare susceptibility modifying alleles in *APP*, *TREM2*, and *PLD3*. Several other large-scale NGS projects are currently underway so that additional discoveries can be expected over the coming years.

Key words Alzheimer's disease, Next-generation sequencing, Rare variant association, Genome-wide association study, GWAS

1 Introduction

Similar to many other adult-onset human disorders, Alzheimer's disease (AD)—a slowly progressive neurodegenerative disease of the brain eventually resulting in cognitive impairment and dementia—represents a "genetically complex trait". This term alludes to the fact that a person's liability for AD is the result of a combination of heritable (e.g. genetic) and non-heritable (e.g. environmental) factors. Twin studies suggest that the contribution of the former probably outweighs the latter for the vast majority of AD cases [1]. In some rare familial forms of AD, genetics plays the predominant role through the effect of extremely infrequent amino acid substituting mutations in genes such as *APP* (β-amyloid precursor protein [APP]), *PSEN1* (presenilin 1 [PS1]), and *PSEN2* (presenilin 2 [PS2]). Despite their rarity, mutations in these genes have been instrumental in clarifying the molecular mechanisms underlying AD pathogenesis where the aberrant production of the β-amyloid (Aβ)

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_17,

peptide represents a crucial event [2]. Intracellularly, Aβ is cleaved from APP by the sequential action of two enzymatic activities, i.e. β-secretase (encoded by the *BACE1* [β-amyloid cleavage enzyme 1] gene) and γ-secretase (whose catalytic site is made up of PS1 and PS2). By identifying AD-causing mutations in both the precursor of Aβ (i.e. *APP*) and the enzymes involved in its production (i.e. *PSEN1* and *PSEN2*), genetics has supported the "amyloid hypothesis" of AD which posits that dysregulated Aβ triggers the development and perhaps progression of the disease [2]. For recent reviews on AD genetics *see* refs. 3–5.

Mutations in *APP*, *PSEN1*, and *PSEN2* only account for a small fraction (≪5 %) of all AD cases, which I will refer to as "Mendelian AD" due to the almost complete penetrance and mostly autosomal-dominant mode of transmission of implicated DNA sequence changes. The vast majority of AD cases, however, is actually of a "non-Mendelian" nature. The predisposition for this type of AD is the result of a combined action of dozens, if not hundreds or thousands, of common DNA sequence variants (i.e. polymorphisms) of small effect (i.e. odds ratios [ORs] typically ≪2) and, hence, incomplete penetrance. Over 30 years of research have investigated thousands of DNA polymorphisms in hundreds of putative AD candidate genes to find genetic risk factors underlying non-Mendelian AD [6]. With one notable exception, i.e. the apolipoprotein E gene (*APOE*) on chromosome 19 [7], these studies have not resulted in establishing firm disease associations until the advent of high-throughput microarray genotyping technology allowed the completion of genome-wide association studies (GWAS) in ~2008. These GWAS have finally resulted in a number of highly convincing AD association findings (*see* the 'AlzGene' database for a summary of these studies: http://www.alzgene.org [8]). Collectively, however, GWAS associations currently explain no more than half of the disease heritability, i.e. the proportion of phenotypic variance that can be explained by genetic or epigenetic factors. Interestingly, this situation—i.e. that the combined contribution to liability for disease is only partially explained by findings to emerge from even the most powerful GWAS—is observed for many more genetically complex diseases (and non-disease traits). Several potential hideouts for this "missing heritability" have been proposed [9], including the possibility that it may altogether represent a "phantom" phenomenon [10].

The recent development of extremely powerful, massively parallel DNA sequencing technologies now allows to systematically screen individual genomes for DNA sequence variation at base-pair resolution, enabling researchers to address the "missing heritability" question (and many other questions) empirically (Fig. 1; Table 1). Owing to their potential to revolutionize human genetics research, the term "next-generation sequencing" (NGS) has been coined for these methods. In this review, I summarize the findings of the first studies specifically applying NGS to the field of AD

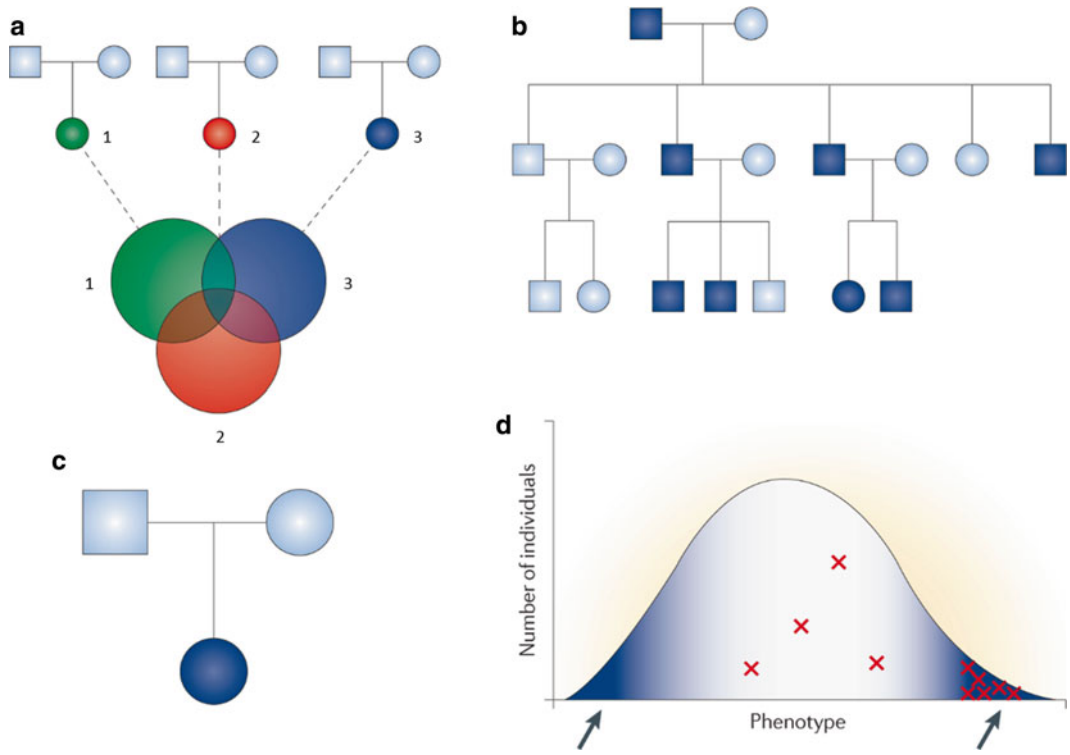


Fig. 1 Strategies for finding disease-causing rare variants using exome sequencing. Four main strategies are illustrated. (**a**) Sequencing and filtering across multiple unrelated, affected individuals (indicated by the *three colored and numbered circles*). This approach is used to identify novel variants in the same gene (or genes), as indicated by the *shaded region* that is shared by the three individuals in this example. (**b**) Sequencing and filtering among multiple affected individuals from within a pedigree (*shaded circles and squares*) to identify a gene (or genes) with a novel variant in a shared region of the genome. (**c**) Sequencing parent–child trios for identifying de novo mutations. (**d**) Sampling and comparing the extremes of the distribution (*arrows*) for a quantitative phenotype. As shown in panel **d**, individuals with rare variants in the same gene (*red crosses*) are concentrated in one extreme of the distribution. Figure and legend reprinted with permission from Macmillan Publishers Ltd: Nature Reviews Genetics [18], copyright (2011)

Table 1
Variant detection in complex traits classified by allele frequency

Variant class	Minor allele frequency	Implication for analysis	Example in AD
Very common	Between 5 and 50 %	GWAS	*APOE*, *CLU*, *CR1*, *PICALM*
Less common	Between 1 and 5 %	GWAS	n.a. (not available)
Rare (but not private)	Less than 1 % but still polymorphic in one or more major human population	GWAS; co-segregation in families	*APP* (Ala673Thr), *PLD3*, *TREM2*
Private	Restricted to probands and immediate relatives	Co-segregation in families; linkage	AD-causing mutations in *APP*, *PSEN1*, *PSEN2*

GWAS genome-wide association study. See text for explanation of gene symbols. Adapted for this review with permission from Macmillan Publishers Ltd: Nature Reviews Genetics [19], copyright (2010)

genetics research. The focus lies exclusively on studies using NGS for genome resequencing (whole or in part) and does not cover projects utilizing other NGS applications, such as transcriptome or methylome sequencing.

2 Next-Generation Sequencing to Identify Novel Disease Genes

There already exist a large number of excellent reviews on the technical details of the currently available NGS platforms [11, 12], as well as on relevant theoretical (e.g. [13–16]), practical (e.g. [17–20]), and analytical (e.g. [21–23]) considerations when embarking on an NGS-based resequencing project (*see* also Fig. 1). Accordingly, these topics will be skipped here except for a few general remarks. For instance, one important issue to keep in mind is that the *generation* of NGS-based large-scale sequencing data has become relatively straightforward thanks to the availability of highly optimized operating procedures developed by the manufacturers of today's NGS instruments. On the other hand, efficient and appropriate *management* and *interpretation* of the resulting sequence data is not nearly as straightforward for most laboratories outside highly specialized genome centers. To a large part this is due to the sheer amount of information created. For instance, a single human genome consists of ~3.2 billion base-pairs (Gbp), each of which needs to be covered at least 30- to 35-fold in order to confidently differentiate between wild-type allele and mutation [20], yielding a minimum of ~100 Gbp per DNA sample per experiment.

Another, possibly even more challenging aspect is that potentially "functional" DNA sequence variants occur at much higher frequency in the general population than originally anticipated (even though overall they may still be classified as "rare", i.e. displaying a minor allele frequency [MAF] $\ll$1 %) [9, 24]. The important conclusion to draw is that not every amino-acid changing nucleotide substitution found in an affected individual or observed to cosegregate with disease status in a given family also automatically represents the underlying disease-causing variant. This situation has been referred to as the "narrative potential" of individual genomes [15], meaning that assigning a disease-related narrative to potential mutations in anyone's genome sequence is relatively easy (simply owing to the high frequency of these sequence changes) but statistically often poorly justified. Thus, in order to avoid publishing "genomic fairytales", researchers need to go to great lengths to ensure that a presumed connection between a pinpointed DNA sequence variant and onset of disease is in fact genuine. For GWAS findings this typically entails to provide consistent association evidence in several independent data sets which, when analyzed alone or in combination, pass a certain threshold of statistical support (typically a *P*-value below 5×10^{-8}). Neither of these requirements can be applied to NGS-based genetics studies in a straightforward fashion.

First, the variants identified are often exceedingly rare (if not altogether "private", i.e. restricted to one founder and its family members), so that a sufficient number of carriers—affected by disease or not—may be difficult to come by for any individual laboratory (*see* also Table 1). Second, there currently exist no firm guidelines on the statistical interpretation of "rare variant" association findings. As a matter of fact, there currently exist no firm rules on how to establish evidence in favor of a genetic association between disease status and specific rare variants. Possible analysis strategies include variant-specific tests (similar to those used for common variants in GWAS), gene-specific tests (e.g. by pooling discovered variants within loci in affected vs. unaffected individuals) with and without pre-defined allele frequency thresholds, and network/gene-set analyses (e.g. by pooling association evidence across genes by their presumed or proven functional connection).

As will become clear in the following paragraphs, many of the issues briefly touched upon above have already been encountered in the few available NGS-studies conducted in the field of AD genetics. Thus, some of the reported gene findings outlined below can be assigned greater credibility than others. In a sense, this situation is not unlike that encountered during the pre-GWAS candidate gene era of AD genetics research. The field will likely remain in this state until firm criteria on the analysis and interpretation of NGS resequencing data have been established. I close this introduction by quoting from a recently published, highly interesting essay on the theoretical framework of rare-variant associations in human diseases [13], which concluded that "for very late-onset diseases like Alzheimer's […] common variant association studies would likely be the better strategy" to identify the most important disease genes. This is due to the fact that rare variant associations will be a comparatively infrequent occurrence in these diseases for a number of reasons discussed in [13]. Time will tell, whether or not these considerations and conclusions will prove to be correct.

3 Next-Generation Sequencing in Alzheimer's Disease Research to Date

Owing to the fact that NGS technologies have only become available (and affordable) outside highly specialized genome centers from ~2010, the literature reporting first results of their application to AD is still rather limited. For the purpose of this review, NCBI's "PubMed" database (http://www.ncbi.nlm.nih.gov/pubmed) was queried using the keywords "[alzheimer* AND ((next generation sequencing OR NGS OR deep sequencing) OR (exome sequencing OR WES) OR (whole genome sequencing OR WGS))]" which yielded a total 103 publications on April 15, 2014. Of these, only 15 reported data relevant to this review, i.e. bona fide NGS-based resequencing in at least one cohort of AD patients. Together with a few additional relevant publications identified via

other sources, these studies represent the "core" findings discussed in more detail below (Table 2). Notwithstanding the relative paucity of NGS studies in AD, the already available papers have collectively applied the full range of different NGS strategies and thus provide a timely starting point for a first critical assessment. Note, that this field is advancing very rapidly, so that readers are advised to consult the 'AlzGene; database' (http://www.alzgene.org [8]) or 'AD and FTD Mutation Database' (http://www.molgen.ua.ac.be/ADMutations/ [25]) for updated summaries of relevant studies published after the day of writing.

3.1 NGS for Resequencing of Alzheimer's Disease Candidate Genes

In an attempt to resolve the contribution of putative functional DNA sequence variants in genes *known* to be associated with disease risk, most early NGS studies in AD either performed deep resequencing of the established early-onset Mendelian AD (and other forms of dementia, particularly FTLD) genes, i.e. *APP*, *PSEN1*, *PSEN2*, *MAPT* and *GRN*, or of loci recently implicated by GWAS, in particular *CLU*, *CR1*, and *PICALM* (Table 2). In addition, there are also a few publications reporting deep resequencing results of loci that had emerged during the "candidate gene" era of AD genetics, i.e. *ABCA1* (encoding ATP-binding cassette, sub-family A [ABC1], member 1) and *NCSTN* (encoding the γ-secretase component nicastrin; Table 2). As will be discussed below, the knowledge gained from these focused, early-adopter NGS studies remains limited. This is in contrast to more systematic projects performing whole exome (WES; Subheading 3.2) or whole genome sequencing (WGS; Subheading 3.3).

3.1.1 NGS for Resequencing Mendelian AD Genes

Several of the first projects have applied NGS to resequence known Mendelian genes in late-onset, non-Mendelian AD cases [26–28]. All of these studies identified novel "rare variants" across the investigated loci leading some authors even to conclude that "rare variants in these genes could explain an important proportion of genetic heritability of AD" [26]. It remains debatable whether this conclusion is indeed justified. First, many of the variants identified within target genes were deemed as "non-pathogenic", i.e. there is currently no compelling evidence from in silico or in vitro assessments to support an impact on pathogenicity (see also 'AD & FTD Mutation Database' for more details). Second and more importantly, sequencing technologies and bioinformatic variant-calling algorithms (and correlated measures such as base-pair coverage, which determine the false-positive and false-negative rate) in AD cases often differed from those applied to controls, potentially biasing the discovery of "rare variants" towards AD populations. This situation arises when AD cases are resequenced in-house and then compared to control genomes derived from public databases, such as the 1000 Genomes Project website (http://www.1000genomes.org) or the ExomeVariant Server (http://evs.gs.washington.edu/EVS/). Possibly the most interesting result

Table 2
Findings from NGS-based resequencing studies in AD

Ref.	Author, year	NGS approach	Study design	Main finding	Overall credibility
[26]	Cruchaga, 2012	Candidate (APP, PSEN1/2, MAPT, GRN)	Family-based (LOAD)	RVs in APP, PSEN1, GRN, MAPT to confer risk	Unclear
[27]	Jin, 2012	Candidate (APP, PSEN1/2, MAPT, GRN)	Family-based (EOAD, LOAD)	RVs in PSEN1, GRN, MAPT to confer risk	Unclear
[28]	Benitez, 2013	Candidate (APP, PSEN1/2, APOE, MAPT, GRN)	Extreme CSF levels	PSEN1-E318G linked to CSF-tau	Low
[29]	Lord, 2012	Candidate (CLU, CR1, PICALM)	AD cases only	Technical issues	–
[36]	Lupton, 2011	Candidate (NCSTN)	Case–control (LOAD)	RVs to confer risk	Low
[37]	Lupton, 2014	Candidate (ABCA1)	Case–control (LOAD)	RVs to confer protection	Low
[39]	Guerreiro, 2012	WES	Family (1 exome)	NOTCH3-R1231C in index patient	Low
[40]	Pottier, 2012	WES	Families (EOAD)	RVs to confer risk	Low/medium
[43]	Cruchaga, 2014	WES	Families and case–control data sets	PLD3 (several RVs to confer risk)	Medium
[50]	Jonsson, 2012	WGS (for GWAS)	Case–control	APP-Ala673Thr to confer protection	High
[51]	Jonsson, 2013	WGS (for GWAS)	Case–control	TREM2-Arg47His to confer risk	High
[58]	Guerreiro, 2013	WGS, WES	Case–control	TREM2-Arg47His (+ other RVs) to confer risk	High

Simplified summary of lead findings of NGS studies discussed in main text. "Overall credibility/relevance" as (subjectively) judged by the author of this review based on the data presented in original publications. See main text for more details on studies and findings. *RVs* rare variants, *LOAD* late-onset AD [also referred to as "non-Mendelian AD" in main text], *EOAD* early-onset AD [also referred to as "Mendelian AD" in main text], *GWAS* genome-wide association study. See text for explanation of gene symbols

from these early NGS studies is the comparatively high proportion of pathogenic *MAPT* and *GRN* variants in individuals clinically diagnosed as "AD" suggesting a larger than anticipated misdiagnosis rate [26] or—less likely—pleiotropy at these loci.

3.1.2 NGS for Resequencing AD Susceptibility Genes

At the day of writing, there is only one published *bona fide* NGS study investigating the established AD GWAS loci [29]. This project performed pooled NGS of sequence capture products targeting *CLU*, *CR1*, and *PICALM* in 96 AD patients and compared the

frequency of identified variants to those listed in public databases. Owing mostly to technical difficulties encountered during the course of the project, no firm conclusion regarding the prevalence of putative functional variants (common or rare) in these loci could be reached. This is in line with a number of conventional, Sanger-based resequencing projects of either all or a subset of the same genes [30–35]. These studies were also unable to conclusively pinpoint "functional" variants within previously reported GWAS loci. The same is true for NGS-based resequencing studies of non-GWAS AD candidate genes, i.e. *NCSTN* [36] and *ABCA1* [37], which have produced little more than anecdotal results (Table 2).

Collectively, this situation is reminiscent of the pre-GWAS "candidate gene era" of AD genetics which has produced a flurry of proclaimed results essentially none of which survived the test of time [38]. The current lack of convincing NGS results in the established AD loci may be due to the same combination of factors that had already bedeviled genetic association studies back then, including small sample size, smaller-than-anticipated effect sizes, failure to provide replication evidence from independent cohorts, and over-interpretation of borderline statistical evidence.

3.2 Whole-Exome Sequencing Studies in Alzheimer's Disease

With respect to "outcome", NGS studies following a more systematic approach—e.g. those focusing on the exome (WES) or whole genome (WGS)—have thus far produced more convincing results than those investigating only a few loci (*see* Subheading 3.1). While a number of publications report to have performed WES on AD cases and sometimes control populations, only the three studies discussed below applied this technology for *bona fide* exome-wide discovery of novel disease genes.

The first study by Guerreiro et al. [39], used WES in a single clinically diagnosed AD patient from Turkey originating from a consanguineous family with a complex medical history, including the presence of both neurological and immunological disorders. Bioinformatic analysis and filtering of the WES data identified 178 candidate missense variants, one of which (p.R1231C in *NOTCH3*; Table 2) was reported as the molecular culprit for the clinical AD-like picture seen in this patient. This mutation, along with more than 130 other mutations in this gene, has previously been reported to cause another neurological disorder: CADASIL (cerebral arteriopathy autosomal dominant with subcortical infarcts and leukoencephalopathy) [39], an inherited cerebrovascular disease with a number of different neurological symptoms. Typically, patients with CADASIL show pronounced white matter abnormalities resulting in characteristic MRI findings. These MRI findings, however, were absent from the patient subject in this study. Unfortunately, DNA from other affected relatives was not available to assess segregation of the variant with disease status. The only other carrier of the p.R1231C mutation was a cognitively normal

son of the index patient, approximately 20 years younger than the dementia onset age in his father. The putative AD-causing *NOTCH3* mutation was absent in more than 300 AD cases and controls from Turkey and elsewhere. While this study demonstrated the power of WES to efficiently generate a near complete status of the "mutational landscape" in a single patient, it did lack some essential supporting evidence to conclusively imply the p.R1231C *NOTCH3* variant as the cause of the complex clinical phenotype observed in this particular patient: e.g. lack of within family disease segregation, independent replication, exclusion of causality of other missense variants present in WES data. Further, owing to the unclear and complex clinical picture it is too early to add *NOTCH3* to the list of AD genes. An alternative interpretation of the available data could simply be—as the authors themselves concede—that p.R1231C is neither pathogenic for CADASIL nor AD.

The second publication [40] highlighted in this section used WES in 14 index cases of autosomal dominant early-onset AD families without mutations in any of the established AD genes *APP*, *PSEN1*, and *PSEN2*. Bioinformatic filtering of the variant calls was based on putative functional impact (i.e. only non-synonymous, splice site, and frameshift indels were retained), "novelty" (i.e. only variants not listed in public databases were retained), and recurrence (i.e. only genes harboring variants resulting from the previous filtering steps in >1 family were retained). This strategy led to a number of potential candidate genes, the most compelling of which was sortilin-related receptor LR11/SorLA (*SORL1*; Table 2) [40]. In its function as a sorting receptor and central regulator of the trafficking and processing of APP, SorLA has represented an AD candidate gene for nearly a decade [41]. Early genetic association analyses suggested *SORL1* to be a late onset Alzheimer's disease (LOAD) risk gene, but were met with a flurry of conflicting data [8]. A recent GWAS meta-analysis revived the topic by reporting common variants in *SORL1* to show genome-wide significant association with AD risk [42], although previous analyses of largely overlapping datasets did not yield such a finding. In the WES study discussed here, *SORL1* was found to harbor previously unknown and putatively functional mutations in 5 of the 14 index patients. Resequencing of *SORL1* in a separate collection of 15 index patients from independent EOFAD families identified two additional novel variants also presumed to be disease-causing [40]. Owing to the lack of available biospecimen, co-segregation with disease could only be demonstrated for one of the seven *SORL1* variants. If indeed genuine, these findings would imply that *SORL1* would be on a par with *PSEN2* as the third most frequently mutated gene responsible for autosomal dominant forms of AD. Furthermore, it would represent the first gene in AD genetics to harbor both rare and disease-causing as well as common susceptibility variants. The lack of functional data clearly and directly supporting an impact of

the identified mutation on protein expression or function, the absence of conclusive segregation evidence in all but one family, and the current nonexistence of independent replication results suggests that more time and scientific evidence is needed before *SORL1* can be counted as an established causal AD gene.

The third and most recent paper to employ WES in AD [43] sequenced a total of 29 individuals from 14 AD families with four or more affected individuals. Filtering based on minor allele frequency (<0.5 %), segregation of candidate variants with disease status within families, and occurrence of the same variant in >1 family, revealed a missense change in the gene encoding phospholipase D3 (*PLD3*; Table 2) on chromosome 19q13.2 [43]. The protein represents a hitherto poorly characterized member of the PLD superfamily of phospholipases. Other members of this superfamily, i.e. PLD1 and PLD2, have been reported to be involved in APP trafficking and synaptic dysfunction [48, 49], making PLD3 a reasonable candidate as well. The PLD3 variant identified by Cruchaga and colleagues, rs145999145, elicits a valine to methionine change at residue 232 (Val232Met) and is present in up to 0.5 % of non-AD individuals of European descent. In AD cases, the frequency of the Met-allele is between 0.6 and 1.3 %, thus approximately doubling the risk for AD in carriers vs. non-carriers [43]. The original family-based finding was subsequently extended to an independent series of more than 11,000 AD cases and controls where it was also found to be associated with risk for AD (OR ~2) and a significant reduction in onset age (between 3 and 8 years). Resequencing of the *PLD3* coding region in ~4,300 AD cases and controls of European descent revealed potential additional rare PLD3 variants which, by means of aggregate analysis ("burden test"), occurred significantly more often in cases than healthy controls. This finding may indicate the presence of other disease associated functional variants at this locus beyond Val232Met. Finally, additional analyses revealed an excess of rare *PLD3* coding variants in AD cases vs. controls in a small collection of individuals of African descent, further supporting the notion that *PLD3* represents a genuine AD locus. The genetic results were accompanied by a range of supporting functional data from human brain samples and transgenic mouse neuroblastoma cell lines suggesting that *PLD3* may exert its pathogenic effects by directly affecting APP processing and Aβ42 and Aβ40 production. Based on these data, the most likely effect of Val232Met (and possibly other variants in *PLD3*) is a loss of function, e.g. by reduced gene expression, and as a result an increase in Aβ42 and Aβ40 production. Since the original publication [43], a number of studies [44–47] were published in an attempt to independently validate the reported findings. However, despite having excellent power to detect the previously reported effect size, none of these papers was able to replicate the

association between AD risk and Val232Ala (or other polymorphisms in PLD3) shedding serious doubt on the notion that PLD3 is in fact a genuine AD gene [44–47].

3.3 Whole-Genome Sequencing Studies in Alzheimer's Disease

At the time of writing, there are no published studies directly applying WGS for the discovery of novel AD genes, but several such studies are currently underway. For instance, WGS data generated as part of the "AD sequencing project" (ADSP) on 584 individuals from 111 multiplex AD families were recently released to the community and are in the process of being analyzed for the presence of AD-associated rare variants in these families. First results from these efforts are expected to be reported early in 2015 (*see* ADSP website for more details: https://www.niagads.org/adsp/). A more extensive WGS study is being conducted as part of the "Alzheimer's Genome Project" (AGP, P.I. Rudolph E. Tanzi at Harvard Medical School), which has recently completed the generation of sequencing data in over 1,500 individuals from 437 multiplex AD families. Bioinformatic workup of the sequence reads is currently underway, and first results are expected to be released in 2015 (for more details *see*: http://www.curealz.org/projects/whole-genome-sequencing).

In addition to *directly* sequencing specific subjects and families of interest for disease gene discovery, WGS can also be used *indirectly*. This strategy was followed by researchers from deCODE genetics who utilized WGS data on a collection of ~2,000 Icelandic individuals in two related projects investigating the role of rare functional variants on AD risk [50, 51]. In both studies, WGS data were used to impute allele status at sites harboring putative functional variants (n ~192,000, including nonsynonymous, frameshift, splice site, and stop gain-loss variants) onto microarray based genome-wide genotype data in more than 3,000 AD patients. These were compared to genotype data imputed from the same WGS panels in ~80,000–110,000 Icelandic healthy control individuals. In essence, the analytic strategy applied in these projects comes down to a GWAS on imputed genotype data specifically enriched for putative functional variants originating from WGS data in the population of interest. Apart from *APOE*, the analyses of the resulting data pinpointed two previously known rare missense variants showing evidence for genome-wide significant association with AD risk: (1) the Ala673Thr substitution in *APP* (decreasing the risk for developing AD in carriers by approximately fivefold) [50], and (2) the Arg47His substitution in *TREM2* on chromosome 6p21.1 (increasing the risk for developing AD by approximately threefold) [51]. Unlike the WES-based results reported in the previous section, both findings have been confirmed by independent laboratories following the original publications.

The first AD association to be revealed by deCODE's "WGS-enriched GWAS strategy" highlighted a previously known SNP

(rs63750847) in *APP* eliciting a threonine to alanine substitution at residue 673 (Ala673Thr) [50]. Interestingly, the minor A-allele was found to be *under* represented in the Icelandic AD cases under study, suggesting a protective effect when compared to the reference G-allele (translating into an OR reduction of about fivefold). In addition to these AD specific findings, the authors also investigated the decline in cognitive function over time in non-AD individuals between 80 and 100 years of age [50]. In line with the observed protection from AD, carriers of the A-allele at rs63750847 performed significantly better in cognitive testing than non-carriers. First in vitro functional data reported in the same study revealed that the protective A-allele (coding for amino acid threonine), which is located within APP's β-cleavage site, significantly reduces the production of sAPPβ and $A\beta_{40/42}$ by approximately 50 % relative to the wild-type G-allele (coding for alanine) [50]. Overall, these data suggest that Ala673Thr exerts a direct effect on BACE1 cleavage of APP. Independent follow-up studies in various populations either confirmed the protective effect of Ala673Thr (e.g. in a Finnish population [52]), or were unable to identify any carriers of the minor threonine allele (e.g. in South-East Asian [53, 54] or North American [55] samples). Of note, A673 was also found in one familial AD case in the NGS-based assessment of known dementia genes by Cruchaga et al. (*see* above and [26]), and in a patient suffering from ischemic cerebrovascular disease [56], a finding that is reminiscent of the protective *APOE* ε2-allele which has also been found—albeit at reduced frequency—in patients suffering from AD and other neurodegenerative diseases [57].

The second AD association identified by the deCODE group was with SNP rs75932628 leading to a an arginine to histidine substitution at position 47 (Arg47His, or R47H) in the gene encoding triggering receptor expressed on myeloid cells 2 (*TREM2*) [51]. Combining association results from a number of different datasets, the minor T-allele at this site was found to significantly increase the risk for AD by approximately threefold. Opposite to the effects observed for the protective variant in *APP* (see above), the AD-associated allele was associated with poorer cognitive function in cognitively healthy individuals aged between 80 and 100 years. Independent confirmation of the association between *TREM2* and AD was reported by another group alongside the deCODE findings [58]. This latter study utilized a combination of DNA sequencing approaches (including reanalysis of WES and WGS data) and reported evidence for the presence of other rare *TREM2* variants in addition to Arg47His to show association with AD. Since these original reports, the Arg47His association with AD has been replicated by several other groups [59–61]. More recently, Arg47His has also been—albeit less consistently—associated with other forms of neurodegenerative disorders (such as PD [62, 63], FTLD [63] and ALS [64]), although these latter findings could not be replicated in independent

datasets [65]. Functionally, TREM2 is likely involved in the body's innate immune system response based on evidence suggesting that the encoded receptor protein has been shown to regulate the phagocytic ability and inflammatory response of microglia [66]. The primary form of resident macrophages in the central nervous system (CNS).

4 Conclusions and Outlook

For the first time in the history of human genetics research, it is now both technically feasible and economically affordable to systematically screen individual genomes for novel disease-causing mutations at base-pair resolution using "next-generation sequencing". Thus far, only relatively few studies have applied these powerful new technologies to search for novel AD-related variants. Notable NGS-based discoveries until early 2014 include the identification of rare susceptibility-modifying alleles in *APP*, *TREM2*, and *PLD3*. Of these, the latter two gene findings are "novel" in the sense that these loci had not previously been linked to AD predisposition and pathophysiology. Several additional large-scale NGS projects are currently underway and many more "*TREM2*-like" discoveries can be expected to emerge from these and other studies over the coming years.

Despite this very exciting and highly promising outlook on AD genetics research made possible by the application of NGS, a few cautionary notes appear justified. First, despite their powerful and maximally exhaustive nature, utilizing NGS in gene discovery efforts does not preclude devising a careful and typically multipronged study design that includes experiments aimed at establishing a firm link between the identified DNA variants and the disease under study. Examples include proving segregation within affected families, independent replication of suspected findings, and carrying out functional experiments. Otherwise, the community will be left with situations currently encountered for *SORL1* or *NOTCH3*, where some evidence suggests an involvement in AD pathogenesis, while other crucial evidence in support of these hypotheses is still lacking. Actually, it can be argued that precisely *because* of the powerful and exhaustive nature of NGS a careful study design and execution is more direly needed than ever before: every single genome analyzed thus far by WGS has been found to contain dozens to hundreds of rare and apparently functional DNA sequence that proved to be without pathogenic consequences [15]. Second, it should be emphasized that in contrast to highly penetrant and disease-causing mutations, e.g. those encountered in *APP* or *PSEN1*, rare-variant associations of modest effect size, e.g. those reported for *TREM2*, are of little value as predictors or diagnostic tools in a clinical setting [61]. This is due to their incomplete penetrance, meaning that a sizable fraction of the

population carries the disease-associated variants without ever developing AD. This situation is not unlike that observed for common variants, e.g. those identified by GWAS. Third, as outlined in the introduction, this review only covers studies utilizing NGS for genome resequencing which is only one of several possible NGS applications. This, of course, does not mean that all or even most of the "missing heritability" in AD can solely be attributed to alterations in the genomic sequence. As a matter of fact, there is both theoretical and empirical evidence suggesting that much of the underlying heritability in AD may be due to alterations beyond the genome sequence, e.g. in epigenetic DNA profiles. Their identification and characterization, however, is more complex and requires the application of other NGS-based technologies, including bisulfite sequencing (to assess DNA methylation patterns), as well as RNA and chromatin immunoprecipitation (ChIP) sequencing. However, as epigenetic profiles are often specific to tissues or even cell-types the choice of appropriate biomaterial becomes crucial, but is difficult to resolve in a brain disease such as AD.

Notwithstanding these limitations, the increasingly widespread application and further development of NGS over the coming years will undoubtedly lead to a vast extension of our knowledge and understanding of the molecular processes underlying the onset and progression of AD and other neurodegenerative diseases. As such they will hopefully pave the way for developing novel therapeutics and biomarkers allowing to effectively prevent or halt the progression of this devastating disease.

Acknowledgements

This work was sponsored by funding from the Cure Alzheimer Fund, the Fidelity Biosciences Research Initiative, and the German Federal Ministry for Education and Research (BMBF grant #16SV5538).

References

1. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174
2. Tanzi RE, Bertram L (2005) Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. Cell 120:545–555
3. Bettens K, Sleegers K, Van Broeckhoven C (2013) Genetic insights in Alzheimer's disease. Lancet Neurol 12:92–104
4. Ridge PG, Ebbert MTW, Kauwe JSK (2013) Genetics of Alzheimer's disease. Biomed Res Int 2013:254954
5. Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. Neuron 68:270–281
6. Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci 9:768–778
7. Strittmatter WJ, Saunders AM, Schmechel D et al (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90:1977–1981

8. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23
9. Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–753
10. Zuk O, Hechter E, Sunyaev SR et al (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 109:1193–1198
11. Mardis ER (2013) Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto, Calif) 6:287–303
12. Eisenstein M (2012) The battle for sequencing supremacy. Nat Biotechnol 30:1023–1026
13. Zuk O, Schaffner SF, Samocha K et al (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111:E455–E464
14. Vinkhuyzen AA, Wray NR, Yang J et al (2013) Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet 47:75–95
15. Goldstein DB, Allen A, Keebler J et al (2013) Sequencing studies in human genetics: design and interpretation. Nat Rev Genet 14:460–470
16. Kiezun A, Garimella K, Do R et al (2012) Exome sequencing and the genetic basis of complex traits. Nat Genet 44:623–630
17. Thomas DC, Yang Z, Yang F (2013) Two-phase and family-based designs for next-generation sequencing studies. Front Genet 4:276
18. Bamshad MJ, Ng SB, Bigham AW et al (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12:745–755
19. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415–425
20. Sims D, Sudbery I, Ilott NE et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132
21. Wang S, Xing J (2013) A primer for disease gene prioritization using next-generation sequencing data. Genomics Inform 11:191–199
22. Normand R, Yanai I (2013) An introduction to high-throughput sequencing experiments: design and bioinformatics analysis. Methods Mol Biol 1038:1–26
23. Ladouceur M, Dastani Z, Aulchenko YS et al (2012) The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. PLoS Genet 8:e1002496
24. 1000 Genomes Project Consortium, Abecasis GR, Auton A et al (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65
25. Cruts M, Theuns J, Van Broeckhoven C (2012) Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat 33:1340–1344
26. Cruchaga C, Haller G, Chakraverty S et al (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset alzheimer's disease families. PLoS One 7:e31039
27. Jin SC, Pastor P, Cooper B et al (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res Ther 4:34
28. Benitez BA, Karch CM, Cai Y et al (2013) The PSEN1, p.E318G variant increases the risk of Alzheimer's disease in APOE-ε4 carriers. PLoS Genet 9:e1003685
29. Lord J, Turton J, Medway C et al (2012) Next generation sequencing of CLU, PICALM and CR1: pitfalls and potential solutions. Int J Mol Epidemiol Genet 3:262–275
30. Guerreiro RJ, Beck J, Gibbs JR et al (2010) Genetic variability in CLU and its association with Alzheimer's disease. PLoS One 5:e9510
31. Bettens K, Brouwers N, Engelborghs S et al (2012) Both common variations and rare non-synonymous substitutions and small insertion/deletions in CLU are associated with increased Alzheimer risk. Mol Neurodegener 7:3
32. Ferrari R, Moreno JH, Minhajuddin AT et al (2012) Implication of common and disease specific variants in CLU, CR1, and PICALM. Neurobiol Aging 33:1846.e7–e18
33. Yu J-T, Ma X-Y, Wang Y-L et al (2013) Genetic variation in clusterin gene and Alzheimer's disease risk in Han Chinese. Neurobiol Aging 34:1921.e17–e23
34. Jiang T, Yu J-T, Tan M-S et al (2014) Genetic variation in PICALM and Alzheimer's disease risk in Han Chinese. Neurobiol Aging 35:934.e1–e3
35. Ma X-Y, Yu J-T, Tan M-S et al (2014) Missense variants in CR1 are associated with increased risk of Alzheimer' disease in Han Chinese. Neurobiol Aging 35:443.e17–e21
36. Lupton MK, Proitsi P, Danillidou M et al (2011) Deep sequencing of the Nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease. PLoS One 6:e17298
37. Lupton MK, Proitsi P, Lin K et al (2014) The role of ABCA1 gene sequence variants on risk

of Alzheimer's disease. J Alzheimers Dis 38: 897–906

38. Bertram L (2011) Alzheimer's genetics in the GWAS era: a continuing story of "replications and refutations". Curr Neurol Neurosci Rep 11:246–253
39. Guerreiro RJ, Lohmann E, Kinsella E et al (2012) Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease. Neurobiol Aging 33:1008.e17–e23
40. Pottier C, Hannequin D, Coutant S (2012) High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset alzheimer disease. Mol Psychiatry 17:875–879
41. Andersen OM, Reiche J, Schmidt V et al (2005) Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. Proc Natl Acad Sci U S A 102:13461–13466
42. Lambert J-C, Ibrahim-Verbaas CA, Harold D et al (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45:1452–1458
43. Cruchaga C, Karch CM, Jin SC et al (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505:550–554
44. Heilmann, Stefanie, Dmitriy Drichel, Jordi Clarimon, Victoria Fernández, André Lacour, Holger Wagner, Mathias Thelen, et al. 2015. "PLD3 in Non-Familial Alzheimer's Disease." Nature 520 (7545): E3–5. doi:10.1038/nature14039.
45. Hooli, Basavaraj V., Christina M. Lill, Kristina Mullin, Dandi Qiao, Christoph Lange, Lars Bertram, and Rudolph E. Tanzi. 2015. "PLD3 Gene Variants and Alzheimer's Disease." Nature 520 (7545): E7–8. doi:10.1038/nature14040.
46. Lambert, Jean-Charles, Benjamin Grenier-Boley, Céline Bellenguez, Florence Pasquier, Dominique Campion, Jean-Francois Dartigues, Claudine Berr, Christophe Tzourio, and Philippe Amouyel. 2015. "PLD3 and Sporadic Alzheimer's Disease Risk." Nature 520 (7545): E1. doi:10.1038/nature14036.
47. Van der Lee, Sven J., Henne Holstege, Tsz Hang Wong, Johanna Jakobsdottir, Joshua C. Bis, Vincent Chouraki, Jeroen G. J. van Rooij, et al. 2015. "PLD3 Variants in Population Studies." Nature 520 (7545): E2–3. doi:10.1038/nature14038.
48. Cai D, Zhong M, Wang R et al (2006) Phospholipase D1 corrects impaired betaAPP trafficking and neurite outgrowth in familial Alzheimer's disease-linked presenilin-1 mutant neurons. Proc Natl Acad Sci U S A 103: 1936–1940
49. Oliveira TG, Chan RB, Tian H et al (2010) Phospholipase d2 ablation ameliorates Alzheimer's disease-linked synaptic dysfunction and cognitive deficits. J Neurosci 30: 16419–16428
50. Jonsson T, Atwal JK, Steinberg S et al (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96–99
51. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116
52 Kero M, Paetau A, Polvikoski T et al (2013) Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. Neurobiol Aging 34:1518.e1–e3
53. Liu Y-W, He Y-H, Zhang Y-X et al (2014) Absence of A673T variant in APP gene indicates an alternative protective mechanism contributing to longevity in Chinese individuals. Neurobiol Aging 35:935.e11–e12
54. Ting SKS, Chong M-S, Kandiah N et al (2013) Absence of A673T amyloid-β precursor protein variant in Alzheimer's disease and other neurological diseases. Neurobiol Aging 34: 2441.e7–e8
55. Bamne MN, Demirci FY, Berman S et al (2014) Investigation of an amyloid precursor protein protective mutation (A673T) in a North American case-control sample of late-onset alzheimer's disease. Neurobiol Aging 35:1779.e15–e16
56. Peacock ML, Warren JT Jr, Roses AD et al (1993) Novel polymorphism in the A4 region of the amyloid precursor protein gene in a patient without Alzheimer's disease. Neurology 43:1254–1256
57. Lill CM, Liu T, Schjeide BMM et al (2012) Closing the case of APOE in multiple sclerosis: no association with disease risk in over 29 000 subjects. J Med Genet 49:558–562
58. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127
59. Pottier C, Wallon D, Rousseau S et al (2013) TREM2 R47H variant as a risk factor for early-onset alzheimer's disease. J Alzheimers Dis 35:45–49
60. Benitez BA, Cooper B, Pastor P et al (2013) TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiol Aging 34:1711.e15–e17
61. Bertram L, Parrado AR, Tanzi RE (2013) TREM2 and neurodegenerative disease. N Engl J Med 369:1565

62. Benitez BA, Cruchaga C, United States–Spain Parkinson's Disease Research Group (2013) TREM2 and neurodegenerative disease. N Engl J Med 369:1567–1568
63. Rayaprolu S, Mullen B, Baker M et al (2013) TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. Mol Neurodegener 8:19
64. Cady J, Koval ED, Benitez BA et al (2014) TREM2 Variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis. JAMA Neurol 71:449–453
65. Lill CM, Rengmark A, Lasse P, Fogh I, Shatunov S, Sleiman PM, Wang LS et al. (2015) The role of TREM2 R47H as a risk factor for Alzheimer's Disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's sisease." Alzheimer's & Dementia; doi:10.1016/j.jalz.2014.12.009
66. Neumann H, Takahashi K (2007) Essential role of the microglial triggering receptor expressed on myeloid cells-2 (TREM2) for central nervous tissue immune homeostasis. J Neuroimmunol 184:92–99

Chapter 18

Pooled-DNA Sequencing for Elucidating New Genomic Risk Factors, Rare Variants Underlying Alzheimer's Disease

Sheng Chih Jin, Bruno A. Benitez, Yuetiva Deming, and Carlos Cruchaga

Abstract

Analyses of genome-wide association studies (GWAS) for complex disorders usually identify common variants with a relatively small effect size that only explain a small proportion of phenotypic heritability. Several studies have suggested that a significant fraction of heritability may be explained by low-frequency (minor allele frequency (MAF) of 1–5 %) and rare-variants that are not contained in the commercial GWAS genotyping arrays (Schork et al., Curr Opin Genet Dev 19:212, 2009). Rare variants can also have relatively large effects on risk for developing human diseases or disease phenotype (Cruchaga et al., PLoS One 7:e31039, 2012). However, it is necessary to perform next-generation sequencing (NGS) studies in a large population (>4,000 samples) to detect a significant rare-variant association. Several NGS methods, such as custom capture sequencing and amplicon-based sequencing, are designed to screen a small proportion of the genome, but most of these methods are limited in the number of samples that can be multiplexed (i.e. most sequencing kits only provide 96 distinct index). Additionally, the sequencing library preparation for 4,000 samples remains expensive and thus conducting NGS studies with the aforementioned methods are not feasible for most research laboratories.

The need for low-cost large scale rare-variant detection makes pooled-DNA sequencing an ideally efficient and cost-effective technique to identify rare variants in target regions by sequencing hundreds to thousands of samples. Our recent work has demonstrated that pooled-DNA sequencing can accurately detect rare variants in targeted regions in multiple DNA samples with high sensitivity and specificity (Jin et al., Alzheimers Res Ther 4:34, 2012). In these studies we used a well-established pooled-DNA sequencing approach and a computational package, SPLINTER (short indel prediction by large deviation inference and nonlinear true frequency estimation by recursion) (Vallania et al., Genome Res 20:1711, 2010), for accurate identification of rare variants in large DNA pools. Given an average sequencing coverage of 30× per haploid genome, SPLINTER can detect rare variants and short indels up to 4 base pairs (bp) with high sensitivity and specificity (up to 1 haploid allele in a pool as large as 500 individuals). Step-by-step instructions on how to conduct pooled-DNA sequencing experiments and data analyses are described in this chapter.

Key words Next-generation sequencing, NGS, Rare variants, Alzheimer's disease, Pooled-DNA sequencing

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_18,

1 Introduction

Alzheimer's disease (AD) is a complex disease with an estimated phenotypic heritability (the percentage of the variation in a trait that is due to variation in genetic factors) up to 80 % [1], which suggests that understanding genetic effects on AD can provide significant insights into developing effective therapeutics approaches to delay disease onset and eventually cure AD. In the past decade, identification of genes and genetic variation contributing to human diseases became cheap and feasible due to advances in our understanding of the human genome, rapid development of technologies to integrate the genome, and methods to analyze genomic data. The development of genome-wide association studies (GWAS) has facilitated the scan of genetic markers across the human genome to reveal disease-associated genetic loci. Recent GWAS have identified several loci affecting AD risk and implicated several biological pathways involved in disease processes [2–5]. However, the genetic variants identified using GWAS are common (MAF >1 %) and are usually with small effect sizes (odds ratio (OR) between 0.8 and 1.2). Moreover, common variants identified in GWAS only explain ~23 % phenotypic heritability [6], suggesting that additional genetic variants remain undiscovered. Several studies have suggested that a significant fraction of heritability may be explained by low-frequency (minor allele frequency (MAF) of 1–5 %) and rare-variants that are not contained in the commercial GWAS genotyping arrays [7–9]. These variants can also have relatively large effects on risk for developing human diseases [10–15]. In 2013, two independent studies [16, 17] used next-generation sequencing technology to identify a rare variant, p.Arg47His, in *TREM2* significantly associated with risk for Alzheimer's disease, with an OR similar to that of an individual carrying one *APOE* ε4 allele (OR = 3 ~ 4) [18]. Similarly, by analyzing whole-exome sequencing in large late-onset AD families and follow-up genotyping in several independent datasets, a rare variant, p.Val232Met, in *PLD3* was found to increase AD risk by 2.1-fold [19]. Additionally, our previous work has demonstrated that rare functional variants in *APP*, *PSEN1-2*, *MAPT* and *GRN* can be identified in AD patients and could account for a significant portion of risk for AD, which was not detected by GWAS [11, 20, 21]. In these studies, we utilized a powerful and cost-effective methodology called pooled-DNA sequencing [22, 23], which combines experimental and computational strategies to quantify and to detect rare variants associated with human diseases.

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations. We do not add sodium azide to the reagents.

2.1 DNA Quantification and Pooling Components

1. TE buffer (20×) (Life Technologies Corporation): 200 mM Tris–HCl, 20 mM EDTA, pH 7.5.
2. Quant-iT™ PicoGreen dsDNA reagent (200×) (Life Technologies Corporation): Solution in DMSO. Store at 4 °C.
3. Lambda DNA standard: 100 μg/mL in TE. Store at 4 °C.
4. 384-well black microplates (VWR, International Inc.).
5. Fluorescence microplate reader.

2.2 PCR Amplification, Purification, Quantification, and Pooling Components

1. dNTP mix: 10 mM each.
2. Forward primer: Stock solution at 10 μM in ultrapure water (Final concentration 200 nM).
3. Reverse primer: Stock solution at 10 μM in ultrapure water (Final concentration 200 nM).
4. 10× Pfu Ultra HF Reaction Buffer (Agilent Technologies, Inc.).
5. Betaine (5 M solution in water) (Sigma-Aldrich Co. LLC) (Final concentration 1 M).
6. Pfu Ultra High-Fidelity DNA Polymerase (2.5 U/μL) (Agilent Technologies, Inc.).
7. 100 bp DNA ladder (New England Biolabs Inc.).
8. GeneMate LE Agarose (BioExpress).
9. Agencourt AMPure XP PCR Purification (Beckman Coulter, Inc.).
10. TBE buffer: Prepare a 10× stock solution in 1 L of ultrapure water: 108 g of Tris base; 55 g of boric acid; 40 mL of 0.5 M EDTA (pH 8.0). The 0.5× working solution is 45 mM Tris–borate/1 mM EDTA.

2.3 Library Preparation Components

1. T4 Ligase (40k U/mL) (New England Biolabs Inc.).
2. T4 PNK (10k U/mL) (New England Biolabs Inc.).
3. End Repair Reaction Buffer (10×) (New England Biolabs Inc.).
4. 10× Pfu Ultra HF Reaction Buffer (Agilent Technologies, Inc.).
5. 100 bp DNA ladder (New England Biolabs Inc.).
6. GeneMate LE Agarose (BioExpress).
7. Covaris microtube (Covaris, Inc.).
8. TruSeq Resuspension Buffer (Illumina, Inc.).

9. TruSeq End Repair Mix (Illumina, Inc.).
10. TruSeq A-Tailing Mix (Illumina, Inc.).
11. TruSeq DNA Ligase Mix (Illumina, Inc.).
12. TruSeq DNA Adapter Index (Illumina, Inc.).
13. TruSeq Stop Ligase Mix (Illumina, Inc.).
14. TruSeq PCR Primer Cocktail (Illumina, Inc.).
15. TruSeq PCR Master Mix (Illumina, Inc.).
16. MinElute PCR Purification Kit (Qiagen).
17. Ethidium Bromide 10 mg/mL (Life Technologies Corporation).

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 DNA Quantification and Pooling

1. Use Invitrogen Quant-iT™ dsDNA Kit to determine the DNA concentration. It is important to accurately quantify the stock solution. If the stock DNA solution concentration exceeds 100 ng/μL, make a dilution to a final concentration of less than 100 ng/μL. It is important to use high-quality DNA (with a 260/280 ratio between 1.8 and 2). Otherwise, the yield of PCR experiments may be not ideal. Low-quality DNA will lead to false negative results and low sensitivity.
   (a) Prepare a 1× TE solution by diluting the 20× TE buffer 20-fold with nuclease-free water.
   (b) Prepare a DNA standard curve: make tenfold serial dilutions to create a six-point standard curve ranging from 100 to 0.001 ng/μL by diluting the bacteriophage lambda DNA (provided at 100 μg/mL in the Quant-iT™ PicoGreen Kits) with a 1× TE solution.
   (c) Transfer 10 μL of each dilution from standard curve solutions and 10 μL of unknown DNAs into the wells of a 384-well black microplate (VWM International, LLC). Mix well by pipetting up and down.
   (d) Prepare 1× Quant-iT™ dsDNA reagent by diluting the 200× Quant-iT™ dsDNA reagent 200-fold with 1× TE solution.
   (e) Add 10 μL of 1× Quant-iT™ dsDNA reagent to each well (total mix volume is 20 μL with the standards and the unknown DNA). Mix well and incubate for 5 min at room temperature protected from light.
   (f) Measure the fluorescence using a spectrofluoreometer or fluorescence microplate reader and standard flurorescein wavelengths (excitation: 460 nm; emission: 530 nm).

(g) Fit a linear (or any appropriate) model using the fluorescence emission intensity against the standard curve concentrations. Then, extrapolate each DNA concentration based on the fitted model.

2. Combine an equal amount of DNA (~150 ng) (*see* **Note 1**) per individual (*see* **Note 2**) in a 5-mL microcentrifuge tube.

3.2 PCR Amplification, Purification, Quantification, and Pooling

1. Design the PCR primers (*see* **Note 3**).
2. Prepare the PCR reaction mix, on ice following the steps in Table 1 (*see* **Notes 4** and **5**).
3. Pre-program the thermal cycler using the program in Table 2.
4. Start the thermal cycling program.

Table 1
Protocol for Pfu PCR amplification[a]

Component	Amount	Final concentration
10× Pfu Ultra HF reaction buffer	5 µL	1×
Betaine (5 M)	10 µL	1 M
dNTP mix, 10 mM each	1 µL	200 µM each
Forward primer (10 µM)	0.5 µL	0.2 µM
Reverse primer (10 µM)	0.5 µL	0.2 µM
Pooled DNA (20 ng gDNA copies/sample)	Variable (*see* **Note 4**)	<0.5 µg/50 µL
Pfu Ultra High-Fidelity DNA polymerase (2.5 U/µL)	0.8 µL	2 U/50 µL
Nuclease-free water to a final volume of:	50 µL	

[a]Please add the reagents in this order

Table 2
Thermal cycling conditions for Pfu DNA Polymerase-mediated PCR amplification

Temperature	Time	Number of cycles
93 °C	2 min	1 cycle
93 °C 61 °C 68 °C	30 s 30 s 2 min	40 cycles
68 °C	10 min	1 cycle
4 °C	Constant	1 cycle

5. Once complete, run a gel to confirm that the PCR amplicon size is correct.
   (a) Prepare a 2 % agarose gel with ethidium bromide using 1× TBE Buffer.
   (b) Mix 2.5 μL of 2× Gel Loading dye and 5 μL of PCR amplicon products.
   (c) Prepare 100 bp DNA ladder (New England Biolabs Inc.) by mixing 4 μL of distilled water, 1 μL of 6× Blue Loading Dye, and 1 μL of DNA ladder.
   (d) Put the agarose gel in a gel electrophoresis unit and add 1× TBE Buffer.
   (e) Load the samples and ladder solutions onto the gel.
   (f) Run the gel at 150 V constant voltage for 45 min.
   (g) Read the gel on a Dark Reader Ultra violet trans-illuminator.
6. Purify PCR amplicons with Ampure XP Beads and elute in 40 μL nuclease-free water.
   (a) Incubate the Ampure XP beads at room temperature for 15 min.
   (b) Vortex beads until they are well dispersed.
   (c) Transfer 90 μL of beads to each well of the PCR plate containing 50 μL of the PCR amplicons. Mix gently and thoroughly and vortex briefly.
   (d) Incubate the PCR plate with the mixed solutions at room temperature for 5 min.
   (e) Place the PCR plate on the magnetic stand for 5 min until the solution is clear.
   (f) Remove the clear supernatant from each well of the PCR plate and discard.
   (g) While the PCR plate is on the magnetic stand, add 500 μL of 70 % ethanol to each well.
   (h) Let the PCR plate settle for 1 min and then remove and discard ethanol from each well.
   (i) Repeat **steps** (**g**) and (**h**).
   (j) While the PCR plate is on the magnetic stand, air dry the beads at room temperature for 15 min and then remove the plate from the magnetic stand.
   (k) Add 40 μL of nuclease-free water to beads. Mix and vortex well and incubate at room temperature for 5 min.
   (l) Place the PCR plate on the magnetic stand for 5 min until the solution is clear.
   (m) Transfer the clear supernatant from each well of the PCR plate to a new tube or well. Be careful to not remove the beads.

7. Use Invitrogen Quant-iT™ dsDNA Kit to determine the concentration of each purified amplicon (*see* Subheading 3.1, **step 1**) (*see* **Note 6**).
8. Pool an equal amount of each PCR amplicon (~3.5×10^{11} molecules per amplicon) (*see* **Note 7**) in a 1.7 mL microcentrifuge tube.
9. Add the positive control and negative control amplicons to each pool (*see* **Note 8**).

3.3 Ligation

1. Prepare the reaction mix, on ice (*see* Table 3) (*see* **Note 9**).
2. Gently pipette the entire volume up and down to mix thoroughly.
3. Incubate the mix at 20 °C for 24 h followed by 65 °C for 20 min and hold at 4 °C.
4. Purify with Ampure XP beads and elute in 134 μL nuclease-free water (*see* Subheading 3.2, **step 6**).
5. Save 2 μL of the ligated product in a tube to run on a gel later.

3.4 Fragmentation

1. Bring volume up to 130 μL with Nuclease-free water.
2. Transfer 130 μL to Covaris microtube.
3. Fragment the ligated products using the settings in Table 4.
4. Once complete, purify with 234 μL of Ampure XP Beads and elute in 55 μL of nuclease-free water (*see* Subheading 3.2, **step 6**).
5. Run a gel with 5 μL of clean fragmented product and 2 μL of clean ligated product previously stored (*see* Subheading 3.2, **step 5**).

Table 3
Protocol for ligation[a]

Component	Volume (μL)	Final concentration
10× End repair buffer	35	1×
T4 Ligase (40k U/mL)	5	200 units
T4 PNK (10 U/μL) = (10k U/mL)	10	100 units
Pooled amplicon (30 ng/μL)	172	5 μg
Polyethylene glycol (50 % w/v)	106	15 %
Nuclease-free water	22	–
Total	350	–

[a]Please add the reagents in this order

Table 4
Protocol for the Covaris sonicator

Setting	Value
Duty cycle	10 %
Intensity	4
Cycles per burst	200
Treatment time	55 s
Temperature	7.8 °C

3.5 End Repair

1. Mix 10 μL of Resuspension Buffer, 40 μL of thawed End Repair Mix and 50 μL of clean fragmented products in the PCR plate. Gently pipette the entire volume up and down to mix thoroughly.
2. Incubate the mix in a thermal cycler with the plate sealed for 30 min at 30 °C.
3. Once complete, purify the end-repaired products with 160 μL of Ampure XP Beads and elute in 17.5 μL of nuclease-free water (*see* Subheading 3.2, **step 6**).

3.6 Adenylate 3′ Ends

1. Add 2.5 μL of Re-suspension Buffer to each well of the PCR plate.
2. Add 12.5 μL of A-Tailing Mix to each well of the PCR plate. Gently pipette the entire volume up and down to mix thoroughly.
3. Incubate the mix in a thermal cycler with the plate sealed for 30 min at 37 °C.

3.7 Adaptor Ligation

1. Add 2.5 μL of Re-suspension Buffer to each well of the PCR plate.
2. Add 2.5 μL of Ligation Mix to each well of the PCR plate.
3. Add 2.5 μL of the appropriate DNA Adapter Index to each well of the PCR plate. Gently pipette the entire volume up and down to mix thoroughly (*see* **Note 10**).
4. Incubate the mix on the thermal cycler with the plate sealed for 10 min at 30 °C.

3.8 Add Stop Ligase Mix STL

1. Add 5 μL of Stop Ligase Mix to inactivate the ligation. Gently pipette the entire volume up and down to mix thoroughly.
2. Purify the sample with 51 μL of Ampure XP Beads and elute in 22 μL of nuclease-free water (*see* Subheading 3.2, **step 6**).

3.9 Purify Ligation Products

1. Prepare a 2 % agarose gel (with ethidium bromide) using 1× TBE buffer and wide combs.
2. Add 4 μL of 6× Gel Loading dye to each well of the PCR plate.
3. Prepare 100 bp DNA ladder (New England Biolabs Inc.) by mixing 4 μL of distilled water, 1 μL of 6× Blue Loading Dye, and 1 μL of DNA ladder.
4. Put the agarose gel in a gel electrophoresis unit and add 1× TBE Buffer.
5. Load the samples and ladder solutions onto the gel.
6. Run the gel at 150 V constant voltage for 45 min.
7. Read the gel on a Dark Reader UV transilluminator
8. Weigh the empty 1.5 mL microcentrifuge tubes.
9. Excise a band from the gel spanning the width of each lane ranging in size from 300 to 600 bp using a clean new scalpel. Use the DNA ladder as a guide.
10. Weigh the tubes containing gel slices and follow the instructions in the MinElute Gel Extraction Kit to purify each sample, eluting in a MinElute column with 12 μL of nuclease-free water.

3.10 Enrich DNA Fragments

1. Add 3 μL of purified ligation products from each gel extraction to each well of the PCR plate.
2. Add 5 μL of thawed PCR Primer Cocktail to each well of the PCR plate.
3. Add 25 μL of thawed PCR Master Mix to each well of the PCR plate. Gently pipette the entire volume up and down to mix thoroughly. Then seal the PCR plate.
4. Pre-program the thermal cycler using the program in Table 5.
5. Place the sealed PCR plate on the thermal cycler and start the program.

Table 5
Thermal cycling conditions for enrichment

Temperature	Time	Number of cycles
98 °C	30 s	1 cycle
98 °C 60 °C 72 °C	10 s 30 s 30 s	12 cycles
72 °C	5 min	1 cycle
4 °C	Constant	1 cycle

6. Once complete, purify with 50 μL of Ampure XP Beads and elute in 30 μL of nuclease-free water (*see* Subheading 3.2, **step 6**).

3.11 Preparation for Illumina Sequencing

1. Use Invitrogen Quant-iT™ dsDNA Kit to determine the concentration of each pool (*see* Subheading 3.1, **step 1**) (*see* **Note 6**).
2. Dilute each Ligation enriched pool down to 10 nM with appropriate amount of nuclease-free water.
3. Combine an equal volume of each pool in one tube for Illumina Hiseq/Miseq.

3.12 Sequencing Data Analysis (See Note 11)

In order to have an accurate call, it is necessary to have at least 30× coverage per allele and individual. If this coverage is not reached, we recommend that additional sequencing data should be generated.

1. Compress the file to reduce computer time for alignment (*see* **Note 12**) (Optional).
2. Align reads back to a reference genome (*see* **Note 13**).
3. Tag aligned reads (*see* **Note 14**).
4. Generate and visualize the error model (*see* **Note 15**).
5. Merge tagged files (*see* **Note 16**) (Optional).
6. SPLINTER detection of rare variants (*see* **Note 17**).
7. Variant filtering (*see* **Note 18**).

3.13 Variant Validation

To validate the variants called by pooled-DNA sequencing, direct-genotyping can be performed using the existing genotyping methods, e.g., the Sequenom iPLEX platform (Sequenom, San Diego, CA, USA), the KASPar Genotyping Assay (LGC, Middlesex, UK) or the Taqman SNP Genotyping Assays (Life Technologies Corporation, NY, USA).

4 Notes

1. We find that 94 samples per pool is an appropriate pool size resulting in good sensitivity and specificity in identifying rare variants. Also, this pool size makes it easy to perform direct genotyping.
2. The pools can be indexed, combined, and sequenced in one lane. We found that in one lane of the HighSeq2500, it is possible to sequence 40 KB (around 5 genes) in 4,000 samples (42 pools) with a mean coverage > 30 per individual and allele.
3. The genomic sequences can be downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov) or the UCSC Genome Browser (http://genome.ucsc.edu). We suggest using the web-based Primer3 tool (http://bioinfo.ut.ee/primer3)

for the design of PCR primers. We have found that amplicons of 600–2k bp are typically ideal, if possible. Smaller amplicons may be necessary and will work fine, but amplicons larger than 1.5k bp, while achievable, typically do not amplify as robustly as those around 1k bp or less. The input conditions for Primer3 to obtain optimal functioning primers are as follows: (1) Primer minimum size = 19; (2) Optimum size = 25; (3) Maximum size = 30; (4) Minimum Tm = 64 °C; (5) Optimum Tm = 70 °C; (6) Maximum Tm = 74 °C; (7) Maximum Tm difference = 5 °C; (8) Minimum GC content = 45; (9) Maximum GC content = 80; (10) Number of return = 20; (11) Maximum 3′ end stability = 100.

4. In order to include 20 copies per individual, the amount of pooled DNA required for PCR is equal to (7 pg/gDNA) × (20 copies) × (Number of individuals in the pool).
5. We find most exons of genes of interest could be amplified with PfuUltra High-Fidelity (Stratagene) DNA Polymerase. If Pfu doesn't work, try Phusion Polymerase 2× Mastermix (Sigma) or 2× Extensor Long Range PCR Master Mix (Thermo Fisher Scientific Inc, USA).
6. We suggest extracting 5 μL of PCR amplicon and making a tenfold dilution with nuclease-free water. Then use Quant-iT™ PicoGreen dsDNA reagent to quantify the dilution and to calculate the original amplicon concentration.
7. In order to contain 3.5×10^{11} molecules, the volume (μL) of amplicon solution needed to be added to the pool is equal to $(3.5 \times 10^{11}) \times (1/6.02 \times 10^{23}) \times (660) \times (1/10^{-9}) \times$ (Amplicon length (bp)) × (1/Amplicon concentration (ng/μL)).
8. SPLINTER requires the presence of two components: a negative control and a positive control. A negative control (1–2 kb of cloned plasmid DNA) is used to generate a run-specific error model. We include a 1,276 bp region of the pUC19 plasmid. We use the first 800 bases to train our algorithm and the remaining 476 bases as a test set. Use a positive control consisting of a synthetic DNA library simulating an artificial pool with mutations engineered at a known position and frequency. These mutations consist of single nucleotide substitutions. We use the consensus sequence of the 72-bp exon 9 from TP53 (RefSeq accession no. NM_000546) as the "wild-type" insert into a pGEM-T Easy vector (Promega). A panel of different variations of this consensus sequence containing single, double, and 4-bp indels, as well as single nucleotide substitutions (Transitions and Transversions in all the possible combinations) is combined with the "wild-type". The positive control must be engineered according to the pool size, since the ratio between the wild-type and each mutation present in the library will determine the sensitivity of the variant detection.

9. The volume of the DNA, the PEG, water, and mix can be modified. Always use the minimum volume possible.
10. TruSeq DNA Sample Preparation Kit (Illumina, Inc.) provides several indexed adapter sequences to multiplex multiple pooled libraries for hybridization onto a flow cell.
11. The SPLINTER program can be used to perform read alignment, allele frequency estimation, and variant calling. You can find the software downloads and details of the program at http://www.ibridgenetwork.org/wustl/splinter. SPLINTER is a command line program and is better run in a UNIX/Linux environment.
12. Read files should be converted into SCARF format or compressed. The compression step is optional but it can significantly reduce computation time and storage space. Use the following command line for data compression.

    ./RAPGAP_read_compressor_v3.pl [Read file in FASTQ or SCARF format] > [Compressed read file]

    E.g.: ./RAPGAP_read_compressor_v3.pl run_1_1_sequence.fq.gz > RAPGAP_run_1_1_sequence.fastq
13. The SPLINTER aligner can perform gapped alignment by aligning the read file to the FASTA reference sequence file which contains the target sequences and positive and negative controls. We recommend allowing 3 bp-mismatch since the most common form of indels has a length less than 3 bp. Additionally, the quality of alignment is inversely proportional to the number of mismatches allowed. Use the following command line for alignment.

    ./RAPGAPHASH5d [Compressed read file] [Reference sequence in FASTA format] [Number of mismatches allowed] > [Aligned read file]

    E.g.: ./RAPGAPHASH5d RAPGAP_run_1_sequence.fastq Reference_Sequence.txt 3 > Aligned_RAPGAP_run_1_sequence.txt.
14. This step gives the aligned read file a unique tag in order to recognize reads coming from the same sequencing run and is important for combining multiple lanes from the same pool. Each sequencing run generates a unique error profile and can be distinguished from the tags and thus different tags should be given to aligned read files generated by different pools or machine run. Use the following command line for read tagging.

    ./RAPGAP_alignment_tagger.pl [Aligned read file] [Tag] > [Aligned tagged file]

    E.g.: ./RAPGAP_alignment_tagger.pl Aligned_RAPGAP_run_1_sequence.txt Tag1 > Aligned_Tag1_RAPGAP_run_1_sequence.txt.
15. Use the following command line to generate the error model.

    ./EMGENERATOR4 [Aligned tagged file] [Negative control sequence file] [Error model output file name] [5′ most

base of the negative control to be used] [3′ most of the negative control to be used] [Include unique reads only?] [Alignment edits cutoff] [Enter pseudocounts?]

E.g.: ./EMGENERATOR4 Aligned_Tag1_RAPGAP_run_1_sequence.txt PCMV_sequence.txt EM_Aligned_Tag1_RAPGAP_run_1_sequence 0 800 y 3 y

Use the following command line to visualize the error model.

./error_model_tabler_v4.pl [Error model 0th order file] [Output file name]

E.g.: ./error_model_tabler_v4.pl EM_Aligned_Tag1_RAPGAP_run_1_sequence_0 EM_Aligned_Tag1_RAPGAP_run_1_sequence.pdf

See Fig. 1 for a typical error-model plot. By visualizing the error model, you can determine the cycles to skip in the following SPLINTER analyses. The error rate cutoff is defined by one mismatch divided by the number of chromosomes (which is 2 times the number of individuals in a pool).

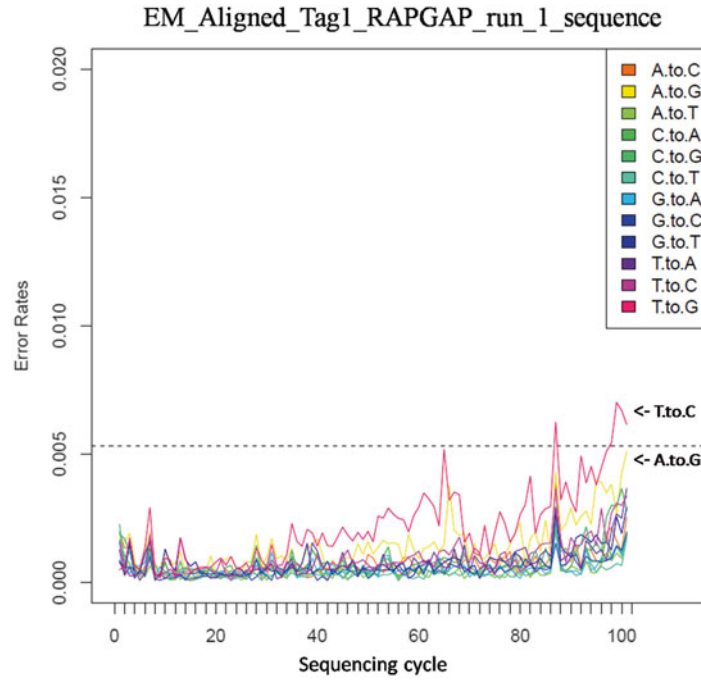


Fig. 1 The error model from negative control reads. The error model was calculated using the negative control reads with a read length of 102 bp (cycle) generated from Illumina HiSeq 2500. The *x*-axis represents the sequencing cycle and the *y*-axis represents the negative-control error rate. The *dash line* indicates the error-rate cutoff defined by one mismatch divided by the number of chromosome (which is two times the number of individuals in a pool). The sequencing cycle with an error rate exceeding the cutoff will be excluded from the SPLINTER analyses. Based on this error model, the first 98 cycles were used for SPLINTER analyses while skipping cycles 65 and 87

16. This step is optional (if you used a single-end sequencing strategy this step is unnecessary). Use the following command line to merge two tagged files.

    cat [Aligned tagged file one] [Aligned tagged file two] > [Output file name]

    E.g.: Assume you have two tagged files from the same pool named "Aligned_Tag1.1_RAPGAP_run_1_1_sequence.txt" and "Aligned_Tag1.3_RAPGAP_run_1_3_sequence.txt"

    cat Aligned_Tag1_RAPGAP_run_1_1_sequence.txt Aligned_Tag1_RAPGAP_run_1_3_sequence.txt > Aligned_Tag1.1_3_RAPGAP_run_1.1_3_sequence.txt.

17. Use the following command line to call rare variants.

    ./SPLINTER6t [Aligned tagged file] [Reference sequence in FASTA format] [2nd order error model file] [Number of read bases to be used] [Read bases or cycles to be excluded] [P-value cutoff] [Use unique reads?] [Alignment edits cutoff] [Pool size from the available options] [Print out the absolute coverage per strand?] > [Output file name]

    This analysis below was based on the error model shown in Fig. 1. The first 98 cycles were used but cycles 65 and 87 were skipped due to surpassing the error rate cutoff.

    E.g.: ./SPLINTER6t Aligned_Tag1_RAPGAP_run_1_sequence.txt Reference_Sequence.txt EM_Aligned_Tag1_RAPGAP_run_1_sequence_2 98 65,87 -2 y 3 1000 y > SPLINTER_Aligned_Tag1_RAPGAP_run_1_sequence_98cycles_skip65_87.txt.

18. The filtering process will normalize areas in the sequence that has been subjected to high sequencing coverage. The presence of positive controls, which consist of a DNA fragment with site-specific mutations artificially spiked-in at one allele frequency for one person, allows the calculation of a *p*-value cutoff in order to maximize sensitivity and specificity. Use the following command line to filter rare variants.

    ./splinter_filter_v3.pl [SPLINTER file] [Positive control hits] [Stringency parameter] > [SPLINTER filtered file]

    E.g.: ./splinter_filter_v3.pl SPLINTER_Aligned_Tag1_RAPGAP_run_1_sequence_98cycles_skip65_87.txt positive_ctrl.txt 0.0001 > Filter_SPLINTER_Aligned_Tag1_RAPGAP_run_1_sequence_98cycles_skip65_87.txt

    A good filtering step is necessary to narrow down the number of variants to validate. The positive and negative controls should be used to model the filtering steps. All the expected variants from the positive control should be properly called, and none should be called from the negative control.

Acknowledgements

This work was supported by grants from the National Institutes of Health (P30-NS069329, R01-AG044546 and P01AG003991, and R01NS085419), the Alzheimer Association (NIRG-11-200110) and Barnes Jewish Foundation. This research was conducted while C.C. was a recipient of a New Investigator Award in Alzheimer's disease from the American Federation for Aging Research.

References

1. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174
2. Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41: 1088–1093
3. Naj AC, Jun G, Beecham GW et al (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset alzheimer's disease. Nat Genet 43:436–441
4. Hollingworth P, Harold D, Sims R et al (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43:429–435
5. Lambert JC, Heath S, Even G et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094–1099
6. So HC, Gui AH, Cherny SS et al (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 35:310–317
7. Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40:695–701
8. Schork NJ, Murray SS, Frazer KA et al (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19: 212–219
9. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124–137
10. Rivas MA, Beaudoin M, Gardet A et al (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43:1066–1073
11. Cruchaga C, Chakraverty S, Mayo K et al (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset alzheimer's disease families. PLoS One 7: e31039
12. Haller G, Kapoor M, Budde J et al (2014) Rare missense variants in CHRNB3 and CHRNA3 are associated with risk of alcohol and cocaine dependence. Hum Mol Genet 23:810–819
13. Nejentsev S, Walker N, Riches D et al (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324:387–389
14. Momozawa Y, Mni M, Nakamura K et al (2011) Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. Nat Genet 43:43–47
15. Emison ES, Garcia-Barcelo M, Grice EA et al (2010) Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 87:60–74
16. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127
17. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116
18. Saunders AM, Strittmatter WJ, Schmechel D et al (1993) Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. Neurology 43:1467–1472
19. Cruchaga C, Karch CM, Jin SC et al (2014) Rare coding variants in Phospholipase D3 (PLD3) confer risk for Alzheimer's disease. Nature 505:550–554
20. Jin SC, Pastor P, Cooper B et al (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res Ther 4:34

21. Benitez BA, Karch CM, Cai Y et al (2013) The PSEN1, p.E318G variant increases the risk of Alzheimer's disease in APOE-epsilon4 carriers. PLoS Genet 9:e1003685
22. Vallania FL, Druley TE, Ramos E et al (2010) High-throughput discovery of rare insertions and deletions in large cohorts. Genome Res 20:1711–1718
23. Vallania F, Ramos E, Cresci S et al (2012) Detection of rare genomic variants from pooled sequencing using SPLINTER. J Vis Exp (64). pii: 3943. doi: 10.3791/3943

Chapter 19

New Genome-Wide Methods for Elucidation of Candidate Copy Number Variations (CNVs) Contributing to Alzheimer's Disease Heritability

Kinga Szigeti

Abstract

The complexity of human genetic variation has been extended by the observation of abundant and widespread variation in the copy number of submicroscopic DNA segments. The discovery of this novel level of genome organization opened new possibilities concerning the genetic variation that may confer susceptibility to or cause disease. Copy number variants (CNVs) influence gene expression, phenotypic variation and adaptation by altering gene dosage and genome organization. Concordant with the common disease common variant hypothesis these structural variants are now subject to interrogation for disease association. Alzheimer's disease (AD) is a progressive neurodegenerative disease with an estimated heritability of 60–80 %. Large scale genome-wide association studies (GWAS) using high frequency single nucleotide polymorphism (SNP) variants identified ten loci which do not account for the measured heritability. To find the missing heritability systematic assessment of all mutational mechanisms needs to be performed. Between the powerful SNP-GWAS studies and the planned Whole Genome Sequencing projects the contribution of copy number variation (CNV) to the genetic architecture of AD needs to be studied fully.

Key words Copy number variation, CNV, Alzheimer's disease, Heritability, GWAS

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia and leads to unrelenting cognitive decline [1]. With increased longevity the prevalence of AD in the elderly represents a major public health problem. The heritability of AD is estimated at 60–80 % [2]. Several large scale genome-wide association studies (GWAS) using high frequency variants identified ten loci including APOE with a combined population attributable fraction of 0.51–0.6 [3]. To find the missing heritability systematic assessment of all mutational mechanisms needs to be performed.

The advent of whole-genome scanning methods revealed widespread variation in the copy number of submicroscopic DNA segments. Copy number variation (CNV) is defined as a DNA

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_19,

segment that is 1 kb or larger and is present at variable copy number in comparison with the reference genome [4]. CNVs are a group of structural variants and can be classified as deletions, duplications, deletions and duplications at the same locus, multi-allelic loci, and complex rearrangements.

CNVs are major contributors to genetic variance, thus it is conceivable that they may contribute to the heritability of disease [5]. CNVs influence gene expression, phenotypic variation and adaptation by altering gene dosage [5]; 18 % of the gene expression traits are associated with CNVs [6].

CNVs have been identified in Mendelian disease and were found to be associated with complex neurological traits. Duplication of amyloid precursor protein (*APP*) causes autosomal dominant early-onset AD with cerebral amyloid angiopathy [7], duplication and triplication of α-synuclein (*SNCA*) causes familial Parkinson disease [8], and lamin B1 (*LMNB1*) duplication causes leukodystrophy [9], all confirmed by segregation of the disease phenotype with the CNV in autosomal dominant families. CNV GWAS studies implicated several candidate loci contributing to the AD phenotype [10–17].

The recombination events resulting in CNVs may be frequent. At the whole genome level about 0.3 % of biallelic CNV genotypes exhibit Mendelian discordance in parent-offspring trios [5, 18]. End tissue mosaicism could add additional complexity and introduce overlap between CNV states [19].

CNV studies leveraging the single nucleotide polymorphism (SNP) arrays used in traditional GWAS face multiple challenges, including variable coverage per platform, batch effects, and limited resolution due to inferior dynamic range [20]. To overcome these difficulties, the first iteration of CNV analyses of SNP arrays in AD applied very similar workflows, concentrating on high stringency calls [10–17]. The majority of the studies performed genotyping on the Illumina platform with a coverage in the 600 k range, except the Translational Genomics Research Institute (TGEN) study which used the Affymetrix 6.0 array with two million probes (Table 1). The analysis methods were strikingly similar. To comply with the high stringency inferred CNV principal shared in all the studies, CNVs were excluded from the analysis based on number of probes, size and overlap with CNV variant regions or segmental duplications, sometimes even based on frequency (Table 2). These studies investigated only the tip of the iceberg with high specificity but low sensitivity for CNV detection. Importantly, these association studies addressed the question whether rare, large CNVs contribute to the genetic architecture of AD; however, due to the very low allele frequencies of these large variants, most studies were not powered and very large sample sizes are needed over 10,000 cases and controls. Variants with higher frequencies blur the CNV calls, as the Kernel distributions overlap due to the fact that the derived

Table 1
Published CNV GWAS studies; case-control

Study	Platform	Input DNA	AD	MCI	Control
GERAD	Illumina 610-quad	200 ng	3,260	0	1,290
ADNI	Illumina Human610-Quad	NA	288	183	184
Caribbean Hispanics	Illumina HumanHap 650Y	NA	559	0	554
Duke	Illumina Human Hap550K	NA	331	0	368
TGEN	Affymetrix 6.0	NA	1,022	0	595
NCRAD	Illumina Human610-Quad	NA	711	0	171

Table 2
Outline of the methodology applied in published CNV GWAS studies; case-control

Study	LogR calculation	Reference file	Segmentation algorithm	Model	CNV exclusion
GERAD	BeadStudio	Not mentioned	PennCNV	Hidden Markov Model	<20 probes, <100 kb, density <1/15 kb, >50 % overlap with segdup
ADNI	GenomeStudio	Not mentioned	PennCNV	Hidden Markov Model	<10 probes, overlap with centromeric and immunoglobulin regions
Caribbean Hispanics	BeadStudio	Not mentioned	QuantiSNP, iPattern, PennCNV, CNVpartition	Multiple	<5 probes, <100 kb, overlap with centromeric and immunoglobulin regions, 50 % overlap with segdup, >1 % frequency
Duke	BeadStudio	Not mentioned	PennCNV	Hidden Markov Model	10 SNPs, 50 % overlap with previously published regions
TGEN	Unknown	Not mentioned	PennCNV	Hidden Markov Model	10 SNPs, 50 % overlap with centromeric, telomeric and immunoglobulin regions
NCRAD	GenomeStudio	Not mentioned	PennCNV	Hidden Markov Model	<10 probes, likelihood ratio <10, centromeric, immunoglobulin

reference genome tends to deviate from the diploid state, both of which increase the genotyping error rate and thus decreases power. These are harder to study on the SNP arrays and have been eliminated in the first iteration.

An alternative analysis strategy is to use segmentation only to reduce the dataset where events may occur, perform the test of association on the numeric segmented data and validate the CNV calls if a replicated association signal was detected [16, 17]. This approach detects association signals from smaller events that would have been discarded when performing the high confidence calls and overcomes the need to determine exact dosage, which is often problematic at common CNV loci as the reference may deviate from the diploid state. This approach signifies a screen and requires diligent validation and replication.

As GWAS studies are performed with increasing sample sizes [3, 21–23] it is becoming clear that in disorders with marked genetic heterogeneity, where the marker specific risk is low in case-control sets, it is difficult to identify the true positives from the false positives and to replicate the results [24, 25]. In addition, case-control design in AD suffers from additional confounders, such as misclassification bias due to age-dependent penetrance. To further empower association studies, quantitative endophenotypes may replace traditional case-control designs, examples being age at onset analysis or using expression quantitative trait loci (eQTL). Genetic variation, both single nucleotide variations (SNV) and copy number variations (CNVs), contribute to changes in gene expression. In some cases these variations are meaningfully correlated with disease states [26].

2 Methods

Three major high-resolution methods are currently available for detection of gene dosage at the genome level [27, 28]. Deep sequencing methodologies can detect CNVs, but is prohibitively expensive for whole genome studies of large patient cohorts. Most of the available copy number data has been collected using array comparative genomic hybridization (aCGH) or derived from SNP arrays (Table 3). Both approaches are subject to rapid technical improvements. The methodology in SNP arrays includes an amplification step which reduces the resolution of the CN calls. Another important difference is the derivation of CNV state in relation to a reference genome: while aCGH uses a single genome in every experiment as a common denominator (1 to 1 comparison), the SNP arrays use a bioinformatically generated reference genome from multiple cases (1 to average comparison). In aCGH the labeling is controlled at every single array, while in a SNP array the reference value will depend on the normalization efficiency and the

Table 3
Comparison of SNP array and array comparative genome hybridization principles

	aCGH	SNP array
Design	Main application	Secondary application
Probes empirically tested	Yes	No
Amplification step	No	Yes
Reference sample	Intraexperimental	Extraexperimental, reference mean of >40 samples
Interarray variability	Compensated for by reference sample	Compensated for by normalization
Intraarray variability	Compensated for by normalization	Compensated for by normalization
Optimization for sensitivity and specificity	For CNV	For SNP calling

allele frequency of any given CNV. Whole genome and exon sequencing methods are in development [29]. However, cohorts of AD with adequate sample sizes are not available yet.

3 Data Analysis Workflows

Several analysis workflows have been proposed and used. The first iteration workflows focused on high stringency calls optimizing specificity over sensitivity, and these often aim for redundancy to further enhance true positives. Other workflows focus on sensitivity over specificity, enhance resolution and apply complementary methods to fully explore datasets with the implied necessity for replication in additional cohorts with orthogonal methods. The selection of the workflow depends on the research question and requires orthogonal validation methods with locus specific high throughput assays (e.g. PCR or long range PCR for breakpoint, qPCR, TaqMan assay or multiplex ligation-dependent probe amplification (MLPA) for dosage).

3.1 Workflow for SNP Array Data

3.1.1 Quality Control (QC)

1. Experimental quality control: Most major platforms developed QC packages independently and these are incorporated in the data capturing process. Most of these QC parameters focus on signal to noise ratio as this is the key for the segmentation algorithms. For the Affymetrix arrays contrast QC and median absolute pair-wise differences are calculated, while for the Illumina arrays mean, median and standard deviation of the logR ratios and the B allele frequency are determined. Arrays with CNV calls more than two SD from the mean are eliminated as this reflects uneven baseline with false positive CNV calls.

2. Data quality control: The parallel detected SNP alleles allow additional QC measures including SNP call rate (exclude <97 %), gender mismatch (by X chromosome logR ratio), and related or duplicate samples (Pi > 0.95) by determining IBD with PLINK software using the genotype data.

3.1.2 Population Substructure/Admixture by the SNP Dataset

The concomitant SNP detection allows the determination of underlying substructure or admixture using principal component analysis (PC). Most of the commercially available software packages incorporate this feature and the most commonly used Eigensoft package is available open access. Primary analysis focuses on Caucasian subjects; in addition the PCs are used as covariates in the statistical model.

3.1.3 LogR Ratio Calculation

The log2 ratio calculation is one of the key elements especially for the Affymetrix arrays, where the analyst can define the samples contributing to the derived reference genome. Due to batch effects, using within study samples as reference genome enhances data quality and many more arrays (up to 155 more) will pass QC measures. As a first pass the reference file is generated from all controls in the given dataset. Second pass the reference for the complete analysis is generated from the top 100 DLRS control samples to optimize the elimination of noise. Normalization of logR data is performed by cRMAv2 (Bioconductor). The logR ratio data is subjected to numeric principal component analysis (GoldenHelix) and corrected for the number of PCs which yields a QQ plot devoid of inflation.

3.1.4 Numeric Array Data or Segmentation

1. Numerical array data: Quantile normalized numeric data is used in the analysis as independent variable. This approach suffers from marked multiple testing burden but has the advantage of highest resolution.
2. Segmented numeric data: Normalized, PC corrected numeric data is segmented to identify probes where a CNV is detected in any of the samples in the set. The segmentation results in a reduced dataset while maintaining the advantages of the numeric data without binned CNVs.
3. Inferred CNVs: Different algorithms give different results even on a single dataset. The algorithms that are developed for a certain platform derived data appear to perform better than the generic CNV calling algorithms [20]. CNV calls are collapsed into regions and CNVs called by either algorithms or both algorithms are entered into the analysis depending upon the goals regarding sensitivity and specificity. Validation of CNV calls with aCGH is depicted in Fig. 1. The two algorithms used generated distinct but overlapping CNV calls; sev-

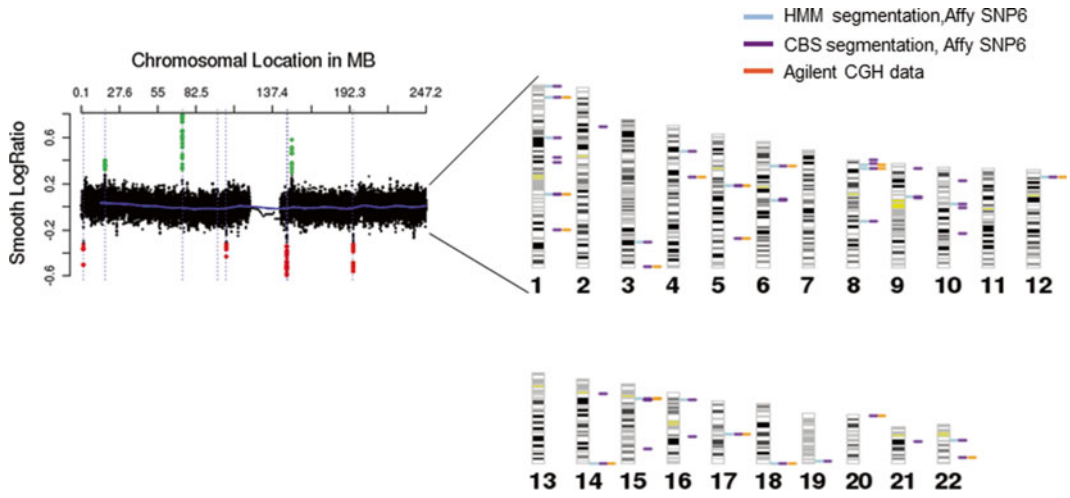


Fig. 1 Validation of CNV calls inferred by a Circular Binary Segmentation and a Hidden Markov Model by aCGH. Segmentation on 50 samples was performed using the Hidden Markov Model (HMM) algorithm implemented in Genotyping Console and the Circular Binary Segmentation (CBS) algorithm DNAcopy implemented in R. We used aCGH to reference the CNV calls to a gold standard in the same subjects. The number of events ascertained by the two segmentation algorithms was 2,282. CBS generated 2,060 CNV calls in the 50 subjects, while HMM generated 1,264 calls. 1,042 calls were overlapping between the two algorithms. There were 1,018 CBS only and 222 HMM only calls. aCGH validated a high percentage of single algorithm calls and all the double algorithm calls in regions where coverage was comparable. This suggests that the two segmentation algorithms are complimentary. The CNV calls are depicted to the right of the karyotype: HMM (*light gray*/*blue*), CBS (*dark gray*/*purple*) and aCGH (*medium gray*/*orange*) from *left* to *right*

eral of the single algorithm calls were validated by the aCGH. This suggests that some of the algorithms are complementary. A recent head to head comparison of various CNV calling algorithms from data captured on various platforms suggests that the algorithm developed for a specific dataset performs best; generic or algorithms developed for a different dataset has lower specificity and sensitivity [20].

3.1.5 *Test of Association*

1. Numerical array data as independent variable: This approach searches for genomically contiguous regions where CN state has an effect on case-control status. To enhance the analysis a "thin and bin" approach is applied.

   Thinning and Binning: Every other oligo is sampled to divide the data in half. In each half, K genomically adjacent oligos are binned and case-control association is performed on the mean CNV state within each thinned bin. False discovery rate (FDR) values for each thin bin p value is calculated, and the q-values for the CNV's coefficient from lowest (near 0) to highest (near 1) in each half is ranked. $K=2$ and $K=100$ is

tested empirically. The *K* at which maximum concordance is attained with FDR *q* values less than 0.05 in each data half is selected. The direction of effect (sign of the beta coefficient) is verified to be concordant.

Effects of moderate size: The case control association is performed on the entire dataset removing the thinning but retaining data aggregation into *K* oligo bins. FDR *q* values are calculated.

2. Segmented numeric data as independent variable: Appropriate statistical methods including *T*-test, or various regression models are used depending upon the dependent variable.
3. Inferred CNVs as independent variable: Appropriate statistical methods including *T*-test, or various regression models are used depending upon the dependent variable.

3.1.6 Visualization of Log2 Ratio Data

1. Kernel distribution: These distribution plots delineate the separation of the various copy number states and assess the probability of the CNV events.
2. Log2 ratio data as genomic location: These visualization strategies depict the size of the CNV, the signal to noise ratio, the number of probes covering the CNV and the consistency of adjacent log2 ratios, also assessing probability.

3.2 Workflow for aCGH Data

At the present time only Agilent microarrays are available. The aCGH workflow is similar, although there are a few distinct features of the data, including 1 to 1 comparison, higher dynamic range and the potential to customize. Data from AD cohorts on aCGH is extremely limited and additional studies are needed.

3.2.1 Quality Control (QC)

The QC parameter for aCGH is MAPD, and values < 0.3 fulfill stringent criteria; even up to 0.35 yields good quality segmentation. Arrays with CNV calls more than two SD from the mean are eliminated. Data with gender mismatch (by X chromosome logR ratio) are eliminated.

3.2.2 Population Substructure/Admixture by the SNP Dataset

Agilent microarray design may incorporate SNPs; however, adding SNPs to the Agilent design reduces the density of the copy number probes, thus degrades resolution. For most of the AD sample collections at least one type of SNP GWAS data is available and that can be used for population substructure/admixture analysis.

3.2.3 Log2 Ratio Calculation

Normalized log-ratio data is generated with the manufacturer's microarray scanner and quantification software (CGH analytics, Agilent). The log2 ratio is calculated between the sample and an intraexperimental control sample, a 1 to 1 comparison.

3.2.4 Numeric Array Data or Segmentation

Similar to the SNP arrays numeric, numeric segmented and inferred CNV calls can be used in downstream analyses.

1. Numerical array data: Quantile normalized numeric data is used in the analysis as independent variable.
2. Segmented numeric data: Numeric PC corrected data is segmented to identify probes where a CNV is detected in any of the samples in the set. The segmentation results in a reduced dataset while maintaining the advantages of the numeric data without binned CNVs.
3. Inferred CNVs: The Agilent package CNV calling algorithm is used. The high dynamic range results in superior accuracy compared to SNP arrays, especially for multi copy gains. Due to the dynamic range and the uniformity of the data (mostly Agilent at this point, since Nimblegene stopped manufacturing aCGH) algorithm development is stable. The sensitivity and specificity data reflects that 70 % of CNV events detected with three probes are validated by orthogonal methods and over 90 % of CNV events detected with five probes are validated by orthogonal methods. CNV calls are collapsed into regions and entered into the analysis as independent variable.

3.2.5 Test of Association

1. Numerical array data as independent variable: This approach searches for genomically contiguous regions where CN state has an effect on case control status. To enhance the analysis a "thin and bin" approach is applied, as described in the SNP array workflow.
2. Segmented numeric data as independent variable: Appropriate statistical methods including *T*-test, or various regression models are used depending upon the dependent variable.
3. Inferred CNVs as independent variable: Appropriate statistical methods including *T*-test, or various regression models are used depending upon the dependent variable.

3.2.6 Visualization of Log2 Ratio Data

1. Kernel distribution: These distribution plots delineate the separation of the various copy number states and assess the probability of the CNV events.
2. Log2 ratio data as genomic location: These visualization strategies depict the size of the CNV, the signal to noise ratio, the number of probes covering the CNV and the consistency of adjacent log2 ratios, also assessing probability.

For all statistical methods multiple testing correction is applied. The Bonferroni correction appears too conservative; the FDR approach and simulation, performing at least 1,000 permutations, are reasonable alternatives. As the analysis is redundant with the three types of data, arguable controlling for multiple testing burden in each results (numeric, numeric segmented and inferred CNVs) is sufficient. Validation of the CNV calls with orthogonal methods is

important to assess the locus specific genotyping error rate. Well-powered replication studies with a locus specific orthogonal method is necessary to replicate the association.

4 Concluding Notes

For Alzheimer's disease, only SNP array datasets are available currently, thus we discuss the caveats for these platforms and their analysis methods. The development of normalization and segmentation algorithms are a rapidly evolving field and vigilant monitoring is recommended. It is worthwhile to evaluate the novel algorithms on a subset of data that has orthogonal validated events for a couple of regions. Size and density of probes enhance for true positives but reduce resolution. Deletions are easier to detect due to the difference of distance from 0 between log2 of 1/2 versus log2 of 3/2. Common CNV regions further reduce the dynamic range as the calculated reference diploid genome are likely not diploid for that specific region. For common CNVs binning into CNV calls results in a very high genotyping error rate (Fig. 2). For these regions the numeric data or the numeric segmented data is superior for power. All GWAS studies require replication on independent datasets with alternative methods.

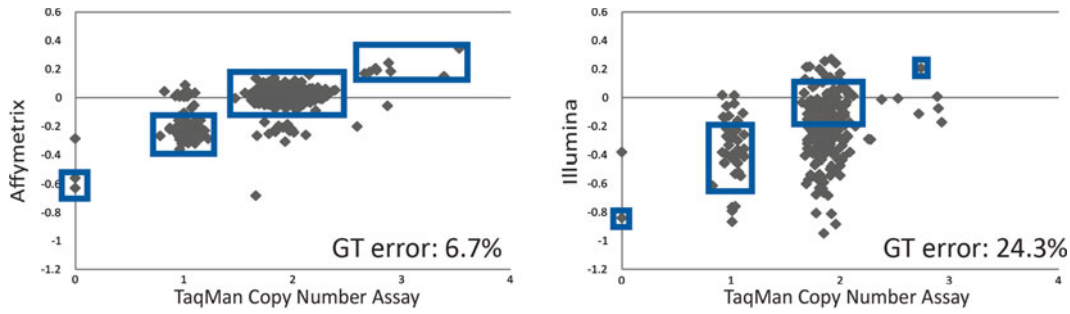


Fig. 2 Genotyping error rates for a frequent (10 %) variant, CHRFAM7A on the Affymetrix 6.0 and Illumina 610 arrays. CNVs were inferred based on the Kernel distributions of the segmented numeric data for the Affymetrix and the numeric data for the Illumina array as this latter dataset failed the segmentation algorithm. Breakpoint specific TaqMan assay was performed on the same samples to assess genotyping error rate. *Rectangles* represent concordant calls between the SNP array and the breakpoint specific TaqMan assay. The striking genotyping error rate for the Illumina 610 array further emphasizes the risks involved in assigning exact dosage; in these situations retaining the numeric values without binning implies a lower genotyping error rate, thus increases power

Acknowledgements

This work was supported by an Alzheimer Association New Investigator Research Grant to K.S.

References

1. Kukull WA, Higdon R, Bowen JD et al (2002) Dementia and Alzheimer disease incidence: a prospective cohort study. Arch Neurol 59: 1737–1746
2. Gatz M, Pedersen NL, Berg S et al (1997) Heritability for Alzheimer's disease: the study of dementia in Swedish twins. J Gerontol A Biol Sci Med Sci 52:M117–M125
3. Naj AC, Jun G, Beecham GW et al (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset alzheimer's disease. Nat Genet 43:436–441
4. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7:85–97
5. Redon R, Ishikawa S, Fitch KR et al (2006) Global variation in copy number in the human genome. Nature 444:444–454
6. Stranger BE, Forrest MS, Dunning M et al (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848–853
7. Rovelet-Lecrux A, Hannequin D, Raux G et al (2006) APP locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–26
8. Singleton AB, Farrer M, Johnson J et al (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841
9. Padiath QS, Saigoh K, Schiffmann R et al (2006) Lamin B1 duplications cause autosomal dominant leukodystrophy. Nat Genet 38:1114–1123
10. Chapman J, Rees E, Harold D et al (2013) A genome-wide study shows a limited contribution of rare copy number variants to Alzheimer's disease risk. Hum Mol Genet 22:816–824
11. Swaminathan S, Huentelman MJ, Corneveaux JJ et al (2012) Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. PLoS One 7:e50640
12. Swaminathan S, Kim S, Shen L et al (2011) Genomic copy number analysis in Alzheimer's disease and mild cognitive impairment: an ADNI study. Int J Alzheimers Dis 2011: 729478
13. Swaminathan S, Shen L, Kim S et al (2012) Analysis of copy number variation in Alzheimer's disease: the NIALOAD/NCRAD family study. Curr Alzheimer Res 9:801–814
14. Heinzen EL, Need AC, Hayden KM et al (2010) Genome-wide scan of copy number variation in late-onset alzheimer's disease. J Alzheimers Dis 19:69–77
15. Ghani M, Pinto D, Lee JH et al (2012) Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics. G3 (Bethesda) 2:71–78
16. Shaw CA, Li Y, Wiszniewska J et al (2011) Olfactory copy number association with age at onset of Alzheimer disease. Neurology 76:1302–1309
17. Szigeti K, Lal D, Li Y et al (2013) Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease. J Alzheimers Dis 33:517–523
18. Conrad DF, Keebler JE, DePristo MA et al (2011) Variation in genome-wide mutation rates within and between human families. Nat Genet 43:712–714
19. McConnell MJ, Lindberg MR, Brennand KJ et al (2013) Mosaic copy number variation in human neurons. Science 342:632–637
20. Pinto D, Darvishi K, Shi X et al (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat Biotechnol 29: 512–520
21. Harold D, Abraham R, Hollingworth P et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41:1088–1093
22. Seshadri S, Fitzpatrick AL, Ikram MA et al (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303:1832–1840
23. Lambert JC, Heath S, Even G et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41: 1094–1099

24. Ku CS, Loy EY, Pawitan Y et al (2010) The pursuit of genome-wide association studies: where are we now? J Hum Genet 55: 195–206
25. Florez JC (2008) Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. J Clin Endocrinol Metab 93:4633–4642
26. Nicolae DL, Gamazon E, Zhang W et al (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 6:e1000888
27. Carter NP (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 39(7 Suppl):S16–S21
28. Scherer SW, Lee C, Birney E et al (2007) Challenges and standards in integrating surveys of structural variation. Nat Genet 39(7 Suppl):S7–S15
29. Duan J, Zhang JG, Deng HW, Wang YP (2013) Comparative studies of copy number variation detection methods for next-generation sequencing technologies. PLoS One 8:e59128

Chapter 20

RNA-Sequencing to Elucidate Early Patterns of Dysregulation Underlying the Onset of Alzheimer's Disease

Bei Jun Chen, James D. Mills, Caroline Janitz, and Michael Janitz

Abstract

With its ability to perform rapid transcriptome profiling and profound transcriptomic analysis powered by high-throughput sequencing at a high resolution with deep coverage, the advent of RNA sequencing technology, RNA-Seq, outperforms other methods in the field, such as microarrays, and has changed our way of performing transcriptomic investigation. Protocols for preparing libraries for RNA-Seq using the Illumina and Roche 454 sequencing platforms are included in this chapter. Common steps for library preparation in both platforms include RNA fragmentation, cDNA synthesis, adaptor ligation, and PCR amplification of cDNA strands. Illumina adopts solid-phase bridge PCR amplification, while 454 uses water-in-oil emulsion-based PCR amplification. Despite differences in the PCR amplification step, both platforms employ the same sequencing-by-synthesis technology for the sequencing process. Application of the RNA-Seq technique in the context of dysregulation of the transcriptome in Alzheimer's disease is also discussed.

Key words RNA-Seq, Transcriptome, Alzheimer's disease, cDNA library preparation

1 Introduction

The transcriptome is the complete repertoire of transcripts in a cell. The diversity and abundance of transcripts reflect the gene expression pattern of the particular cell and are determined by the cell's specific tissue type, its developmental stage, and its physical condition. The transcriptome is the blueprint for the downstream protein synthesis process and serves as a link between encoded DNA sequences and the cell's manifested phenotype.

By comparing differences in gene expression patterns, transcriptomics provides an invaluable tool for studying cell biology. Knowledge of the transcriptomic changes that occur in different conditions, such as diseased versus healthy tissues or during development, is extremely useful for understanding the physiological and pathological processes involved within a cell.

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_20,

There are approximately 300,000 RNA molecules in a human cell [1]. This abundance of transcripts makes genome-wide expression analyses daunting. RNA microarrays first made transcriptomics analyses possible and has remained the main transcriptomic method over the past years [2]. A microarray employs thousands of chip-embedded oligonucleotide probes for the detection of target RNA molecules within a sample. The quantities of each RNA molecule hybridized to the probes are measured according to the color intensity emitted by the fluorescent dye attached to the probes. However, the color intensity-based quantification method is limited, and subtle changes in gene expression across different samples are difficult to capture. Microarrays also require prior knowledge of the genes under investigation, so discovery of novel genes or transcript isoforms is technically complicated. Another problem with microarrays is that cross-hybridization is commonly observed, which presents a source of artifact error.

The recently developed high-throughput next-generation RNA-Seq technique overcomes the inherent limitations of microarrays. Using large-scale, parallel-direct sequencing, RNA-Seq generates short reads of sequences from starting sample materials at a much faster speed compared with the first generation or Sanger sequencing methods. Reads are later mapped to a reference genome or assembled de novo. Mapped reads are assembled into transcripts for downstream data analysis.

Various data analysis tools have been developed to process gigabit-sized RNA-Seq output data, such as Bowtie, TopHat, and Cufflinks, which are all part of the Tuxedo Suite. For bioinformaticians who are less command-line savvy, web-based tools, such as Galaxy (http://galaxyproject.org/), which accommodates the entire Tuxedo Suite software, are also available.

The employment of RNA-Seq-facilitated transcriptome analysis covers a wide range of applications, including the investigation of different RNA expression patterns between various conditions or during developmental stages, the detection of splicing junctions and novel isoforms, the construction of transcript structures such as the 5′ and 3′ UTRs, the detection of RNA editing events, small RNA analysis, and allele-specific polymorphism profiling [3–5].

Currently, there are three main RNA-Seq platforms on the market: Illumina, Roche 454, and SOLiD, and relevant protocols for Illumina and Roche 454 are presented in this chapter. Although these platforms adopt different technologies, the principles behind DNA sequencing are essentially the same: they all rely on signals captured after dNTPs are incorporated into an extending cDNA strand. This principle is known as sequencing-by-synthesis technology.

1.1 Transcriptome Studies in Alzheimer's Disease

Alzheimer's disease (AD) is a complex, progressive neurodegenerative disease that has large social and economic costs throughout the world. While extensive resources have been directed toward understanding the pathogenesis of AD, little progress has been

made in elucidating the underlying mechanisms. Currently, no cure exists, and treatment tools are limited; these tools must be developed to lessen the impact of the impending epidemic.

A small number of AD cases (1–2 %) follow Mendelian inheritance patterns and are caused by mutations in amyloid precursor proteins (*APP*), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*) [6]. The vast majority of AD cases (>95 %) are 'sporadic', meaning they show no familial or geographical associations; however, an estimated 60–80 % of these cases may be genetically determined [7]. While numerous gene candidates have been associated with AD [8, 9], proving causation remains elusive. The traditional molecular methods used to study DNA and proteins are unable to contribute further to our understanding of AD. The human brain is a complex organ, and one of the key elements observed with an increase in tissue and organism complexity is an increase in the number of transcriptional elements and transcriptional regulation methods, including the presence of non-coding RNAs, alternative splicing (AS) and RNA editing [10–12]. These elements must be finely tuned for the brain to function correctly; any disturbances could result in abnormal function. The underlying cause of AD may come from dysregulation within the transcriptome. RNA-Seq is currently the best tool available for a comprehensive analysis of the transcriptome of brain tissue. A review by Sutherland et al. [13] advocates the use of RNA-Seq for transcriptome profiling in AD brains.

Currently, numerous papers suggest that dysregulation of the transcriptome plays a role in AD and other neurodegenerative diseases [14, 15]. More recently, a small number of papers have used RNA-Seq to profile the transcriptomes of different regions of the AD brains from both humans and mice models [16–18]. RNA-Seq has identified an association between altered splicing patterns in the genes diazepam-binding inhibitor (*DBI*) and apolipoprotein E (*APOE*) and AD [17, 18]. The altered splicing patterns in *APOE* observed in AD brain tissue [18] presents an interesting case that highlights the usefulness of RNA-Seq. These findings demonstrate a novel pathway in which the well-known gene *APOE* may be involved in the pathogenesis of AD.

While RNA-Seq has become more widely used in research pertaining to other diseases, such as cancer, it has yet to gain the same traction in profiling the brain transcriptome. This lag may be due to the inherent difficulties with handling human brain tissue; however, these difficulties can be overcome through the use of appropriate techniques and experimental design. Multifaceted diseases, such as AD, likely have complex causative factors. With all of its advantages, RNA-Seq has the potential to detect these factors, leading to the identification of biomarkers and new treatment targets.

1.2 Illumina Sequencing

Illumina currently manufactures the most comprehensive high-throughput sequencing systems, such as its HiSeq series; these systems are able to generate up to three billion bases of single reads or

six billion bases of paired-end reads in a single sequencing run, totaling a data output of 600 gigabits with a maximum read length of 150 bp. While the full cycle for high output sequencing takes approximately 8.5–11 days, a rapid run can be completed in 27 hours.

Different library preparation protocols exist for different applications. The general RNA-Seq library preparation protocol includes the extraction and fragmentation of polyA RNA molecules, synthesis of cDNA by reverse transcription, end repair and addition of adenine bases, adaptor ligation, cDNA fragment size selection, adaptor-ligated cDNA amplification by PCR reactions, and purification of amplified cDNA molecules. The enriched library is then diluted to a 2 nM solution and is ready for downstream sequencing procedures.

Sequencing begins after the cDNA library that is obtained from the previous preparation step is captured and amplified on the surface of the flow cell using a cBot, which is an automated system for clonal amplification of sequencing libraries. Currently, there are two types of flow cells available depending on the sequencing mode used. The standard 'High Output' flow cell contains eight lanes, and a 'Rapid' flow cell contains only two lanes. On the flow cell surface, there are dense lawns of oligonucleotide primers to which the adaptors on cDNA fragments are annealed. CDNA strands are then attached to the flow cell surface and form single-stranded bridge-like cDNA structures on it. CDNA PCR amplification begins after the addition of nucleotides and enzymes and generates double-stranded bridge clusters of cDNA fragments on the cell surface. Double-stranded cDNA is denatured in the following step, and the reverse strands are washed away (Fig. 1).

Sequencing cycles start after deoxynucleotides (dNTPs), primers, and enzymes are added to the solution. These dNTPs are labeled with four removable fluorescent dyes that emit different colors when excited by a laser. DNA synthesis stops after a dNTP has been incorporated into the elongating strand and resumes after the fluorescent dye emits a signal and is consequently cleaved away. The signal is captured by a charge-coupled device (CCD) camera, and the identity of the incorporated nucleotide is determined and recorded. The cycle repeats, and the identities of the following nucleotides in the template sequence are determined in the same way. For paired end fragment sequencing, this procedure is followed by reverse strand sequencing after previous cluster resynthesis of the reverse strand using the HiSeq paired-end module. After the newly synthesized strand is cleaved and removed, the forward strands flip over, hybridize with primers on the lawn, and form single-stranded cDNA bridges. Reverse strands are synthesized and sequenced in the same manner after the forward strands are washed away.

Illumina sequencing technology has a number of advantages. First, the solid-phase bridge amplification feature allows the process to be fully automated. Second, Illumina uses a reversible

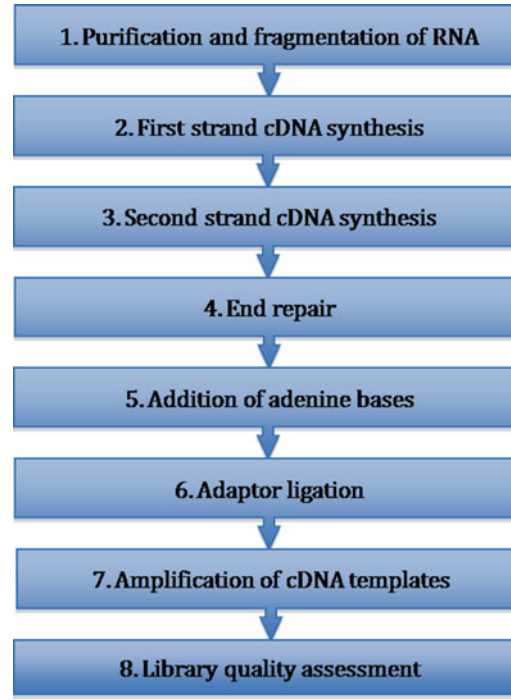


Fig. 1 Illumina RNA-Seq protocol outline for sequencing template preparation

terminator and employs the sequencing-by-synthesis approach. In this setting, single fluorescent signals are captured after each dNTP is incorporated; therefore, error rates are minimized, especially in repetitive and homopolymer regions. Third, the constantly increasing output in sequencing run and read length allowing for simultaneous sequencing of a larger number of samples, which in turn results in a dramatic decrease of sequencing costs. Finally, because DNA polymerase is the only enzyme used in the sequencing step, fewer errors are caused by defects in enzyme activity. However, the reads generated by Illumina are relatively short, and short reads means less overlapping and, hence, less information for transcript assembly.

1.3 Roche 454 Sequencing

454 Life Sciences, a Roche company, is another major player in the sequencing system market.

Suitable 454 systems for transcriptome sequencing include the GS FLX and GL FLX+ series, which are able to generate reads with lengths of 450–700 bp. The typical throughput of these systems ranges from 450 to 700 megabits.

Different library preparation protocols exist for different applications. The general RNA-Seq library preparation protocol includes the extraction, purification, and fragmentation of polyA

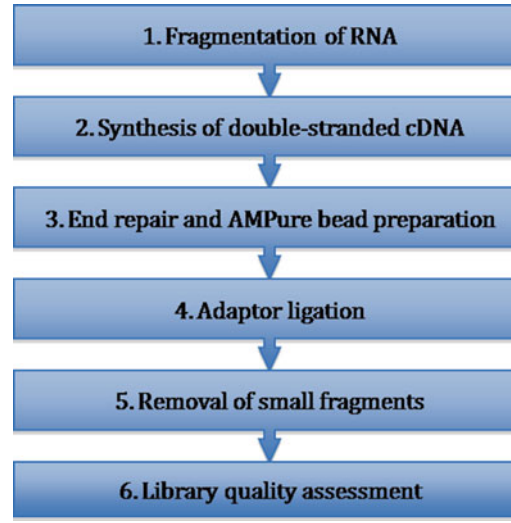


Fig. 2 Roche 454 RNA-Seq protocol outline of the template preparation for sequencing

RNA molecules; the synthesis of cDNA by reverse transcription; end repair; AMPure beads preparation; adaptor ligation; cDNA fragment size selection; library quality assessment; and quantification. The single-stranded cDNA library is then mixed in water with DNA capture beads and enzymes. After emulsion with synthetic oil, each bead is trapped in a droplet, forming a micro-reactor with one single-stranded cDNA fragment that will be amplified within the micro-reactor during the subsequent emulsion PCR (emPCR) amplification step (Fig. 2).

After the emPCR amplification, the beads are screened, washed, and transferred to the PicoTiterPlate, which has 1.6 million wells on its surface. Each well accommodates one and only one bead bound with millions of homologous single cDNA strands. In addition, beads with enzymes for pyrosequencing are present in the wells. Similar to the Illumina methods, Roche 454 also adopts the sequencing-by-synthesis method. During one sequencing run, dNTPs are sequentially added in waves. The nucleotide that is complimentary to the template strand is incorporated into the extending strand by DNA polymerase. A pyrophosphate is released and converted to ATP by sulfurylase. Luciferase utilizes ATP to generate a chemiluminescent signal that is captured by a CCD camera and recorded by the system. For a single nucleotide species, the signal intensity is proportional to the number of consecutive nucleotides that were incorporated.

The Roche 454 sequencing method usually generates longer reads that provide more overlapping information for transcript assembly. There are, however, certain drawbacks in 454

sequencing technology. For example, cDNA samples are amplified using emPCR, after which the samples must be manually transferred to a PicoTiterPlate, which lowers the degree of automation; therefore, the overall template preparation is much more complicated than the preparation for Illumina sequencing. Another issue is related to the chemiluminescence signal intensity that is used to measure how many nucleotides have been incorporated. This ambiguity in homopolymer detection could be problematic in regions that consist of more than eight consecutive homologous nucleotides where the signal falls off quickly [19]. Lastly, because a combination of enzymes is used, the sequencing process is relatively error-prone and less cost-efficient.

2 Materials

2.1 Illumina Sequencing

Samples, reagents and solutions for this protocol prepared according to procedures described in TruSeq RNA Sample Preparation v2 Guide from Illumina [20] (*see* **Note 1**).

1. Sample RNA.
2. Nuclease-free Ultra-pure water.
3. RNA-purification beads.
4. Bead-Binding Buffer.
5. Bead-Washing Buffer.
6. Elute, Prime, Fragment Mix.
7. Elution Buffer.
8. Resuspension Buffer.
9. First-Strand Master Mix.
10. SuperScript II reverse transcriptase.
11. Second-Strand Master Mix.
12. AMPure XP beads.
13. End-Repair Mix.
14. 80 % Ethanol.
15. A-Tailing Mix.
16. RNA-Adaptor Indices.
17. Ligation Mix.
18. Stop-Ligation Buffer.
19. PCR Master Mix.
20. PCR Primer Cocktail.

2.2 Roche 454 Sequencing

Samples, reagents and solutions for this protocol prepared according to procedures described in cDNA Rapid Library Preparation Method Manual, GS FLX+ Series—XL+ (May 2011) from Roche [21].

1. Sample RNA.
2. 10 mM Tris–HCl pH 7.5.
3. 70 % Ethanol.
4. RNA fragmentation solution: $ZnCl_2$ powder, 1 M Tris–HCl pH 7.0, Molecular Biology Grade Water.
5. 0.5 M EDTA pH 8.0.
6. 400 μM Roche primer "random".
7. RNAClean reagent.
8. Molecular Biology Grade Water.
9. cDNA Synthesis System Kit: vial 1: 5× RT-buffer AMV; vial 2: AMV RT, 25 U/μL; vial 3: DTT, 0.1 M; vial 4: Protector RNase Inhibitor, 25 U/μL; vial 7: dNTPs, 10 mM; vial 9: 5× 2nd strand synthesis buffer; vial 10: 2nd strand enzyme; vial 11: T4 DNA polymerase; vial 12: re-distilled water.
10. Rapid Library (RL) Library Prep Kit: RL 10× PNK Buffer, RL ATP, RL dNTP, RL T4 Polymerase, RL PNK, RL Taq Polymerase, RL Adaptor, RL MID Adaptor, and RL Ligase.
11. AMPure beads.
12. TE buffer.
13. Sizing solution.

3 Methods

3.1 Illumina Sequencing

This protocol has been prepared according to procedures described in TruSeq RNA Sample Preparation v2 Guide from Illumina (March 2014) [20].

3.1.1 Purification and Fragmentation of mRNA

1. Start with 0.1–1 μg of total RNA by preparing a dilution of RNA samples with nuclease-free ultra-pure water to a final volume of 50 μL in a 0.3 mL 96-well microplate labeled RNA Bead Plate (*see* **Note 1**).
2. Allow the RNA purification beads to come to room temperature (at least 30 min), and then, mix vigorously to resuspend the beads.
3. Add 50 μL of RNA purification beads to each sample in the well to allow binding between the oligo-dT-tailed magnetic beads and the poly-A mRNA molecules.
4. Seal the plate with the adhesive seal, and mix it for 2 min in a MixMate vortex at 1,600 rpm. Shortly spin the plate (*see* **Note 2**).

5. Transfer the RNA-Bead Plate to a pre-heated thermal cycler at 65 °C (conditions: 5 min at 65 °C with a pre-heated lid at 100 °C followed by a 4 °C hold), and incubate for 5 min to facilitate RNA denaturation and binding between the beads and RNA molecules.
6. After the thermal cycler reaches 4 °C, transfer the microplate to room temperature and continue the incubation for another 5 min.
7. Remove the seal, and transfer all samples to a new 0.3 mL 96-well microplate with round bottom wells, using a multi-channel pipette, for further bead purification. Transfer the RNA-Bead Plate to a magnetic stand for another 5 min to allow the RNA-bound magnetic beads to form pellets on the sides of the wells (*see* **Note 3**).
8. Slowly aspirate and discard the supernatant being careful not to disturb the bead pellets.
9. Remove the plate from the magnetic stand, and add 200 μL of Bead-Washing Buffer to each well to wash away unbound RNA.
10. Mix the plate for 2 min in a MixMate vortex at 850 rpm.
11. Place the RNA-Bead Plate on a magnetic stand at room temperature for 5 min to allow the RNA-bound magnetic beads to form pellets on the sides of the wells.
12. Slowly aspirate and discard the supernatant, which contains mostly ribosomes and other RNA species.
13. Remove the RNA-Bead Plate from the magnetic stand, and add 50 μL of Elution Buffer to each well.
14. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
15. Transfer all samples from the RNA-Bead Plate to a new 0.2 mL 96-well microplate, seal the plate with a Microseal 'B' Adhesive seal and incubate it for 2 min in a thermal cycler pre-heated to 80 °C (conditions: 2 min at 80 °C with a heated lid at 100 °C, followed by a 4 °C hold) to facilitate the elution of the bound mRNA from the beads.
16. Remove the RNA-Bead Plate form the thermal cycler when it reaches 25 °C, and transfer all of the samples to a new 0.3 mL 96-well microplate with round-bottom wells using a multi-channel pipette for further bead purification.
17. Add 50 μL of Bead-Binding Buffer to each well in the plate. The Bead-Binding Buffer allows the specific binding of mRNA and reduces non-specific binding of rRNA to the beads.
18. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
19. Incubate the RNA-Bead Plate at room temperature for 5 min.
20. Transfer the RNA-Bead Plate to the magnetic stand and leave for 5 min at room temperature.

21. Slowly aspirate and discard the supernatant.
22. Remove the RNA-Bead Plate from the magnetic stand, and add 200 μL of Bead-Washing buffer to each well of the plate.
23. Mix the plate for 2 min in a MixMate vortex at 850 rpm.
24. Transfer the RNA-Bead Plate to the magnetic stand and leave for 5 min at room temperature.
25. Slowly aspirate and discard the supernatant.
26. Remove the RNA-Bead Plate from the magnetic stand, and add 19.5 μL of the Elute, Prime, and Fragment Mix to each well of the plate. The Elute, Prime, and Fragment Mix contains reagents that elute, fragment, and prime RNAs within a single solution.
27. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm; then, transfer all samples to a new 0.3 mL 96-well microplate labeled RNA-Fragmentation Plate using a multichannel pipette.
28. Seal the plate with the adhesive seal, transfer it to a pre-heated thermal cycler at 94 °C (conditions: 8 min at 94 °C with a pre-heated lid at 100 °C followed by a 4 °C hold) and incubate for 8 min.
29. Remove the RNA-Fragmentation plate from the thermal cycler when the cycle is completed and spin briefly.

3.1.2 First Strand cDNA Synthesis

Now that the RNA molecules have been fragmented and primed with random hexamers, they are ready for first strand cDNA synthesis using reverse transcriptase and random primers.

1. Remove the adhesive seal from the RNA-Fragmentation Plate (after **step 29** in Subheading 3.1.1), and transfer all samples to a new 0.3 mL 96-well microplate with round-bottom wells using a multichannel pipette for further bead purification.
2. Place the RNA-Fragmentation Plate on the magnetic stand for 5 min.
3. Slowly aspirate and transfer 17 μL from each well to the corresponding wells of a new 0.3 mL 96-well microplate labeled cDNA Plate.
4. Prepare the Final First-Strand Master Mix (FFSMM) by mixing 7 μL of the First-Strand Master Mix (FSMM) with 1 μL of SuperScript II (SSII) for each sample including 10 % extra e.g., for 40 samples prepare a mix for 44 samples mix as follows 308 μL FSMM +44 μL SSII (total volume 352 μL).
5. Add 8 μL of the Final First-Strand Master Mix to each well of the plate, and seal the plate with the adhesive seal.
6. Mix the plate for 1 min in a MixMate vortex at 1,600 rpm. Spin the plate shortly.

7. Place the cDNA Plate in a thermal cycler, and run the following program:
   (a) Pre-heat lid and set to 100 °C.
   (b) Incubate at 25 °C for 10 min.
   (c) Incubate at 42 °C for 50 min.
   (d) Incubate at 70 °C for 15 min.
   (e) Hold at 4 °C indefinitely.

3.1.3 Second Strand cDNA Synthesis

1. Remove the adhesive seal from the cDNA Plate (after **step 7** in Subheading 3.1.2), and add 25 μL of the Second-Strand Master Mix to each well.
2. Seal the plate with the adhesive seal, and mix it for 1 min in a MixMate vortex at 1,600 rpm.
3. Incubate the cDNA Plate in a thermal cycler at 16 °C for 1 h with the lid closed.
4. Remove the cDNA plate for the thermal cycler, and let it come to room temperature.
5. Vortex to resuspend the AMPure XP beads (use only beads that have been previously brought to room temperature), and add 90 μL of the AMPure XP beads to each well of a new 0.3 mL 96-well round bottom microplate labeled cDNA Clean-Up Plate using a multichannel pipette.
6. Remove the seal, and transfer all samples from the cDNA Plate to the corresponding wells of the cDNA Clean-Up Plate containing the AMPure XP beads.
7. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
8. Incubate the cDNA Clean-Up Plate at room temperature for 15 min.
9. Place the cDNA Clean-Up Plate on a magnetic stand at room temperature for 5 min, and allow the magnetic beads to form a pellet on the side of the wells.
10. Slowly aspirate and discard 135 μL of the supernatant from each well of the cDNA Clean-Up Plate.
11. With the plate remaining on the magnetic stand, add 200 μL of 80 % ethanol to each well without disturbing the beads (*see* **Note 4**).
12. Incubate the plate at room temperature for 30 s; then, slowly aspirate and discard the supernatant from each well.
13. Repeat **steps 11** and **12** for a total of two ethanol washes.
14. Leave the plate on the magnetic stand for 15 min to allow the ethanol to evaporate completely; then, remove the cDNA Clean-Up Plate from the magnetic stand (*see* **Note 5**).

15. Add 52.5 μL of Resuspension Buffer to each well of the cDNA Clean-Up Plate.
16. Mix the plate for 2 min in a MixMate vortex at 1,000 rpm.
17. Incubate the cDNA Clean-Up Plate at room temperature for 2 min.
18. Leave the plate on the magnetic stand at room temperature for 5 min.
19. Transfer 50 μL of the supernatant, which contains the double-stranded cDNA from the cDNA Clean-Up Plate, to a new 0.2 mL 96-well microplate labeled as Insert-Modification Plate.

3.1.4 End Repair

1. Add 10 μL of Resuspension Buffer to each well of the Insert-Modification Plate containing 50 μL of cDNA.
2. Add 40 μL of End-Repair Mix to each well of the Insert-Modification Plate.
3. Seal the plate with the adhesive seal, and mix it for 1 min in a MixMate vortex at 1,600 rpm.
4. Incubate the plate in a thermal cycler pre-heated to 30 °C (conditions: 30 min at 30 °C with a heated lid to 100 °C) for 30 min.
5. Vortex the AMPure XP Beads to resuspend the beads. Add 160 μL of AMPure XP Beads to each well of a new 0.3 mL 96-well microplate with round bottom wells using a multi-channel pipette for further bead purification.
6. Remove the seal from the Insert-Modification Plate after the completed incubation (**step 4**), and transfer all samples to the new Insert Modification Plate containing the AMPure XP beads. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
7. Incubate the Insert-Modification Plate at room temperature for 15 min.
8. Place the Insert-Modification Plate on a magnetic stand at room temperature for 5 min, and allow the magnetic beads to form a pellet on the sides of the wells.
9. Slowly aspirate and discard 127.5 μL of supernatant from each well.
10. Repeat **step 9** once.
11. With the plate remaining on the magnetic stand, add 200 μL of 80 % ethanol to each well without disturbing the beads.
12. Incubate the plate at room temperature for 30 s; then, gently aspirate and discard the supernatant from the wells.
13. Repeat **steps 11** and **12** for a total of two ethanol washes.

14. Leave the plate on the magnetic stand for 15 min to allow the ethanol to evaporate completely.
15. Add 17.5 μL of Resuspension Buffer to resuspend the pellet in the wells.
16. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm, and incubate the plate for 2 min at room temperature.
17. Place the Insert-Modification Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form pellets on the sides of the wells.
18. Transfer 15 μL of the supernatant from the Insert-Modification Plate to a new 0.3 mL 96-well microplate labeled Adapter-Ligation Plate.

3.1.5 Addition of Adenine Bases

1. Add 2.5 μL of Resuspension Buffer to each well of the Adapter-Ligation Plate.
2. Add 12.5 μL of A-Tailing Mix to each well of the Adapter-Ligation Plate.
3. Seal the plate with the adhesive seal, and mix it for 1 min in a MixMate vortex at 1,600 rpm.
4. Incubate the plate in a thermal cycler pre-heated to 37 °C (conditions: 30 min at 37 °C with a heated lid at 100 °C) for 30 min.
5. Remove the Adapter-Ligation Plate from the thermal cycler, and keep it on ice. Immediately proceed to the ligation reaction.

3.1.6 Adapter Ligation

1. Add 2.5 μL of Resuspension Buffer to each well of the Adaptor-Ligation Plate.
2. Add 2.5 μL of Ligation Mix to each well of the Adaptor-Ligation Plate.
3. Add 2.5 μL of the appropriate RNA-Adaptor Index to each well of the Adaptor-Ligation Plate.
4. Seal the plate with the adhesive seal, and mix it for 1 min in a MixMate vortex at 1,600 rpm.
5. Incubate the plate in a thermal cycler pre-heated to 30 °C for 10 min at 30 °C with a heated lid at 100 °C.
6. Add 5 μL of Stop-Ligation Buffer to each well of the Adaptor-Ligation Plate to stop the ligation reaction.
7. Vortex the AMPure XP Beads to resuspend the beads. Add 42 μL of AMPure XP Beads to each well of the new 0.3 mL 96-well microplate with round bottom wells labeled the Adaptor-Ligation Clean-Up Plate using a multichannel pipette.

8. Transfer all of the samples from the Adaptor-Ligation Plate (after **step 7**) to the Adaptor-Ligation Clean-Up Plate, and mix it for 1 min in a MixMate vortex at 1,000 rpm.
9. Incubate the Adaptor-Ligation Clean-Up Plate at room temperature for 15 min.
10. Place the Adaptor-Ligation Clean-Up Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form a pellet on the sides of the wells.
11. Slowly aspirate and discard 79.5 μL of supernatant from each well.
12. With the plate remaining on the magnetic stand, add 200 μL of 80 % ethanol to each well without disturbing the beads.
13. Incubate the plate at room temperature for 30 s; then, aspirate and discard the supernatant from the wells.
14. Repeat **steps 12** and **13** for a total of two ethanol washes.
15. Leave the plate on the magnetic stand for 15 min at room temperature to allow the ethanol to evaporate completely.
16. Add 52.5 μL of Resuspension Buffer to resuspend the pellets in each well.
17. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
18. Incubate the Adaptor-Ligation Clean-Up Plate at room temperature for another 2 min.
19. Place the Adaptor-Ligation Clean-Up Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form a pellet on the sides of the wells.
20. Transfer 50 μL of the supernatant from the Adapter-Ligation Clean-Up Plate to a new 0.3 mL 96-well microplate with round bottom wells labeled the Adaptor Ligation Clean Up (2) Plate.
21. Vortex the AMPure XP Beads to resuspend the beads. Add 50 μL of AMPure XP Beads to each well of the Adaptor Ligation Clean Up (2) Plate containing the purified adapter ligation mix.
22. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
23. Incubate the plate for another 15 min at room temperature.
24. Place the Adapter-Ligation Clean-Up (2) Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form pellets on the sides of the wells.
25. Aspirate and discard 95 μL of supernatant from each well.
26. With the plate remaining on the magnetic stand, add 200 μL of 80 % ethanol to each well without disturbing the beads.
27. Incubate the plate at room temperature for 30 s; then, aspirate and discard the supernatant from the wells.

28. Repeat **steps 26** and **27** for a total of two ethanol washes.
29. Leave the plate on the magnetic stand for 15 min at room temperature to allow the ethanol to evaporate completely.
30. Add 22.5 μL of Resuspension Buffer to resuspend the pellet in each well, and mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
31. Incubate the plate at room temperature for another 2 min.
32. Place the Adapter-Ligation Clean-Up (2) Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form pellets on the sides of the wells.
33. Transfer 20 μL of the supernatant from the Adapter-Ligation Clean-Up (2) Plate to a new 0.3 mL 96-well microplate labeled PCR Plate.

3.1.7 Amplification of cDNA Templates (See Note 6)

1. Let the AMPure XP beads to come to room temperature.
2. Add 5 μL of PCR Primer Cocktail to each well of the PCR plate.
3. Add 25 μL of PCR Master Mix to each well of the PCR plate.
4. Seal the plate with the adhesive seal and mix it for 1 min in a MixMate vortex at 1,600 rpm.
5. Place the cDNA Plate into the thermal cycler using the following program: (1) pre-heat lid and set to 100 °C, (2) incubate at 98 °C for 30 s, (3) incubate for ten cycles at 98 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s, (4) incubate at 72 °C for 5 min, and (5) hold at 10 °C indefinitely (*see* **Note 7**).
6. Vortex the AMPure XP Beads to resuspend the beads. Add 50 μL of AMPure XP Beads to each well of a new 0.3 mL 96-well microplate with round bottom wells labeled PCR Clean-Up Plate.
7. Transfer the contents of each well from the PCR Plate to the corresponding wells of the PCR Clean-Up Plate containing 50 μL AMPure XP beads.
8. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
9. Incubate the PCR Clean-Up Plate at room temperature for 15 min.
10. Place the PCR Clean-Up Plate on a magnetic stand at room temperature for 5 min to allow the magnetic beads to form pellets on the sides of the wells.
11. Aspirate and discard 95 μL of the supernatant from each well.
12. With the plate remaining on the magnetic stand, add 200 μL of 80 % ethanol to each well without disturbing the beads.
13. Incubate the plate at room temperature for 30 s; then, aspirate and discard the supernatant from the wells.

14. Repeat **steps 12** and **13** for a total of two ethanol washes.
15. Leave the plate on the magnetic stand for 15 min at room temperature to allow the ethanol to evaporate completely.
16. Add 40 μL of Resuspension Buffer to resuspend the pellets in each well.
17. Mix the plate for 1 min in a MixMate vortex at 1,000 rpm.
18. Incubate the PCR Clean-Up Plate at room temperature for 2 min.
19. Place the PCR Clean-Up Plate on the magnetic stand at room temperature for 5 min to allow the magnetic beads to form pellets on the sides of the wells.
20. Transfer 38 μL of the supernatant from the PCR Clean-Up Plate to a new 0.3 mL 96-well microplate labeled Target-Sample Plate.

3.1.8 Library Quality Assessment

1. Run 1 μL of the cDNA library sample (from **step 20** in Subheading 3.1.7) on an Agilent Bioanalyzer High Sensitivity DNA chip.
2. The expected average fragment length is approximately 260 bp.

3.2 Roche 454 Sequencing

This protocol has been prepared according to procedures described in cDNA Rapid Library Preparation Method Manual, GS FLX+ Series—XL+ (May 2011) from Roche [21].

3.2.1 Fragmentation of RNA

1. Add Molecular Biology Grade Water to a 200 μL PCR tube containing 200 ng of sample RNA to a final volume of 19 μL.
2. For comparison with fragmented RNA in a later step (**step 14**), transfer 1 μL of the above solution to a new 200 μL tube and add 2 μL of Molecular Biology Grade Water.
3. Add 2 μL of RNA Fragmentation Solution to the remaining 18 μL solution containing the sample RNA, vortex and centrifuge for 2 s.
4. Place the sample into a thermocycler that has been preheated to 70 °C and heat at 70 °C for 30 s, with the heated lid in place.
5. Immediately transfer the tube to ice.
6. Add 2 μL of 0.5 M EDTA pH 8.0 and 28 μL of 10 mM Tris–HCl, pH 7.5 to the tube; vortex and then centrifuge the sample for 2 s.
7. Add 80 μL of RNAClean reagents, which contains SPRI beads; mix well by pipetting up and down ten times; and incubate at room temperature for 10 min.
8. Place the tube onto a 96 ring magnetic particle concentrator (MPC).

9. When the beads in the tube have pelleted on the side of the tube, discard all of the supernatant without disturbing the beads.
10. Wash the beads three times using 200 μL of 70 % ethanol each time, followed by complete remove of the ethanol.
11. Return the tube to the MPC, uncap the tube, and allow the ethanol to evaporate at room temperature for 3 min.
12. Remove the tube from the MPC, add 19 μL of 10 mM Tris–HCl, pH 7.5, vortex for 20 s, and spin for another 2 s.
13. Return the tube to the MPC; when beads have pelleted on the side of tube, transfer the supernatant containing the RNA to a new 200 μL PCR tube.
14. Place the tube on ice, pipette 1 μL to a new 200 μL PCR tube, and add 2 μL of Molecular Biology Grade Water. Compare the sample in this tube with that from **step 2** in a Bioanalyzer; use 1 μL from each tube to confirm the successful fragmentation of the sample RNA.
15. Proceed with the remaining ~17 μL sample.

3.2.2 Synthesis of First Strand cDNA

1. Add 4 μL of 400 μM Roche Primer "random" to the sample from **step 15**, vortex for 10 s, and then spin for 2 s.
2. Incubate the tube for 10 min at 70 °C; then, leave on ice for 2 min.
3. Add the following to the tube on ice, making the total volume to 40 μL: 8 μL vial 1; 4 μL vial 3; 4 μL vial 7; 1 μL vial 4; 2 μL vial 2.
4. Vortex for 2 s; then, spin for 2 s.
5. Incubate the tube at 25 °C for 10 min, followed by 42 °C for 60 min. Then, transfer onto ice.

3.2.3 Synthesis of Second Strand cDNA

1. Add the following to the tube, making the total volume 150 μL: 30 μL of vial 9; 72 μL of vial 12; 1.5 μL of vial 7; 6.5 μL of vial 10.
2. Vortex for 5 s, and spin for 2 s.
3. Incubate at 16 °C for 2 h.
4. Add 20 μL of vial 11, and vortex for 5 s.
5. Incubate at 16 °C for 5 min.
6. Add 17 μL of 0.2 M EDTA, pH 8, vortex for 5 s, and then spin for 2 s.

3.2.4 Double-Stranded cDNA Purification

1. Transfer the sample containing cDNA to a new 1.7 mL tube.
2. Add 300 μL AMPure beads.
3. Vortex for 10 s followed by incubation at room temperature for 5 min.

4. Place the tube onto an MPC.
5. When the beads in the tube have pelleted on the side of the tube, discard the supernatant without disturbing the beads.
6. Wash the beads three times using 800 μL of 70 % ethanol each time, followed by complete removal of the ethanol.
7. Return the tube to the MPC, uncap the tube, and allow the ethanol to evaporate at room temperature for 3 min.
8. Remove the tube from the MPC, add 16 μL of 10 mM Tris–HCl, pH 7.5, vortex for 20 s, and spin for another 2 s.
9. Return the tube to the MPC. When beads have pelleted on the side of tube, transfer the supernatant containing the double-stranded cDNA to a new 200 μL PCR tube and leave on ice.

3.2.5 End Repair

1. To make the End Repair mixture, combine the following reagents from the Rapid Library Prep Kit in a 1.7 mL microcentrifuge tube: 2.5 μL of RL 10× PNK buffer; 2.5 μL of RL ATP; 1 μL of RL dNTP; 1 μL of RL T4 Polymerase; 1 μL of RL PNK; 1 μL of RL Taq Polymerase.
2. Mix well by pipetting up and down; then, place 9 μL in the tube containing cDNA sample.
3. Vortex for 5 s, and then spin for 5 s.
4. Place the tube containing cDNA into a thermocycler, and run the following program: 25 °C for 20 min, 72 °C for 20 min and hold at 4 °C indefinitely.

3.2.6 AMPure Bead Preparation

1. Vortex the bottle containing AMPure beads for 20 s or until the beads are fully resuspended.
2. Pipette 125 μL of AMPure beads into a 1.7 mL microcentrifuge tube.
3. Place the tube on an MPC.
4. When the beads in the tube have pelleted on the side of the tube, discard the supernatant without disturbing the beads.
5. Add 73 μL of TE Buffer to the tube, and vortex for 5 s.
6. Add 500 μL of Sizing Solution to the tube; vortex for 5 s, and then spin for 2 s.
7. Leave the tube on ice.

3.2.7 Adaptor Ligation

1. After the last step of End Repair is completed, add 1 μL of RL Adaptor or RL MID Adaptor to the tube.
2. Add 1 μL of RL Ligase to the tube.
3. Vortex for 5 s, and then spin for 5 s.
4. Incubate at 25 °C for 10 min.

3.2.8 Removal of Small Fragments

1. Add the adaptor-ligated cDNA sample into the tube containing AMPure beads that has been on ice since **step 7** in Subheading 3.2.6.
2. Incubate at room temperature for 5 min.
3. Place the tube on an MPC.
4. When the beads in the tube have pelleted on the side of the tube, discard the supernatant without disturbing the beads.
5. Perform the following two steps: (a) Add 100 μL of TE Buffer to the tube, and vortex for 5 s or (b) Add 500 μL of Sizing Solution, and vortex for 5 s.
6. Incubate at room temperature for 5 min.
7. Return the tube to the MPC. When the beads in the tube have pelleted on the side of the tube, discard the supernatant without disturbing the beads.
8. Repeat **steps 4** through 7 once.
9. With the tube remaining on the MPC, wash the beads three times using 1 mL of 70 % ethanol each time, followed by complete remove of the ethanol.
10. Uncap the tube, and allow the ethanol to evaporate at room temperature for 2 min.
11. Remove the tube from the MPC, add 53 μL of TE Buffer, vortex for 5 s, and spin for another 2 s.
12. Place the tube on the MPC; after the beads form a pellet on one side of the tube, transfer 50 μL of the supernatant containing the cDNA library to a new 1.7 mL tube.

3.2.9 Library Quality Assessment

Run 1 μL of the cDNA library sample (from **step 12** in Subheading 3.2.8) on an Agilent Bioanalyzer High Sensitivity DNA chip. The expected average fragment length should be between 600 bp and 1,200 bp, and less than 10 % of fragments should be shorter than 500 bp.

4 Notes

1. Although the recommended amount of total RNA input ranges from 0.1 to 1 μg, in our hands, only up to 1.5 μg of input material is usually used to avoid library over-amplification, which leads to increased duplication rates.
2. There are some significant differences in the mixing efficiency between different microplate shakers. The advantage of using the Eppendorf MixMate, which utilizes the unique two-dimensional Mix Control technology, ensures efficient mixing during the incubation without any need of further centrifugation of the microplate. Because most of microplate mixing is

performed without a seal it is recommended that a speed of only 1,000 rpm is used to prevent cross contamination.

3. For bead-based purification there are two types of magnetic stands available—Ambion (Cat No AM10050) and Beckman Coulter (Cat No A32782). Both of these magnetic stands allow up to 96 samples to be processed simultaneously; however, for a manual workflow using the Ambion stand in combination with a 0.3 mL 96-well microplate with round bottom wells (e.g., Eppendorf Cat No 0030 601 203) is our first choice. Due to the wider well diameter of both the Ambion stand and the 0.3 mL microplates it is much easier to handle the bead pellet and prevent any disruption of the bead pellet. However, this platform requires an additional sample transfer from the 0.2 mL 96-well microplate that is used for incubations to the 0.3 mL round bottom microplate that is used for bead purification. For an automated bead purification workflow, both 0.2 mL and 0.3 mL microplates can be used.
4. For best results use freshly prepared 80 % ethanol.
5. Removal of all ethanol is important; before proceeding to the next step, always manually inspect the plate. The time required for the ethanol to evaporate completely depends on the ambient temperature of the lab. Avoid over-drying the beads (cracks in the pellet), which can affect the final yield (the captured DNA will not be completely released from over-dried beads).
6. Preparation of the PCR reaction should be carried out in the pre-PCR area; once amplification has been completed, the amplified samples should always remain in the post-PCR area. This is to reduce the chance of PCR product contamination. It is important that the lab is set up in a way that prevents any chance of back contamination.
7. Reducing the number of PCR cycles to 10 instead of 15, which is recommended by Illumina, results in a reduction of the number of duplicate reads in the transcriptome data.

References

1. Velculescu VE, Madden SL, Zhang L et al (1999) Analysis of human transcriptomes. Nat Genet 23:387–388
2. Bertone P, Stolc V, Royce TE et al (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246
3. Marguerat S, Wilhelm BT, Bähler J (2008) Next-generation sequencing: applications beyond genomes. Biochem Soc Trans 36: 1091–1096
4. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264
5. Sultan M, Schulz MH, Richard H et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960
6. Rogaeva E (2002) The solved and unsolved mysteries of the genetics of early-onset alzheimer's disease. Neuromolecular Med 2: 1–10

7. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174
8. Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci 9:768–778
9. Guerreiro RJ, Gustafson DR, Hardy J (2012) The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. Neurobiol Aging 33:437–456
10. Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet 17:100–107
11. Gray MW (2012) Evolutionary origin of RNA editing. Biochemistry 51:5235–5242
12. Mattick JS (2009) The genetic signatures of noncoding RNAs. PLoS Genet 5:e1000459
13. Sutherland GT, Janitz M, Kril JJ (2011) Understanding the pathogenesis of Alzheimer's disease: will RNA-Seq realize the promise of transcriptomics? J Neurochem 116:937–946
14. Courtney E, Kornfeld S, Janitz K, Janitz M (2010) Transcriptome profiling in neurodegenerative disease. J Neurosci Methods 193: 189–202
15. Mills JD, Janitz M (2012) Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. Neurobiol Aging 33:1012.e11–1012.e24
16. Kim KH, Moon M, Yu SB et al (2012) RNA-Seq analysis of frontal cortex and cerebellum from 5XFAD mice at early stage of disease pathology. J Alzheimers Dis 29:793–808
17. Mills JD, Nalpathamkalam T, Jacobs HI et al (2013) RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. Neurosci Lett 536:90–95
18. Twine NA, Janitz K, Wilkins MR, Janitz M (2011) Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 6:e16266
19. Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
20. Illumina TruSeq RNA sample preparation v2 guide. http://support.illumina.com/sequencing/sequencing_kits/truseq_rna_sample_prep_kit_v2.ilmn
21. Roche cDNA rapid library preparation method manual, GS FLX+ series—XL+ (May 2011). http://454.com/downloads/my454/documentation/gs-flx-plus/Rapid-Library-Preparation-Method-Manual_XLPlus_May2011.pdf

Chapter 21

Systems Biology Approaches to the Study of Biological Networks Underlying Alzheimer's Disease: Role of miRNAs

Wera Roth, David Hecker, and Eugenio Fava

Abstract

MicroRNAs (miRNAs) are emerging as significant regulators of mRNA complexity in the human central nervous system (CNS) thereby controlling distinct gene expression profiles in a spatio-temporal manner during development, neuronal plasticity, aging and (age-related) neurodegeneration, including Alzheimer's disease (AD). Increasing effort is expended towards dissecting and deciphering the molecular and genetic mechanisms of neurobiological and pathological functions of these brain-enriched miRNAs. Along these lines, recent data pinpoint distinct miRNAs and miRNA networks being linked to APP splicing, processing and Aβ pathology (Lukiw et al., Front Genet 3:327, 2013), and furthermore, to the regulation of tau and its cellular subnetworks (Lau et al., EMBO Mol Med 5:1613, 2013), altogether underlying the onset and propagation of Alzheimer's disease. MicroRNA profiling studies in Alzheimer's disease suffer from poor consensus which is an acknowledged concern in the field, and constitutes one of the current technical challenges. Hence, a strong demand for experimental and computational systems biology approaches arises, to incorporate and integrate distinct levels of information and scientific knowledge into a complex system of miRNA networks in the context of the transcriptome, proteome and metabolome in a given cellular environment. Here, we will discuss the state-of-the-art technologies and computational approaches on hand that may lead to a deeper understanding of the complex biological networks underlying the pathogenesis of Alzheimer's disease.

Key words miRNA, Systems biology, Alzheimer's disease, Neurodegeneration, Computational biology

1 Introduction

Ribonucleic acid (RNA) molecules have been known since the middle of the twentieth century when Caspersson and Schultz described the presence of pentose nucleotides in the cytoplasm [1, 2]. Nevertheless, following the discovery of the DNA structure [3] and the formulation of the central dogma [4] the RNA has been for long time mainly considered to be the mere intermediary between DNA and protein synthesis, although it was clear that non-coding RNAs (ncRNAs), i.e. RNAs that are not translated into a protein, existed. Examples of ncRNAs are the transfer RNAs (tRNAs) and

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_21,

the ribosomal RNA (rRNAs) that were also discovered in the second half of the twentieth century [5–7].

In recent years the use of a systems approach in the study of the genome (e.g. high-throughput sequencing technologies) has put in evidence that the mammalian transcriptome is much more complex than previously thought. In particular, it emerged that the transcriptome includes a large number of ncRNAs (for a detailed review refer to [8]). The ncRNAs are generally divided according their length in: (1) small ncRNAs (sncRNAs) with length < 100 bases, and (2) long ncRNAs (lncRNAs) with length > 100 bases. Among the sncRNAs, the micro RNAs (miRNAs) have emerged as an important class of regulatory RNAs involved in development and disease.

Here we discuss the role of miRNAs in health and disease with a particular focus on Alzheimer's disease (AD). We also discuss the approach to study miRNAs using a systems biology approach, including the techniques available to detect miRNAs in a systematic way and the computational analysis level to understand the miRNAs in complex networks and mathematical modeling.

1.1 microRNAs (miRNAs)

Mature miRNAs are ca. 20–30 nucleotides long and originate from double-stranded RNA precursors derived from transcripts that fold back on themselves forming characteristic hairpin structures [9]. MicroRNAs are produced endogenously by gene transcription of both sense and antisense DNA as well as from pseudogenes and inverted repeats. The miRNA biogenesis machinery has been extensively investigated in recent years. Detailed information on this can be found in following reviews [10, 11]. Briefly, miRNAs are mainly transcribed by either RNA polymerase II or RNA polymerase III as primary miRNAs (pri-miRNAs) presenting a characteristic stem loop structure. The pri-miRNAs are localized in the nucleus and are processed in pre-miRNAs by a protein complex, the microprocessor, containing the RNAse III DROSHA [12, 13] (Fig. 1). The pre-miRNAs are ca. 65–70 nucleotides long containing the stem loop, and are exported to the cytosol by a nuclear export complex composed by exportin-5/RanGTP [14]. Once in the cytoplasm, the pre-miRNAs undergo final maturation by the DICER RNA endonuclease that cut the pre-miRNAs to result in 20–25 base pairs mature miRNAs (Fig. 1). The mature miRNAs are recruited into the RISC complex containing key proteins of the Argonaute family and the activated RISC complex which participates in post-transcriptional regulation of mRNAs. In the early days, miRNAs were thought to act mainly as negative regulators of gene expression and exert their effects by binding to regions within the 3′ UTR of their target protein-coding mRNAs in a sequence dependent manner [15–18]. The active degradation of miRNAs appears to happen in the P-bodies [19, 20] that function as negative regulators of mRNA by

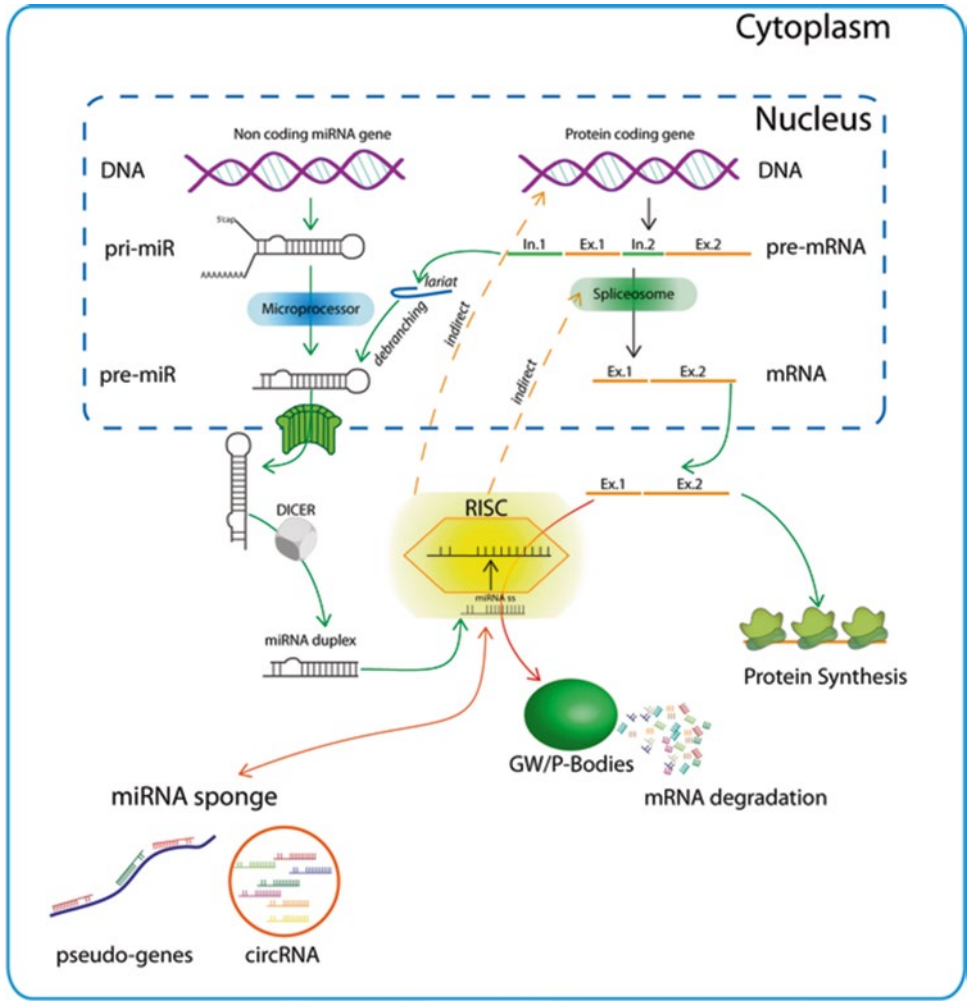


Fig. 1 The biogenesis and function of miRNAs. MicroRNAs (miRNAs) can be generated in the nucleus by the canonical pathway, dependent on RNases II and III and the Microprocessor (including Drosha and DGCR8). Alternatively, miRNAs can also be generated in a microprocessor-independent manner. In this case, short intronic hairpins (mirtrons) are excised by splicing and linearized by lariat debranching. Once the pri-miRs are processed, the resulting pre-miRs are exported to the cytoplasm through exportin 5. In the cytoplasm, the pre-miR is processed by DICER and transformed into a mature miRNA duplex. The duplexes are loaded into the RNA-induced silencing complex (RISC) were one strand is selected. The loaded RISC is the active component executing the messenger RNA (mRNA) translation regulation by binding 3′ UTR region of mRNAs in a sequence specific manner. Messenger RNAs can be sequestered in GW/P-bodies were they are retained or degraded. MicroRNAs themselves also undergo regulation by the miRNA sponge formed by pseudogenes or circular RNAs

sequestering the mRNA in translation inactive compartments. However, recent data reveals that miRNA regulation entails a far more complex system of post-transcriptional control than initially appreciated. In fact, there is a growing evidence that indicates that miRNAs might have alternative ways to control protein synthesis, including positive regulation of protein synthesis by coordinated

release of mRNA, or the control of the mRNA splicing forms, which could have an important role in AD (*see* below).

MicroRNAs are regulated themselves, and recently the concept of the miRNA sponge was postulated as a possible mechanism of miRNA regulation. The sponge concept was introduced in mammalian cells and initially was applied to the use of artificial decoy RNA molecules designed to capture and hence generate loss-of-function phenotypes for given miRNA families [21]. In 2010 it was shown that in mammalian cells pseudogenes could work as miRNA sponges [22], confirming the role of miRNA sponges as a natural occurring mechanism to regulate miRNA activity. The concept of miRNA sponges has been recently expanded to circular RNAs (circRNAs), confirmed by two groups independently [23, 24]. These data show an additional level of miRNA regulation that could play a main role as regulators of mRNA activity and could have implications in pathological conditions. Interestingly, circRNAs have been identified to be expressed in diverse cell types including brain cells, where hundreds of circRNAs have been predicted [24], and could represent an interesting target for future studies on their role in AD.

1.2 miRNAs in Alzheimer's Disease

Alzheimer's disease (AD) is the most common form of dementia accounting for more than 60 % of dementia cases. The target organ of AD is the brain and the disease is characterized by a progressive degenerative condition leading to loss of memory and intellectual abilities. The AD onset correlates with age, which is considered a risk factor. AD is characterized by the presence of amyloid plaques, neurofibrillary tangles, synaptic loss and selective neuronal death [25]. Following the discovery of miRNAs there has been an active progress in understanding the role of miRNAs in the onset and development of AD.

Seven years after the first demonstration of the existence of human miRNAs [26], Lukiw performed the first study investigating miRNA differential expression in AD, where AD-specific changes in the miRNA regulatory system were shown for the first time [27]. Recent years have seen an increasing amount of studies linking miRNAs to AD, and evidence is mounting that miRNAs play an important role in AD. A number of brain specific miRNAs (e.g. miR-124, miR-200 family, miR-29 family, miR-181 family, etc.) have been identified to be deregulated in AD patients, animal models and in vitro cell-based assays (for reviews *see* refs. 28–35). However, it should be stressed that miRNA studies in AD patients so far revealed very little uniformity in miRNA changes [32] and contradictory results in some cases. This indicates the extreme complexity of the miRNA networks controlling cell/tissue identity, and the inherent difficulties to dissect their role in physiology and pathology. From here, two main issues emerge: first, the need to better understand the mechanisms involved in the function of

miRNAs at a "system" level and second, the need to increase the quality of tools to study miRNAs (discussed below) to overcome limitations hampering present studies.

Studies on the role of miRNA deregulation in AD so far have focused mainly on well known target genes or pathways underlying AD pathology, such as amyloid precursor protein (APP), ß-secretase (BACE), presenilin 1 (PSEN1), tau and others, and the role of particular miRNAs modulating these pathways, which are extensively investigated. These approaches however, may hamper the discovery of alternative complimentary miRNA pathways. The miRNA networks are extremely complex, and the validation of the role of miRNA deregulation in AD calls for new holistic, systems biology approaches, including network analysis and mathematical modeling.

As an example of the complexity of miRNAs networks, recent discoveries show that miRNA-dependent mRNA deregulation is probably not the only mechanism by which miRNAs could be involved in the onset and progression of AD. In fact, recent studies indicate that miRNAs could have an active role in the regulation of mRNA splicing. In particular, it has been demonstrated that splicing of APP and tau mRNAs, two main players in AD, is influenced by miRNAs [36, 37]. Thus, miR-124 elevation reduces the content of polypyrimidine tract binding protein 1 (PTBP1) resulting in the shift of APP mRNA splicing towards isoforms containing predominantly exon 7 and 8 (non neuronal isoforms) and a reduction of the mRNA containing exon 15 (prevalent neuronal isoform) [36]. Similarly, miRNAs are involved in control of tau mRNA splicing [37] and tau phosphorylation [38], with miR-132 identified as the factor regulating PTBP2 and hence controlling tau exon 10 splicing in neurons [37].

The role of miRNAs in splicing opens the way to possible additional roles of miRNAs in network and cell identity control in AD. It could be speculated that early alteration of the miRNA networks can participate in the early phase of AD by shifting the protein isoforms content, affecting proteome identity and cell functional status. Another interesting possible interaction is the role of miRNAs as mediators between cellular protein homeostasis and translational control. Whether this or other miRNA functions may act as mediators affecting the regulation of APP splicing, Aβ production, tau missplicing and hyperphosphorylation remains an exciting possibility and a challenge for the systems biology community.

2 Methods

2.1 miRNAs Detection Methods

Increasing interest towards understanding the role of miRNAs in tissue development, homeostasis and cell identity has led to rapid advances in technologies measuring miRNAs expression in recent years. Accurate quantitative and comparative assessment of

miRNA expression profiles allows the understanding of human disease mechanisms and will help the discovery of new drug targets.

Quantification of miRNAs is technically feasible, though still challenging, due to their small size as mature forms, high sequence homology, 5′ and 3′ end polymorphisms, and high dynamic range of expression. Genome-wide miRNA expression profiling to investigate global differential expression can be based on three principal technologies: (1) miRNA microarray hybridization [39, 40], (2) reverse transcription quantitative real-time PCR (RT-qPCR) [41, 42] using TaqMan™ low-density arrays (TDLAs) [43] and, (3) massive parallel next-generation sequencing (NGS) approaches of miRNA-specific cDNA libraries/small RNA libraries [44, 45]. All of them are available by several commercial suppliers. The method of choice depends on the type of application, final goal of the study and finally, on overall costs, precision and accuracy of the methodology and sample quantity [46, 47].

2.2 Technical Considerations and Challenges in the Field

Although miRNAs represent a quite abundant class of transcripts, overall miRNAs constitute only a small fraction (0.01 %) of the total cellular RNA [48]. The miRNA expression levels vary among species, tissues and cell types and are regulated in a temporal and spatial manner within single cells and tissues. Also, low abundant miRNAs may be missed in technologies with low sensitivity, such as hybridization-based techniques (e.g. microarray analysis).

The active miRNA species in vivo are the mature miRNAs. These exert their actions via base pairing of their seed regions to the 3′ UTR of their target mRNAs. Consequently, it is desirable to detect and quantify mature miRNAs rather than the population of miRNA precursors (i.e. pri- and pre-miRNAs species). Alternatively, when the analysis of the entire subpopulations of miRNA species in a given cell type or tissue is needed, this allows to address the whole regulation of miRNA biogenesis at the level of transcription, nuclear export and Dicer activity, resulting in the final mature forms.

As a matter of principle, the technology to be used has to be carefully selected depending on the final objective, type of data output and content of information. Methodology-based limitations of each technique have to be considered as well (*see* below).

Mature miRNAs arise from both the 5′ and 3′ arms of the hairpin precursor, an expression analysis of a pre-miRNA population for instance does not allow to assign the predominant mature miRNA products in a given cell/tissue state.

MicroRNAs are grouped into families of nearly identical isoforms according to their sequence homology [49]. Mature family members show either nearly identical sequences differing by 1–3 nucleotides (nt) only (e.g. let-7b and let-7c), or they may derive from slightly different precursor genes often localized on different chromosomes, resulting in identical mature miRNA sequences (e.g. let-7a-1 and let-7a-2) [50]. These (nearly) identical sequences

mean technical limitations need to be faced concerning the specificity of pre-designed mature miRNA probes in microarrays and TaqMan® probe-based RT-qPCR assays. The fact is that identical mature miRNAs produced by different genetic loci can only be differentiated by assessing their respective miRNA precursors. Toward this end, distinct/specific TaqMan® MGB probes can be designed that anneal to the loop portion of the miRNA precursor allowing the isoform-specific detection and quantification of mature miRNAs [50].

Finally, detected miRNAs may vary (from their reference sequence) due to the variability in 3′ and 5′ processing, or due to (enzymatic) modifications such as single-nucleotide 3′ extensions [51, 52]. In these cases, next-generation sequencing might be considered as the technology of choice, to assess the biological relevance and functional significance of these so-called isomiRs (*see* below).

2.3 miRNA Microarrays

Microarray technology was first applied to miRNA studies in 2004, in which 40mer miRNA-specific oligonucleotide probes were designed to distinct miRNA precursor and/or mature forms and spotted onto a biochip/solid support. Respective 5′ biotin-labeled complementary cDNA targets were generated by reverse transcription of sample RNAs under investigation (including its miRNAs) using a biotinylated oligonucleotide primer followed by array hybridization, staining and signal detection [53, 54]. This technology has been further extended to a high-throughput miRNA microarray expression-profiling platform suitable for global analysis of miRNA expression [55].

Several microarray methodologies have been developed during the last years [56–60]. Main differences between miRNA microarray technologies lay within the oligonucleotide probe design, solid support probe immobilization chemistry, target labeling of the sample and the signal detection method on the chip [39, 40, 55]. However, the common theme used in the field is the fluorescent labeling of the target in a biological RNA sample, followed by hybridization to capture probes on the array. Methodologies based on direct labeling techniques of the target miRNAs, either by chemical modifications or via T4 RNA ligase-mediated dinucleotide coupling, allow straight measurements of mature miRNA targets without the supplementation of an amplification step for label incorporation, which is prone to certain biases. Thus, manipulation of the sample remains minimal, although a relative higher amount of input sample is required [55, 56, 60, 61]. Technical variations of target labeling include a linear amplification step after adapter ligation, in order to incorporate the fluorescent label either in the sense or antisense strands of miRNA-derived target cDNAs. Since the small RNA sample is amplified before array hybridization, this technique is suitable for samples originated from low starting material [58].

Although microarrays are suitable methods for high-throughput analysis of global miRNA expression, certain disadvantages and limitations of the microarray technology should be considered: (1) Still, there is the need for a relatively high concentration of the target within the input sample, to mediate efficient hybridization and signal generation, (2) the detection of low abundant targets suffers from poor sensitivity and, (3) compared to RT-qPCR and NGS technologies, microarrays show a lower dynamic range of detection of differential expression, which points towards the need for post-array validation using more sensitive assays, such as RT-qPCR [62].

Another challenge of miRNA profiling using DNA-oligonucleotide-based microarrays is the difficulty to optimize and ensure homogeneous probe-target hybridization conditions due to the small size of target miRNAs, with distinct melting temperatures (T_m) over the array. Toward this end, locked nucleic acid (LNA)-modified capture probes have been developed allowing T_m-normalized hybridization conditions with high affinity and specificity, and have been successfully applied for instance in the miChip microarray platform [61, 63]. LNAs are synthetic bicyclic high-affinity RNA/DNA analogs in which the furanose ring in the sugar-phosphate backbone is chemically locked in an N-type (C3′-endo) conformation. LNAs mediate increased thermostability when introduced into oligonucleotides and have been successfully used in Northern blot and in situ hybridization assays [64, 65]. The miChip platform has been shown to monitor accurately and sensitively miRNA expression without prior need for RNA size selection or miRNA amplification [61, 63, 66].

In general, any application of technologies using probe-based detection methods, i.e. microarrays and RT-qPCR approaches, restricts the identification and quantitative assessment of miRNAs only to those previously discovered, sequenced and annotated. Thus, new yet unidentified miRNAs cannot be detected.

The majority of miRNA microarray analyses published in the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo) have been performed on commercial platforms provided by multiple suppliers [67]. Commercial platforms provide highly standardized experimental conditions and protocols allowing the comparative analysis of the relevant abundance of miRNAs between two states [67]. However, cross-platform comparison analyses are difficult, and the risk of platform withdrawal by vendors has to be considered during study design.

Of crucial importance, a quality control assessment is essential to prove cross-sample comparability. The performance of internal controls present on the microarray allows monitoring potential geographic biases present on the chip.

Finally, processing of raw data and the applied normalization strategy have a significant impact on the final results. Data normalization is essential for obtaining accurate and comparable results,

and the strategy to be applied needs to be carefully considered. For microarray data, global normalization methods based on the assumptions that (1) global distribution is expected to be similar for all samples and, (2) the number of up- and downregulated features is similar [67] are frequently used.

2.4 Reverse Transcription Quantitative Real Time PCR (RT-qPCR)

RT-qPCR-based miRNA expression analysis and quantification has been introduced since 2004 [41, 68]. It represents a sensitive approach of expression analysis for both mature and precursor miRNAs and can be extended to a medium-throughput approach by utilizing TaqMan® low-density expression array cards or other microfluidic assay systems [43, 69]. Here, the main emphasis is on the detection and relative quantification of miRNAs via TaqMan® assays. Other RT-qPCR-based technologies such as poly(A) tailing-based and direct RT-based SYBR-Green assays (including LNA-technology), or primer-extension PCRs are discussed in detailed articles elsewhere [42, 70–73].

Due to the hairpin structure of miRNA precursors and the short length of their mature forms, miRNAs are challenging to be specifically amplified and quantified by PCR. Briefly, in the first step, total RNA is reverse transcribed to cDNA using either random hexamer primers (generally to detect miRNA precursors) or miRNA-specific stem-loop primers (for the detection of mature miRNAs). In the subsequent quantitative real-time PCR, miRNA-specific amplification and quantification can be achieved by using gene-specific primers. To specifically amplify pre-miRNA specimen, forward and reverse primers for qPCR are designed to hybridize to the respective 5′ and 3′ parts of the stem portion of the miRNA hairpin. Hence, caution needs to be exercised for providing miRNA specificity due to close homology of distinct miRNA family member isoforms [50, 68, 74]. Yet TaqMan® minor groove binding (MGB) probes designed to the loop portion of the miRNA precursor have been shown to provide specificity and allow the differentiation between distinct miRNA isoform precursors in qPCR assays [50, 68, 74].

By using stem-loop primers during the RT reaction in conjunction with miRNA-specific TaqMan® probes during the subsequent qPCR step, mature miRNAs can be accurately and isoform-specifically quantified [41]. Stem-loop RT primers provide higher specificity for discriminating similar miRNAs, inhibit their hybridization to miRNA precursors via their double-stranded stem, give higher stability to the miRNA-DNA heteroduplexes therefore increasing RT efficiency, and finally, upon unfolding their stem-loop structure they add an extra downstream sequence resulting in a longer RT product, serving as proper templates during the qPCR step [74]. During the subsequent real-time TaqMan® assay, the miRNA-specific forward primer in conjunction with a distinct TaqMan® probe, provide high specificity and high sensitivity to the assay. Therefore, stem-loop RT-based TaqMan® miRNA assays are

widely considered as the gold-standard method for qPCR-based techniques providing a dynamic range of seven orders of magnitude [41, 42].

TaqMan® probe-based RT-qPCR assays have been adapted to a 384-well format [43, 74] in conjunction with an Applied Biosystems 7900HT Sequence Detection System, and even extended to the TaqMan® low-density array (TDLAs) assay format allowing medium-throughput analysis [43, 75]. Towards this end, this technology has been complemented by the introduction of an optional preamplification step into the workflow by the use of a specific set of PCR primers facilitating the detection of low-expressed miRNAs. The development of Megaplex™ RT and PreAmp primer pools (panel A and panel B) in conjunction with TaqMan® array cards allows the detection and quantification of up to 380 distinct miRNAs per panel, with high sensitivity and accuracy [43]. TaqMan® array cards are microfluidic devices containing 384 reaction chambers that are preloaded by the manufacturer with distinct panels of miRNA TaqMan® assays. Panel A allows the analysis of well-studied, broadly expressed miRNAs of interest to the scientific community (for both human and rodent), whereas panel B targets miRNAs not extensively characterized and known to be more narrowly expressed. The sample/TaqMan® master mixes are administered and loaded onto the TaqMan® array cards through fill-ports followed by subsequent centrifugation and completed by a sealing step which collapses the microfluidic channels [43]. The real-time qPCR assays for TaqMan® array cards are run on the ABI Prism® 7900HT sequence detection system (Applied Biosystems, USA) or on alternative compatible instruments [43]. This format for miRNAs expression and quantification analysis via RT-qPCR TaqMan® assays, available through Applied Biosystems, is reviewed elsewhere [62].

Using the OpenArray® real-time PCR platform allows even a further increase in sample throughput. Up to 3 samples per OpenArray® plate can be analyzed, performing up to 818 TaqMan® assays per sample [43].

Standard procedures and how to perform and publish real-time qPCR experiments have been compiled in the "Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)" in 2009 [76].

A critical issue after data acquisition is the data analysis workflow, which includes inspection of the raw data, evaluation of quality and robustness/reliability, data processing towards results, and finally, statistical analysis. Two main methods exist to present quantitative gene expression data, absolute and relative quantification. Absolute quantification is based on a standard curve derived from a reference sample and calculates the absolute quantity of a single nucleic acid target sequence within an unknown sample, expressed as copy number per cell [42, 77]. In relative quantification,

RT-qPCR data are processed relative to reference targets, i.e. internal or endogenous controls, which remain unaltered during experimental conditions. Due to variations in sample quantity, quality and variable PCR efficiency affecting absolute quantification, relative quantification is most widely used [42, 77]. However, several endogenous controls have to be considered and even a prior screening approach might be necessary in order to identify appropriate and suitable endogenous controls.

The most common method to present relative expression is the comparative Ct method, also known as the 2(-Delta Delta C(T)) or $2^{-\Delta\Delta Ct}$ method [78], which includes the assumptions that the efficiency of the PCR is close to 1 and the PCR efficiency of the target gene of interest is similar to the internal/endogenous control, and which presents data results as "fold-change", or differential expression [77]. Derivation of the formula, final equations towards data normalization and fold-change expression calculation, including statistical analysis, are described in detail in excellent articles [42, 77, 78] and in Applied Biosystems User Bulletins (e.g. http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040980.pdf).

In miRNA RT-qPCR assays, the accuracy of the final results expressed as "fold-change" or differential expression is largely dependent on proper data normalization. For global miRNA profiling only a few candidate reference miRNAs have been reported suitable for data normalization [79]. These reference miRNAs might be severely affected by the experimental system and the study design. Usually, other small non-coding RNAs are used for normalization, such as U6 small nuclear RNA (RNU6) or small nucleolar RNAs (e.g. small nucleolar RNA C/D box 87 (SNORD87)) [80]. Mestdagh and colleagues successfully introduced the mean expression value in a given sample to normalize high-throughput miRNA RT-qPCR data to the scientific community and provided a workflow for proper data normalization of global miRNA profiling [80].

2.5 Next-Generation Sequencing (NGS)

Next-generation sequencing (NGS) or "massive parallel sequencing" refer to the high-throughput DNA sequencing technologies available which, in marked contrast to automated Sanger sequencing, allow the parallel sequencing of large numbers of different input samples/DNA templates. NGS technology can also be used to sequence the RNA population, in order to analyze the transcriptome profile and its expression status of protein-coding genes or to investigate the population of small RNAs. Recently, NGS-based profiling techniques have been extended towards the analysis of global miRNA expression profiles [81]. NGS technologies allow a depth of sequencing that extends the dynamic range of miRNA expression analysis to up to six orders of magnitude. Hence, these technologies offer the potential to identify low abundant miRNAs

and to dissect subtle differences in miRNA expression between samples. Furthermore, deep sequencing approaches have a unique advantage, that is, to allow the identification of novel miRNAs. The first report of a massively parallel sequencing approach of miRNAs was published in 2008 surveying the global expression status of the miRnome in human embryonic stem cells [52]. In contrast to capillary sequencing, these technologies allow a high-throughput approach and thus facilitate in-depth global analyses of the miRnome of distinct cellular states or tissues. In depth overviews about the currently most frequently used platforms and sequencing technologies are provided in recent review articles [81–83]. In the context of global miRNA profiling, two main platforms are regularly used, the Illumina (Genome Analyzer or HiSeq2000) and the SOLiD platforms which use short-read sequencing-by-synthesis technologies based on either bridging amplification or emulsion PCR [81]. miRNA libraries are constructed out of total RNA or small RNA enriched RNA samples by passing through sequential steps of 3′ and 5′ adapter ligation reactions, first-strand reverse transcription for cDNA library generation followed by (optional) PCR amplification and size selection/purification of resulting DNA fragments [84]. A streamlined protocol for RNA-Seq towards miRNA expression analysis is provided by Illumina, i.e. the Illumina® TruSeq ®Small RNA Sample Preparation protocol (http://supportres.illumina.com/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqsmallrna/truseq-small-rna-sample-prep-guide-15004197-f.pdf). The advantage of this protocol is that barcode (so-called "index") sequences are added to the library molecules during the PCR amplification step and not via the adapter sequences, which was shown to result in a significant ligation bias [85]. Index sequences of the amplified library are sequenced separately in the same channel/region of the instrument. Later, obtained sequence reads can be computationally sorted based on their indices. This approach in which individual miRNA libraries can be marked by unique indices allows highly parallel sequencing of a large number of samples within the same lane in a time and cost saving manner.

The data analysis pipeline after the actual sequencing run is a critical issue. The initial base calling is usually conducted by proprietary software on the sequencing platform. Obtained reads derived from miRNA cDNA libraries are first processed by removing the 3′ adapter sequences which would interfere with proper sequence alignment and mapping to a reference database or genome. For read mapping of RNA-seq data to reference databases, numerous software packages have been developed which are continuously undergoing further improvement and which are applicable for miRNA sequence data, e.g. Bowtie and SOAP among others [81, 86]. Reference databases for miRNA mapping include miRBase [87], deepBase [88] and microRNA.org, among others (Table 1),

Table 1
Most relevant frequently utilized miRNA annotation databases

Name	Description	Website	References
miRBase	Primary online repository for all miRNA sequences and annotations miRNA target aggregation service provided	http://www.mirbase.org/	[89]
deepBase	Annotation and mining of miRNAs. Integrative, and versatile web graphical interface facilitating transcriptomic research and novel miRNA discovery	http://deepbase.sysu.edu.cn/	[88]
microRNA.org	Comprehensive resource of miRNA target predictions and expression profiles	http://www.microrna.org/	[152]
miRGen2.0	Integrated database of miRNA gene transcripts, transcription factor binding sites, miRNA expression profiles and single nucleotide polymorphisms associated with miRNA	http://diana.imis.athena-innovation.gr/DianaTools/index.php	[153]

with miRBase representing the state of the art online repository of all miRNA sequences and annotations, subjected to regular updates and releases [89]. Mapping against other non-coding RNA databases obtained from other sources, such as GenBank, ENSEMBL and the UCSF genome browser constitute an alternative strategy to annotate miRNA deep sequencing data [90, 91], and whole genome assemblies may also being used as reference regarding miRNA expression profiling, towards identification of novel miRNAs. Along with this, and fostered by the increasing volume of miRNA de novo deep-sequencing data, new miRNA prediction tools have been recently developed, such as miRDeep/miRDeep2 [92], miRanalyzer [93], miRDeep* [94], miRTRAP [95] and MIReNA [81, 96, 97] (Table 2). These tools facilitate the discovery and expression profiling analysis of novel miRNAs from NGS data.

In order to gain insight into molecular mechanisms of action, a relevant step is the identification of miRNAs with differential expression among biological samples. Towards this end, normalization of NGS data derived from different samples is a critical issue. Different normalization strategies based on distinct features and assumptions are used in the field [98], such as RPKM (reads per kilobase of exon model per million mapped reads) [99], TMM (trimmed mean of M-values) [100], or upper quartile normalization [101]. To evaluate and quantify differential miRNAs expression between biological samples from preprocessed NGS data [98] many methodologies and software tools have been developed.

Table 2
Selected computational tools for miRNA prediction and data analysis [97, 156]

Name	Description	Website	References
miRDeep	Use of a probabilistic model of miRNA biogenesis to predict novel miRNAs	https://www.mdc-berlin.de/8550208/de/research/research_teams/systems_biology_of_gene_regulatory_elements/projects/miRDeep	[92]
miRanalyzer	Free web-server tool for processing small-RNA data. Prediction of previously unknown microRNAs by data mining	http://bioinfo5.ugr.es/miRanalyzer/miRanalyzer.php	[93]
miRDeep*	Integrated application tool for miRNA identification from RNA-Seq data. Identification of precursor miRNAs. Expression level analysis	http://www.australianprostatecentre.org/research/software/mirdeep-star	[94]
MIReNA	Validation of pre-miRNAs with high sensitivity and specificity. Detection of new miRNAs by homology from known miRNAs	http://www.lgm.upmc.fr/mirena/index.html	[96]
miRExpress	Generating miRNA expression profiles without the need for sequenced genomes. Discovery of novel miRNAs	http://mirexpress.mbc.nctu.edu.tw/	[154]
mirTools	Discovery of novel miRNAs and piRNAs. Detection and profiling of various types of ncRNAs. Differential expression analysis	http://centre.bioinformatics.zj.cn/mirtools/	[155]

The most commonly used open-source software packages include DEGSeq [102], based on a Poisson distribution of tag count data, or edgeR [100], DESeq [103] and baySeq [104] following a negative binominal distribution of read counts [98].

Although NGS is considered the most promising technology to perform global differential miRNA expression analysis, technical limitations need to be considered. Thus, there are a number of confounding factors, such as sample preparation, efficiency of adapter ligation or amplification biases that affect sequencing coverage and therefore sequence representation and read distribution [105]. For instance, GC content might have a considerable impact on clonal amplification of cDNA libraries prior to sequencing by synthesis [106]. In this context, it is worth mentioning the impact of the method used for small RNA library preparation, ranging from a poly(A) tailing-based method to adapter ligation-mediated procedures, on the final miRNA read distribution and miRNA frequencies [107]. Moreover, systematic and computational biases exist due to variations in read alignment and mapping tools (*see* Table 3 for a list of sequence alignment tools), choice of reference databases, and read count data normalization strategies [108]. Of note, depending on the experiment short reads (25–100 bp) usually obtained by NGS platforms might constitute a downside, since read mapping constrains may occur depending on the short-read mapping software applied [109]. Moreover, downstream computational tools and statistics applied to perform differential expression analysis between biological samples may impact the final results.

Table 3
Most frequently used sequence alignment tools (short reads) [81, 86, 109]
(*see* also: http://en.wikipedia.org/wiki/List_of_sequence_alignment_software)

Name	Description	Website	References
Bowtie	Ultrafast and efficient alignment program; Burrows-Wheeler transform-based indexing of the reference genome	http://bowtie-bio.sourceforge.net/index.shtml	[157]
BWA	Alignment of short reads to large reference sequences; Burrows-Wheeler transform-based; supports paired-end mapping	http://bio-bwa.sourceforge.net/	[158]
SOAP	Alignment program for efficient gapped and ungapped alignment, based on a Burrows-Wheeler transformation compression index; compatible with both single and paired-end reads	http://soap.genomics.org.cn/	[159, 160]
TopHat	Alignment of RNA-Seq data to a genome in order to identify exon-exon splice junctions	http://tophat.cbcb.umd.edu/manual.shtml	[161]

In future, deep sequencing approaches performed to detect and quantify the miRnome of cells, tissues and organisms will substantially contribute to an increasing knowledge of miRNA expression, regulation and function, and their role in complex diseases. However, in order to gain solid conclusions and knowledge, rather than mere big data accumulation, certain challenges need to be faced. First, an adjustment of the level of confidence is required to correctly annotate a short sequence read as a true miRNA. Second, high computational power and the development of appropriate standard software tools is necessary to deal with the increasing data volume produced [110], since the increasing data volume will lead to the discovery of hundreds of novel miRNAs to be mapped and annotated [89]).

The big data volume and recurrent patterns of read mappings will provide valuable knowledge on the relative abundance of mature miRNAs in a given context, and will give insight into the miRNA precursors and the contribution of 5′ and 3′ hairpin arms giving rise to distinct miRNA species. Finally, it will allow a closer look into miRNA isoforms, and the overall abundance and functional significance of isomiRs, which vary in mature miRNA length.

2.6 Detection Methods. Concluding Remarks

The recent advances in deep sequencing technologies have not only resulted in a dramatic increase in the rate of discovery of novel miRNAs [89]. They are also facilitating the identification of modifications of existing miRNA entries, such as miRNA post-transcriptional editing or terminal nucleotide additions. In comparison, microarray and RT-qPCR based technologies would require frequent additions and changes of miRNA microarray probe designs and TaqMan® assay strategies, becoming more ineffective, less suitable for global approaches. The rapid progress and development in the field is reflected by the continuous need for refinement of miRNAs databases such as miRBase [89].

At the time of writing, an increasing body of evidence indicates that miRNA expression analysis and quantification are significantly influenced by the choice of technology, and are particularly dependent on related methodologies [46, 47, 111]. A clear challenge in the field is the standardization of miRNA expression data collection and data analysis tools.

Finally, of crucial importance, different miRNA sources, sample preparation, extraction and enrichment methods, not covered in this introductory chapter, have a considerable impact on miRNA profiling results, particularly in the context of biomarker identification and characterization. There is the strong need for standardized sample acquisition, storage and processing methodologies that allow cross-experiment and cross-platform comparative analyses and validation [111].

3 Computational Systems Biology

As discussed previously, it is reasonable to think that miRNAs exert their functions through regulating cellular networks. Therefore, a systems biology approach represents an important tool to understand miRNAs in neurodegenerative diseases. A system-level understanding of a cell, or of higher units of biological organization can be derived from insight into three key points and their properties [112]:

1. Network construction. The identity of the components that constitute the biological system.
2. Network inference and analysis. The interaction among these components.
3. Modeling. The dynamic behavior of these components.

In this section we introduce the three points and discuss their use.

3.1 Networks Construction

The recent advent of high-throughput/high-content technologies (e.g. high-content screening or deep sequencing) has generated an abundance of data on system elements and interactions that can be used for computational analysis. The data produced by those technologies often represent global patterns of systems (e.g. transcriptome of a tissue). However, to understand the function of the single elements, it is often beneficial to conceptualize those data as a system of interacting elements. Therefore, it is useful to organize the knowledge obtained on a system on a formalized network map. Network maps are graphical representations where nodes represent the molecular components within a cell and their direct or indirect interaction is represented by an edge (Fig. 2). The final goal of creating a network map is to capture the qualitative relationships between the observed measurements. For reviews on network construction see [112–114]. While a good regulatory map is a valuable tool and can provide a wealth of knowledge in itself, these relationships also form the basis for any future mathematical modeling.

The solid base to build a reliable and informative network is the use of high quality experimental data. Those data can be enriched with additional information obtained by textbook knowledge, personal knowledge and experience, and data from online resources. Several databases are available for the purpose of network construction. Regarding miRNAs these databases can be split into two categories.

1. Predicted information.
2. Curated (experimentally verified) information.

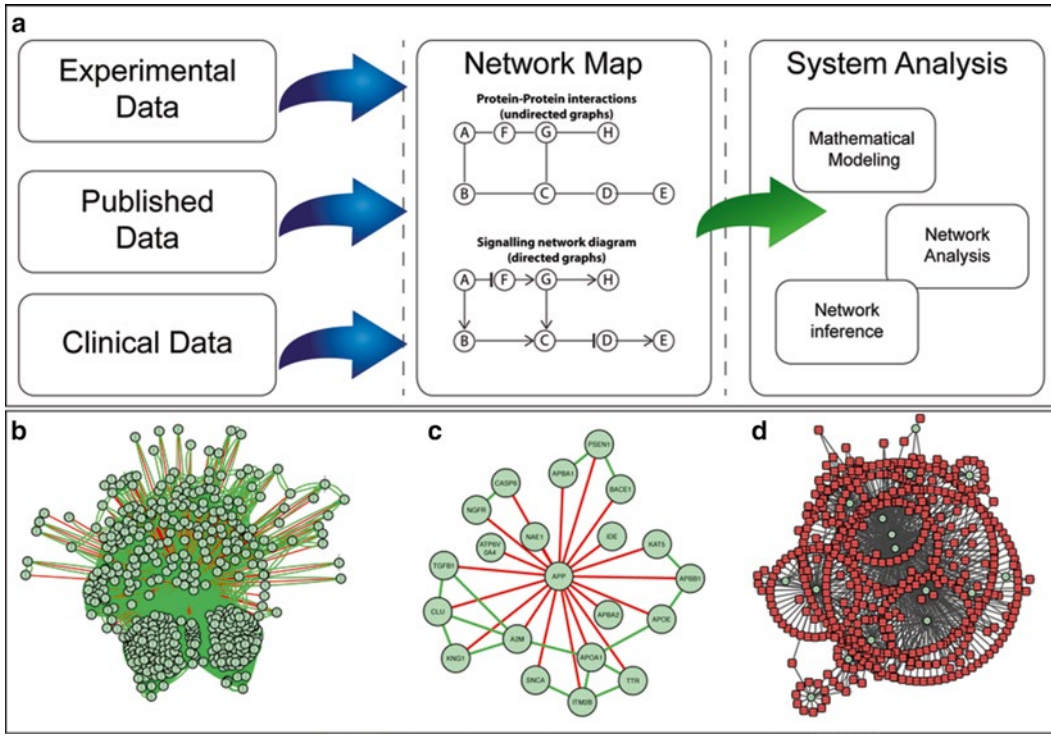


Fig. 2 Systems biology of miRNAs. (**a**) System analysis of miRNAs. To build a formal system approach, comprehensive data from experiments, published data, clinical data or from other sources, need to be collected. From here, the data need to be structured in a network map where the relationships between the different components are defined (for example, a protein-protein interaction or signaling network). The network map can be used to formalize network inference/analysis and predictive modeling. (**b**–**d**) Example of complexity of network construction for the amyloid precursor protein (APP) interactome network. Proteins (*green circles*); miRNA (*red squares*); direct interactors of APP (*red lines*); interactions between interactors of APP (*green lines*); miRNA interactions (*black lines*). Panel **b** shows all interactors of APP as predicted using StringDB (query for ensemble ID of APP, human, limit:9999, confidence > 0.800). To reduce the complexity of the protein-protein interaction in panel **c** the number of APP interactors has been limited to 15 (StringDB, query for ensemble ID of APP (ENSG*****), human, limit:15, confidence > 0.800). In panel **d** miRNAs have been predicted for the 15 closer interactors of APP and the resultant network has been constructed

While both resources are valuable for the study of miRNAs, they also have inherent limitations. Prediction databases for example, suffer from a high rate of false positives [115], and curated information is often sparse and incomplete. It is therefore highly recommended that the researchers make a critical analysis of the caveats for each database and implement actions to avoid results misinterpretation. An example of a miRNA prediction database is TargetScan [116–118], while mirBase [49, 87, 89, 119] and TarBase [120] contain experimental data. At the time of writing the ncRNA website (http://www.ncRNA.org) includes a comprehensive list of available miRNA databases. There are other commercially available curated databases and tools, such as Ingenuity (http://www.ingenuity.com) with high reliability and completeness.

MicroRNAs are translation regulators and one of their key characteristics is that a single miRNA targets many mRNAs. Similarly, regulated mRNAs are targeted by several miRNAs. Therefore, due to this multiple targets modality, miRNA-mRNA interactions generate large networks. For example, starting with a single miRNA of interest (mmu-miR-101a), we can predict 755 interacting mRNAs and, going only one step further, each of these mRNAs yielded ~75 interacting miRNAs. The result is a network containing ~1,500 nodes and ~58,000 edges (Fig. 2).

Additionally, there are several layers that add complexity to the emerging network such as translation regulation, protein-protein interactions (PPI), transcription factor activity and promoter accessibility. Combined, these biological concepts result in system networks which are highly enriched in network motifs, such as feedback- and feed-forward-loops, target hubs and clusters. It becomes obvious that manual evaluation of such networks is not feasible and the system analysis of miRNAs is a prime field for computational studies. At this point it is important to stress that functional miRNA analysis needs always to take into consideration the mRNA counterpart, as the miRNA-mRNA tandem represents the essential unit of the miRNA network/system.

3.2 Network Inference and Analysis

Network inference refers to the collection of approaches to predict (infer) new relationships (edges) in a given network. There is a wide variety of methods and algorithms available (e.g. regression analysis, Bayesian inference), and the researcher has to accurately choose the methods best fitting the scientific question under investigation. However, as emerged during the Dialogue for Reverse Engineering Assessments and Methods 3 (DREAM3) challenge [121, 122] aimed to analyze genetic and biological networks, currently there is no unique algorithm able to best address and solve relevant problems related to different biological networks [122]. Hence, the investigator needs to test different methods in order to check the best solution for the given question.

The great challenge of network inference is the issue of underdetermination, i.e. having not enough constrains to specify a unique solution. In practice, the number of possible solutions greatly exceeds the number of available data points and there is no unique solution given the gathered evidence [123]. For this reason, it is necessary to simplify the problem as much as possible, e.g. making assumptions, to reduce the number of possible solutions. In these cases a number of conceptual and computational methods can help to reduce the pool of possible outcomes and validate their efficiency [124]. For example, many biological networks are usually sparse, meaning that the number of edges for each node is low, i.e. assumption of sparseness of biological networks [125] and regression analysis [126].

The quality of the input data is important for correct network inference. Also, the existing research bias in publicly available data,

with the majority of data coming from studies focusing on a relatively low number of proteins (e.g. most often studied Alzheimer-related proteins like APP, ß-secretase or tau) needs to be kept in mind. Despite these caveats, the prediction of previously unknown interactions in a network is a very useful tool. A comparison of over 30 network inference approaches for different conditions can be found in [127].

Network analysis deals with the extraction of information out of a given network. Whether the network has been self-constructed, downloaded, or mined for, is of secondary importance. The most critical factor is the quality of the original data which determines the results and information extracted from the analysis. As biological networks are very diverse and highly complex, here again there is no single best approach or algorithm that suits all network types and questions. Hence, we will give an overview of the main methods and their use in network analysis. The choice and application of a specific method will depend greatly on the given network and the ultimate goal.

In its roots, network analysis derives from the mathematical branch of graph theory, where a graph consists of "nodes" and "edges" describing the relationship between these nodes. Leonard Euler laid the foundation of graph theory in a paper published in 1741 called "The Seven Bridges of Koenigsberg" [128]. Today, graph theory is being used in a wide variety of fields, including social sciences, linguistics, economics, physics and biology. Networks from all of these fields share common properties and, thus, many of the approaches to analyze a given graph can be applied to all. However, biological networks do have properties reflecting their specific nature. For instance, in contrast to a random network in which all nodes are connected by randomly placed edges, a biological network is a scale-free network. Mathematically speaking, a scale-free network is a topological connected graph or network with the property that the number of links k (degree) originating from a given node exhibits a power-law distribution.

$$P(k) \sim k^{-y} \left(k : \text{degree.}\ \ y : [2 < y < 3]\right)$$

Compared to a random network's Poisson distribution, a power-law distribution is right-skewed, effectively meaning that there is a higher occurrence of nodes having an above-average number of edges, so called hub nodes. A second property of these networks is that they have a degree of self-similarity, a feature that can also be used for mathematical modeling and simulation.

The analysis of a given network can be done using several statistical features. However, using the main network properties (i.e. degree, distance, diameter, clustering coefficient and betweeness) can already convey meaning to individual nodes. Finding the mentioned "hub nodes" is often the first step in analyzing a network. To this end the *degree* of each node is determined. The degree is

the number of edges a node has and, thus, is a direct feature of the "hub-ness" of each node. The *distance* between two nodes *u* and *v* is defined as the shortest path connecting them. The mean of the distances in a network and the maximum distance between two nodes (the *diameter*) are measures of the size of the network. Small diameter networks are referred to as "small world" networks [129], which are characterized by the fact that two arbitrary nodes in the network can be connected through relatively short paths. In biology, metabolic networks are an example of such networks and this property might well serve a biological function [130].

The *clustering coefficient* is the ratio of existing edges in a neighborhood to the maximum possible edges that could exist in that neighborhood. Small world networks usually have a high clustering coefficient. *Betweenness* is a measure of the importance of a node, assuming that functional relationships between any two nodes follow the shortest path connection. As such, betweenness is defined mathematically as the fraction of shortest-paths connecting all pairs of nodes that run through the given node.

Network analysis allows to go beyond individual node properties e.g. by looking at the position of a node inside the network. The topology of a network is represented by the distribution of edges between the given nodes, which may reflect functional relationships. Network *motifs* are recurring sub-graphs that are statistically overrepresented compared to random networks. In biological networks, these motifs represent such concepts as auto-regulation, feed-forward and feed-back loops or single-input modules [131]. It has to be noted that the implicit assumption "structure implies function" of this approach has been questioned at least once [132]. A different concept, originally introduced in 1949 by Luce and Perry for social networks is that of *cliques* [133]. A clique is a group of people all knowing each other. From a graph theoretic point of view, a clique is a complete sub-graph, i.e. a part of a graph in which all nodes are interconnected. Cliques have been used in PPI networks to find groups of closely interacting proteins [134] and form the basis of power graphs, an approach to compress graphs through the use of cliques [135].

Graphlets are predefined sub-graphs consisting of a given number of nodes. These sub-graphs represent all of the possible structures that can be formed by the given number of nodes. Due to the exponential increase in possible structures, the maximum number of nodes considered is usually around 5, although there is no theoretical limit to the number of nodes. The analysis of a given network then works by finding these graphlets inside the larger network and calculating statistical measures based on these graphlets. A standard application for graphlet-based analysis is similarity comparison of two given networks [136]. By comparison of sub-graphs inside a network it is also possible to perform functional prediction of uncharacterized nodes [137].

Together with this, there are several algorithms and tools to compute larger sub-graphs, usually based on high connectivity measures. These larger sub-networks are not motifs, cliques or graphlets, but instead are groups of nodes standing out, related to the measurement criteria used, for instance connectivity (degree).

In terms of visualization and analysis of networks there is a plethora of software and programming tools available, ranging from basic programming libraries for C++ (SNAP) or Python (NetworkX), over interpreter-based modules for MATLAB or R programs, to fully integrated open source software platforms like Cytoscape (http://www.cytoscape.org) that can be enhanced and tweaked via plugins.

3.3 Predictive Modeling

A reliable (inferred) network combined with appropriate analyses is a great tool to gain new insights into biological processes and/or predict novel functions of its members. However, a wealth of additional information can be gained by using the given network as a basis for mathematical modeling. Eykhoff [138] defined a mathematical model as "a representation of the essential aspects of an existing system which presents knowledge of that system in usable form". For instance, simulation of a transcription regulation network (TRN) can predict the performance of a network over time and/or its reaction to perturbation. Both of these go well beyond the static snapshot information that a single transcriptome or deep sequencing analysis usually provides.

Mathematical models can be as simple [139] or complex [140] as desired, where the level of detail depends largely on the posed question. On the other hand, especially for models of large interaction networks, the level of detail is limited by the rapidly growing complexity of the system, both in terms of designing the mathematical model and of computing the finished simulation framework. Thus, the choice of the mathematical method used for a given network is influenced strongly by considerations of accuracy and computational efficiency.

Differential equation systems are the most common methods for modeling regulatory networks. They allow simulating the concentrations of network components (mRNAs, proteins, miRNAs) in a continuous and deterministic way. However, it has been argued that this does not reflect biological reality [141]. Furthermore, the necessary parameters (e.g. kinetic rate constants, etc.) are unknown or may be derived from in vitro experiments and might not match the "true" values within a cell [142]. This is a general caveat of modeling biological systems. Approaches have been developed to work around this problem [143].

Stochastic methods like, for instance, Monte Carlo simulations incorporate a probabilistic and discreet component to model the behavior of a network [144]. Probability distributions describing the amount of each component within the system are introduced

and drawn from, to determine the change of the system over time. Conceptually, this method is closer to biological reality as, for instance, protein binding is actually being described as the probability that two proteins collide within a given time. The disadvantage of this type of simulations is usually the computational efficiency, since it is necessary to repeat each simulation a large number of times. A closer look at several approaches to stochastic modeling of gene expression can be found in [145], and an example of the whole process for a gene regulatory network including miRNAs is provided in [146].

A third common modeling method used for regulatory networks are Boolean networks. In basic Boolean network models, each node (gene, mRNA, protein or microRNA) has an associated value of 0 or 1 indicating, for instance, activity or abundance (on/off, high/low). This approach is strongly simplified, but has the advantage of being computationally extremely simple, allowing the computation of large networks. An introduction to Boolean network can be found in [147] and real world examples in [148]. Finally, for an overview of computational methods on regulatory networks the reader can refer to de Jong's work [149].

In complex neurodegenerative diseases like AD it may not be sufficient to model the dynamics of a single cell (e.g. a neuron) alone. Alzheimer's disease affects the whole brain at tissue level, in different tissues and regions, and has been reported to spread starting from the transentorhinal cortex [150]. Supra-cellular models, with interplay between different cells may be relevant, which can be performed using a multiscale modeling approach. As the name implies, multiscale refers to the concept of including different size scales in a single model like, for instance, gene regulatory networks of individual cells (cell scale) and cell-cell interactions of the same cells (tissue scale). An overview of this approach applied to cancer can be found in [151].

4 Conclusions

In this chapter we have discussed the role of miRNAs in AD and the use of computational systems biology methods for their investigation. It is clear that miRNAs can rarely be considered acting in isolation or having an effect on a single defined pathway only, but rather acting at a "system" level involving transcriptional, post-transcriptional and signaling regulation in a cell/tissue specific manner. Hence, systems biology is the best approach to investigate the role of miRNAs in the onset and development of AD. Some limitations in the technologies for miRNAs detection (e.g. sensitivity issues) and computational methods (e.g. database accuracy, sequence prediction validity) are still hampering the full exploitation of the miRNAs networks in systems biology approaches.

To overcome these limitations there is the need to increase the understanding of miRNA's (and other non-coding RNA's) mechanisms, and to sharpen present technologies to investigate miRNAs at an experimental and computational level. We expect the next years to bring exciting developments on the role of miRNAs in AD and other neurodegenerative diseases, and a stronger integration among experimental, clinical and computational biology approaches, in what we consider true systems biology.

Acknowledgements

We would like to thank Prof. Eckhard Mandelkow for his encouragement in writing this chapter.

References

1. Caspersson T, Schultz J (1940) Ribonucleic acids in both nucleus and cytoplasm, and the function of the nucleolus. Proc Natl Acad Sci U S A 26:507–515
2. Caspersson T, Schultz J (1939) Pentose nucleotides in the cytoplasm of growing tissues. Nature 143:602–603
3. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738
4. Crick FH (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163
5. Holley RW, Apgar J, Everett GA et al (1965) Structure of a ribonucleic acid. Science 147:1462–1465
6. Palade GE (1955) A small particulate component of the cytoplasm. J Biophys Biochem Cytol 1:59–68
7. Ban N, Nissen P, Hansen J et al (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905–920
8. Esteller M (2011) Non-coding RNAs in human disease. Nat Rev Genet 12:861–874
9. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297
10. Gomes AQ, Nolasco S, Soares H (2013) Non-coding RNAs: multi-tasking molecules in the cell. Int J Mol Sci 14:16010–16039
11. Winter J, Jung S, Keller S et al (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 11:228–234
12. Wu H, Xu H, Miraglia LJ, Crooke ST (2000) Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. J Biol Chem 275:36957–36965
13. Lee Y, Ahn C, Han J et al (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425:415–419
14. Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res 32:4776–4785
15. Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25–33
16. Bagga S, Bracht J, Hunter S et al (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell 122:553–563
17. Olsen PH, Ambros V (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol 216:671–680
18. Chekulaeva M, Filipowicz W (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Curr Opin Cell Biol 21:452–460
19. Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol 7:719–723
20. Brengues M, Teixeira D, Parker R (2005) Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. Science 310:486–489

21. Ebert MS, Sharp PA (2010) Emerging roles for natural microRNA sponges. Curr Biol 20: R858–R861
22. Poliseno L, Salmena L, Zhang J et al (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465:1033–1038
23. Memczak S, Jens M, Elefsinioti A et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495: 333–338
24. Salzman J, Gawad C, Wang PL et al (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 7:e30733
25. Suh YH, Checler F (2002) Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. Pharmacol Rev 54:469–525
26. Pasquinelli AE, Reinhart BJ, Slack F et al (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408:86–89
27. Lukiw WJ (2007) Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus. Neuroreport 18:297–300
28. Delay C, Mandemakers W, Hebert SS (2012) MicroRNAs in Alzheimer's disease. Neurobiol Dis 46:285–290
29. Johnson R, Noble W, Tartaglia GG, Buckley NJ (2012) Neurodegeneration as an RNA disorder. Prog Neurobiol 99:293–315
30. Satoh J (2012) Molecular network of microRNA targets in Alzheimer's disease brains. Exp Neurol 235:436–446
31. Satoh J (2012) Molecular network analysis of human microRNA targetome: from cancers to Alzheimer's disease. BioData Min 5:17
32. Hebert SS, De Strooper B (2009) Alterations of the microRNA network cause neurodegenerative disease. Trends Neurosci 32:199–206
33. Junn E, Mouradian MM (2012) MicroRNAs in neurodegenerative diseases and their therapeutic potential. Pharmacol Ther 133: 142–150
34. Lau P, Bossers K, Janky R et al (2013) Alteration of the microRNA network during the progression of Alzheimer's disease. EMBO Mol Med 5:1613–1634
35. Salta E, De Strooper B (2012) Non-coding RNAs with essential roles in neurodegenerative disorders. Lancet Neurol 11:189–200
36. Smith P, Al Hashimi A, Girard J et al (2011) In vivo regulation of amyloid precursor protein neuronal splicing by microRNAs. J Neurochem 116:240–247
37. Smith PY, Delay C, Girard J, Papon MA et al (2011) MicroRNA-132 loss is associated with tau exon 10 inclusion in progressive supranuclear palsy. Hum Mol Genet 20:4016–4024
38. Hebert SS, Papadopoulou AS, Smith P et al (2010) Genetic ablation of Dicer in adult forebrain neurons results in abnormal tau hyperphosphorylation and neurodegeneration. Hum Mol Genet 19:3959–3969
39. Yin JQ, Zhao RC, Morris KV (2008) Profiling microRNA expression with microarrays. Trends Biotechnol 26:70–76
40. Li W, Ruan K (2009) MicroRNA detection by microarray. Anal Bioanal Chem 394: 1117–1124
41. Chen C, Ridzon DA, Broomer AJ et al (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33:e179
42. Chen C, Tan R, Wong L et al (2011) Quantitation of microRNAs by real-time RT-qPCR. Methods Mol Biol 687:113–134
43. Hurley J, Roberts D, Bond A et al (2012) Stem-loop RT-qPCR for microRNA expression profiling. Methods Mol Biol 822:33–52
44. Hafner M, Landgraf P, Ludwig J et al (2008) Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods 44:3–12
45. Berninger P, Gaidatzis D, van Nimwegen E, Zavolan M (2008) Computational analysis of small RNA cloning data. Methods 44:13–21
46. Baker M (2010) MicroRNA profiling: separating signal from noise. Nat Methods 7:687–692
47. Git A, Dvinge H, Salmon-Divon M et al (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA 16:991–1006
48. Pritchard CC, Cheng HH, Tewari M (2012) MicroRNA profiling: approaches and considerations. Nat Rev Genet 13:358–369
49. Griffiths-Jones S (2004) The microRNA registry. Nucleic Acids Res 32(Database issue):D109–D111
50. Jiang J, Lee EJ, Gusev Y, Schmittgen TD (2005) Real-time expression profiling of microRNA precursors in human cancer cell lines. Nucleic Acids Res 33:5394–5403
51. Landgraf P, Rusu M, Sheridan R et al (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129:1401–1414
52. Morin RD, O'Connor MD, Griffith M et al (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18:610–621

53. Liu CG, Calin GA, Meloon B et al (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc Natl Acad Sci U S A 101:9740–9744
54. Liu CG, Spizzo R, Calin GA, Croce CM (2008) Expression profiling of microRNA using oligo DNA arrays. Methods 44:22–30
55. Liu CG, Calin GA, Volinia S, Croce CM (2008) MicroRNA expression profiling using microarrays. Nat Protoc 3:563–578
56. Babak T, Zhang W, Morris Q et al (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. RNA 10:1813–1819
57. Barad O, Meiri E, Avniel A et al (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res 14:2486–2494
58. Miska EA, Alvarez-Saavedra E, Townsend M et al (2004) Microarray analysis of microRNA expression in the developing mammalian brain. Genome Biol 5:R68
59. Nelson PT, Baldwin DA, Scearce LM et al (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. Nat Methods 1:155–161
60. Thomson JM, Parker J, Perou CM, Hammond SM (2004) A custom microarray platform for analysis of microRNA gene expression. Nat Methods 1:47–53
61. Castoldi M, Schmidt S, Benes V et al (2006) A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA 12:913–920
62. Kramer MF (2011) Stem-loop RT-qPCR for miRNAs. Curr Protoc Mol Biol Chapter 15:Unit 15.10. doi:10.1002/0471142727.mb1510s95
63. Castoldi M, Benes V, Hentze MW, Muckenthaler MU (2007) miChip: a microarray platform for expression profiling of microRNAs based on locked nucleic acid (LNA) oligonucleotide capture probes. Methods 43:146–152
64. Valoczi A, Hornyik C, Varga N et al (2004) Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res 32:e175
65. Nelson PT, Baldwin DA, Kloosterman WP et al (2006) RAKE and LNA-ISH reveal microRNA expression and localization in archival human brain. RNA 12:187–191
66. Castoldi M, Schmidt S, Benes V et al (2008) miChip: an array-based method for microRNA expression profiling using locked nucleic acid capture probes. Nat Protoc 3:321–329
67. De Cecco L, Dugo M, Canevari S et al (2013) Measuring microRNA expression levels in oncology: from samples to data analysis. Crit Rev Oncog 18:273–287
68. Schmittgen TD, Jiang J, Liu Q, Yang L (2004) A high-throughput method to monitor the expression of microRNA precursors. Nucleic Acids Res 32:e43
69. Moltzahn F, Olshen AB, Baehner L et al (2011) Microfluidic-based multiplex qRT-PCR identifies diagnostic and prognostic microRNA signatures in the sera of prostate cancer patients. Cancer Res 71:550–560
70. Shi R, Chiang VL (2005) Facile means for quantifying microRNA expression by real-time PCR. Biotechniques 39:519–525
71. Raymond CK, Roberts BS, Garrett-Engele P et al (2005) Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. RNA 11:1737–1744
72. Benes V, Castoldi M (2010) Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. Methods 50:244–249
73. Redshaw N, Wilkes T, Whale A et al (2013) A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability. Biotechniques 54:155–164
74. Schmittgen TD, Lee EJ, Jiang J et al (2008) Real-time PCR quantification of precursor and mature microRNA. Methods 44:31–38
75. Liu DZ, Tian Y, Ander BP et al (2010) Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. J Cereb Blood Flow Metab 30:92–101
76. Bustin SA, Benes V, Garson JA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–622
77. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3:1101–1108
78. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402–408
79. Peltier HJ, Latham GJ (2008) Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. RNA 14:844–852
80. Mestdagh P, Van Vlierberghe P, De Weer A et al (2009) A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol 10:R64

81. Liu J, Jennings SF, Tong W, Hong H (2011) Next generation sequencing for profiling expression of miRNAs: technical progress and applications in drug development. J Biomed Sci Eng 4:666–676
82. Kircher M, Kelso J (2010) High-throughput DNA sequencing—concepts and limitations. Bioessays 32:524–536
83. Rizzo JM, Buck MJ (2012) Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila) 5:887–900
84. Eminaga S, Christodoulou DC, Vigneault F et al. (2013) Quantification of microRNA expression with next-generation sequencing. Curr Protoc Mol Biol Chapter 4:Unit 4.17. doi:10.1002/0471142727.mb0417s103
85. Hafner M, Renwick N, Brown M et al (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA 17:1697–1712
86. Flicek P, Birney E (2009) Sense from sequence reads: methods for alignment and assembly. Nat Methods 6(11 Suppl):S6–S12
87. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36(Database issue):D154–D158
88. Yang JH, Shao P, Zhou H et al (2010) deepBase: a database for deeply annotating and mining deep sequencing data. Nucleic Acids Res 38(Database issue):D123–D130
89. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39(Database issue):D152–D157
90. Flicek P, Ahmed I, Amode MR et al (2013) Ensembl 2013. Nucleic Acids Res 41(Database issue):D48–D55
91. Meyer LR, Zweig AS, Hinrichs AS et al (2013) The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res 41(Database issue):D64–D69
92. Friedlander MR, Chen W, Adamidi C et al (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 26(4):407–415
93. Hackenberg M, Rodriguez-Ezpeleta N, Aransay AM (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res 39(Web Server issue):W132–W138
94. An J, Lai J, Lehman ML, Nelson CC (2013) miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. Nucleic Acids Res 41:727–737
95. Hendrix D, Levine M, Shi W (2010) miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol 11:R39
96. Mathelier A, Carbone A (2010) MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics 26:2226–2234
97. Li Y, Zhang Z, Liu F et al (2012) Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. Nucleic Acids Res 40:4298–4305
98. Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. Genome Biol 11:220
99. Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628
100. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25
101. Langmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biol 11:R83
102. Wang L, Feng Z, Wang X et al (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26:136–138
103. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106
104. Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422
105. Sendler E, Johnson GD, Krawetz SA (2011) Local and global factors affecting RNA sequencing analysis. Anal Biochem 419:317–322
106. Aird D, Ross MG, Chen WS et al (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12:R18
107. Linsen SE, de Wit E, Janssens G et al (2009) Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods 6:474–476
108. Hansen KD, Irizarry RA, Wu Z (2012) Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13:204–216
109. Trapnell C, Salzberg SL (2009) How to map billions of short reads onto genomes. Nat Biotechnol 27:455–457

110. McPherson JD (2009) Next-generation gap. Nat Methods 6(11 Suppl):S2–S5
111. Chugh P, Dittmer DP (2012) Potential pitfalls in microRNA profiling. Wiley Interdiscip Rev RNA 3:601–616
112. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
113. Albert R (2007) Network inference, analysis, and modeling in systems biology. Plant Cell 19:3327–3338
114. Ma'ayan A (2011) Introduction to network analysis in systems biology. Sci Signal 4:tr5. doi:10.1126/scisignal.2001965
115. Lewis BP, Shih IH, Jones-Rhoades MW et al (2003) Prediction of mammalian microRNA targets. Cell 115:787–798
116. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20
117. Grimson A, Farh KK, Johnston WK et al (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27:91–105
118. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19:92–105
119. Griffiths-Jones S, Grocock RJ, van Dongen S et al (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34(Database issue):D140–D144
120. Vergoulis T, Vlachos IS, Alexiou P et al (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res 40(Database issue):D222–D229
121. Marbach D, Prill RJ, Schaffter T et al (2010) Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci U S A 107:6286–6291
122. Prill RJ, Marbach D, Saez-Rodriguez J et al (2010) Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PLoS One 5:e9202
123. De Smet R, Marchal K (2010) Advantages and limitations of current network inference methods. Nat Rev Microbiol 8:717–729
124. Szederkenyi G, Banga JR, Alonso AA (2011) Inference of complex biological networks: distinguishability issues and optimization-based solutions. BMC Syst Biol 5:177
125. de Hoon MJ, Imoto S, Kobayashi K et al (2003) Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations. Pac Symp Biocomput 8:17–28
126. Kato M, Tsunoda T, Takagi T (2000) Inferring genetic networks from DNA microarray data by multiple regression analysis. Genome Inform 11:118–128
127. Marbach D, Costello JC, Kuffner R et al (2012) Wisdom of crowds for robust gene network inference. Nat Methods 9:796–804
128. Euler L (1741) Solutio problematis ad geometriam situs pertinentis. Commentarii Academiae Scientiarum Petropolitanae 8:128–140
129. Travers J, Milgram S (1969) An experimental study of the small world problem. Sociometry 32:425–443
130. Wagner A, Fell DA (2001) The small world inside large metabolic networks. Proc Biol Sci 268:1803–1810
131. Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461
132. Ingram P, Stumpf M, Stark J (2006) Network motifs: structure does not determine function. BMC Genomics 7:108
133. Luce RD, Perry AD (1949) A method of matrix analysis of group structure. Psychometrika 14:95–116
134. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A 100: 12123–12128
135. Royer L, Reimann M, Andreopoulos B, Schroeder M (2008) Unraveling protein networks with power graph analysis. PLoS Comput Biol 4:e1000108
136. Pržulj N (2007) Biological network comparison using graphlet degree distribution. Bioinformatics 23:e177–e183
137. Milenkovic T, Przulj N (2008) Uncovering biological network function via graphlet degree signatures. Cancer Inform 6:257–273
138. Eykhoff P (1974) System identification: parameter and state estimation. Wiley-Interscience, New York
139. Levine E, Ben Jacob E, Levine H (2007) Target-specific and global effectors in gene regulation by MicroRNA. Biophys J 93: L52–L54
140. Zinovyev A, Morozova N, Gorban AN, Harel-Belan A (2013) Mathematical modeling of microRNA-mediated mechanisms of translation repression. Adv Exp Med Biol 774:189–224
141. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81:2340–2361

142. Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J Biol Chem 276: 10577–10580
143. Brown KS, Hill CC, Calero GA et al (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. Phys Biol 1:184–195
144. El Samad H, Khammash M, Petzold L, Gillespie D (2005) Stochastic modelling of gene regulatory networks. Int J Robust Nonlinear Control 15:691–711
145. Bokes P, King JR, Wood AT, Loose M (2012) Multiscale stochastic modelling of gene expression. J Math Biol 65:493–520
146. Lai X, Bhattacharya A, Schmitz U et al (2013) A systems biology approach to study microRNA-mediated gene regulatory networks. Biomed Res Int 2013:703849
147. Xiao Y (2009) A tutorial on analysis and simulation of boolean gene regulatory network models. Curr Genomics 10:511–525
148. Albert I, Thakar J, Li S et al (2008) Boolean network simulations for life scientists. Source Code Biol Med 3:16
149. de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9:67–103
150. Braak H, Braak E (1995) Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16:271–278, discussion 16:278-284
151. Chaplain MAJ (2011) Multiscale mathematical modelling in biology and medicine. IMA J Appl Math 76:371–388
152. Betel D, Wilson M, Gabow A et al (2008) The microRNA.org resource: targets and expression. Nucleic Acids Res 36(Database issue):D149–D153
153. Alexiou P, Vergoulis T, Gleditzsch M et al (2010) miRGen 2.0: a database of microRNA genomic information and regulation. Nucleic Acids Res 38(Database issue):D137–D141
154. Wang WC, Lin FM, Chang WC et al (2009) miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. BMC Bioinformatics 10:328
155. Wu J, Liu Q, Wang X et al (2013) mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. RNA Biol 10: 1087–1092
156. Gomes CP, Cho JH, Hood L et al (2013) A review of computational tools in microRNA discovery. Front Genet 4:81
157. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
158. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
159. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714
160. Li R, Yu C, Li Y et al (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967
161. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

Chapter 22

The Emerging Role of Metalloproteomics in Alzheimer's Disease Research

Dominic J. Hare, Alan Rembach, and Blaine R. Roberts

Abstract

Metals are increasingly recognized to have an important role in molecular processes underlying Alzheimer's disease (AD). This chapter discusses the current role of metals in AD and expands on the development of metalloproteomics and how the recent advances in analytical technology will allow detailed investigation of metalloproteins. Investigation of individual metalloproteins will yield new mechanistic details about the role of metals in AD.

Key words Metalloproteomics, Alzheimer's disease, Metals, Amyloid, ICP-MS

1 Metal Hypothesis of Alzheimer's Disease

It has been more than 100 years since Alzheimer's disease (AD) was first described and despite significant progress in understanding the pathological hallmarks of the disease, a lack of understanding of the principal mechanism that leads to AD still remains. With an ageing population that is living longer, healthier lives, AD now affects approximately 2 % of the world population and is the third leading cause of death [1]. To date, there are no effective therapies for AD and for this reason it remains a looming major social and economic crisis.

Within the brain, the principal hallmarks of AD are the appearance of extracellular senile plaques, predominantly containing aggregated β-amyloid (Aβ), a small cleaved product from the β-amyloid precursor protein (APP) [2], and neurofibrillary tangles consisting of insoluble intra-neuronal inclusions formed from hyperphosphorylated tau [3]. The vast majority of AD cases are sporadic with patients having no family history, yet, there are some genetic factors that appear to modulate an individual's risk for developing AD [4]. However, age is still the major risk factor for AD with the majority of cases manifesting after the age of 85 (*see* ref. 5 for review).

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_22,

Over the last 40 years considerable interest has grown in the hypothesis that biologically active metals may play a major role in the etiology of AD [6]. Metals are nearly ubiquitous in multicellular organisms and play a critical role in a range of physiological systems. Their unique properties render them vital for a range of reactions that are needed in biological pathways, but when not properly compartmentalized or processed outside controlled physiological parameters, metals can lead to biological damage [7].

The central nervous system (CNS) is particularly vulnerable to perturbations in metal homeostasis because of the risk of oxidative stress and uncoupling of antioxidant defense, therefore they are particularly relevant to neurodegenerative diseases. It has been estimated that approximately 30 % (but possibly more) of all enzymes utilize metals for normal processes [8] however, metal ion concentrations in the CNS are tightly controlled with little to no passive exchange from the circulation to the brain. The majority of metal ions, are highly regulated and in the CNS are bound to ligands, with some loosely bounds metals that are able to be exchanged, in coordinated transfers.

2 Iron, Copper and Zinc in Alzheimer's Disease

The three main transition metals; copper, zinc and iron are of particular interest as they are involved in either synaptic function or modulate the expression of certain pathways [9–12]. Copper is transported in the periphery by chaperones including ceruloplasmin and albumin and crosses the blood brain barrier (BBB) by transporters ATP7A and Ctr1 [9–12]. Zinc is transported in the periphery by metallothioneins and crosses the BBB by transporters ZnT1-10 [12–14]. Iron is transported in the periphery ferritin and crosses the BBB through transferrin [10, 12].

The above mentioned transition metals are essential for life and are involved in a range of pathways and enzymes critical for homeostasis [13]. However, as mentioned, many metals are redox active and therefore have been implicated in potentially driving oxidative stress in AD [15]. This had led to a "metal hypothesis of AD" which has gained support recently as evidence accumulates that Aβ itself drives toxicity through direct interactions with metals.

Perturbed metal homeostasis is evident in the development and progression of AD. Affected AD brain tissues are copper deficient but high in iron and zinc compared to aged-matched controls [16–20], and high concentrations of metals are present in amyloid plaques [21, 22]. Altered metal homeostasis and APP processing in the AD brain is postulated to contribute to aberrant interactions between copper or zinc and Aβ in the vicinity of glutamatergic synapses, resulting in the formation of toxic Aβ oligomers, and consequently, loss of neurons [23].

APP contains copper and zinc binding sites in its N-terminal domain, and the Aβ peptide is capable of binding copper, zinc and iron [24–28]. These metals are also directly associated with the regulation of expression and activity of many key proteins associated with AD. Copper regulates the expression and activities of both APP and the β-secretase, BACE1 [29–32]. Zinc is implicated in modulating the expression of presenilins, catalytic subunits of the γ-secretase complex [33, 34]; inhibition of APP cleavage by γ-secretase [34]; and regulation of α-secretase activities, which are zinc metalloproteases cleaving APP within the Aβ sequence to generate a non-amyloidogenic peptide [35]. Therefore, disruption to normal metal metabolism affects many aspects of APP expression and processing. APP has been assigned functional roles in copper and iron homeostasis. Over-expression of APP or the β-cleaved C-terminal fragment in transgenic animals resulted in a copper-deficient phenotype [36, 37]; whereas genetic ablation of APP in animal and cell culture studies led to copper accumulation [38, 39]. These data led to the hypothesis that APP participates in the copper efflux pathway.

There appears to be an intrinsic relationship between the expression of APP and Aβ and co-localization of metals in the AD brain. Whether metal dyshomeostasis is a downstream consequence of perturbed Aβ isoform accumulation and or failed clearance is still to be determined. New tools are required to interrogate the proteome to evaluate the hypothesis that metals are a critical link in the manifestation of AD and should be considered as potential therapeutic targets. The following section will outline the development of such tools and techniques that will now begin to address some of these questions.

3 Principles of Metalloproteomics and Implications for Alzheimer's Disease

Examining the interactions between metals and Aβ demands analytical techniques capable of highly accurate measurements of both metal species and the Aβ peptide that can be interrelated with a high degree of precision. Recent advances in both separation mechanisms and mass-specific detection techniques have made available a wide range of versatile approaches applicable not only to probing the relationship between metals and Aβ, but also the numerous other cellular processes implicated in AD pathology.

In the post-genomic era, a veritable explosion of so-called 'designer' 'omic' sciences [40] filled the space that would form the foundation of modern systems biology. Defined as the study of an entity in aggregate [41], a range of new 'omic' sciences proliferated in the early twenty-first century, focusing on the protein, lipid, metabolic, transcription factors and a raft of other cellular components, typically in isolation. The metal species found within the cell

was no different, and in 2004 the term '*metallomics*' was concurrently proposed by Haraguchi [42] and Szpunar [43] to describe the study of the combined free metals, metal-binding proteins and metalloenzymes within an organism, referred to as the *metallome*. Haraguchi identified that metallomics' place is 'in symbiosis with genomics and proteomics, because syntheses and metabolic functions of genes (DNA and RNA) and proteins cannot be performed without the aid of various metal ions and metalloenzymes.' As such, a shift in focus from metal species itself to the biomolecule to which it is bound warrants the re-defining of contemporary 'metallomics' as '*metalloproteomics*', which better reflects the ubiquity of biometals in the cell and recognizes the role chaperoning proteins play in metal metabolism.

Central to the principles of metalloproteomics are the analytical techniques required for the comprehensive study of both metal species and related biomolecules. The most significant advances in metalloproteomics have come about from the integration of inductively coupled plasma-mass spectrometry (ICP-MS) into the life sciences, which, when combined with the organic mass spectrometry approaches common to proteomics, provides a powerful tool for studying the 'metalloproteome'. It is estimated that as many of 50 % of the total proteins within this pool contain at least one metal ion [44], with many dependent on this metal ion for function. Relating metal species to protein function is still a major gap in our understanding of the roles metals play in biology. In 2010, Cvetkovic et al. [45] studied the metal-binding species *Pyrococcus furiosus* using ICP-MS, finding a vast range of metalloproteins with no previous known metal cofactor, which has since led to the conclusion that in a more complex organism (like humans) as many as 20 % of the total proteome will have either unexpected metal interactions or contain metal ions with incorrectly predicted metal association [46]. Thus, in concert with the analytical difficulties in probing changes in ubiquitous metal species specific to certain pathological conditions, much of our knowledge regarding metal-binding proteins is either clouded or even incorrect. Thus, the development of new analytical strategies that combines the highly sensitive yet protein-unspecific detection of ICP-MS with traditional proteomic workflows allowing for simultaneous protein structural elucidation and metal quantification (*see* Fig. 1) is highly

Fig. 1 (continued) from each group are then mixed and separated according to traditional two-dimensional liquid chromatography. 10 % of the eluent is analyzed for metal content by ICP-MS, and the remaining 90 % undergoes tandem mass spectrometry. Combining the two data sets provides quantitative information on both protein levels (via stable isotope labeling) and metal levels. Relating protein identification to activity via a metal cofactor provides insight into the functional role of metalloproteins. Reproduced from Lothian et al. [46], with permission (open-access article distributed under the terms of the Creative Commons Attribution License (http://www.frontiersin.org/about/openaccess))

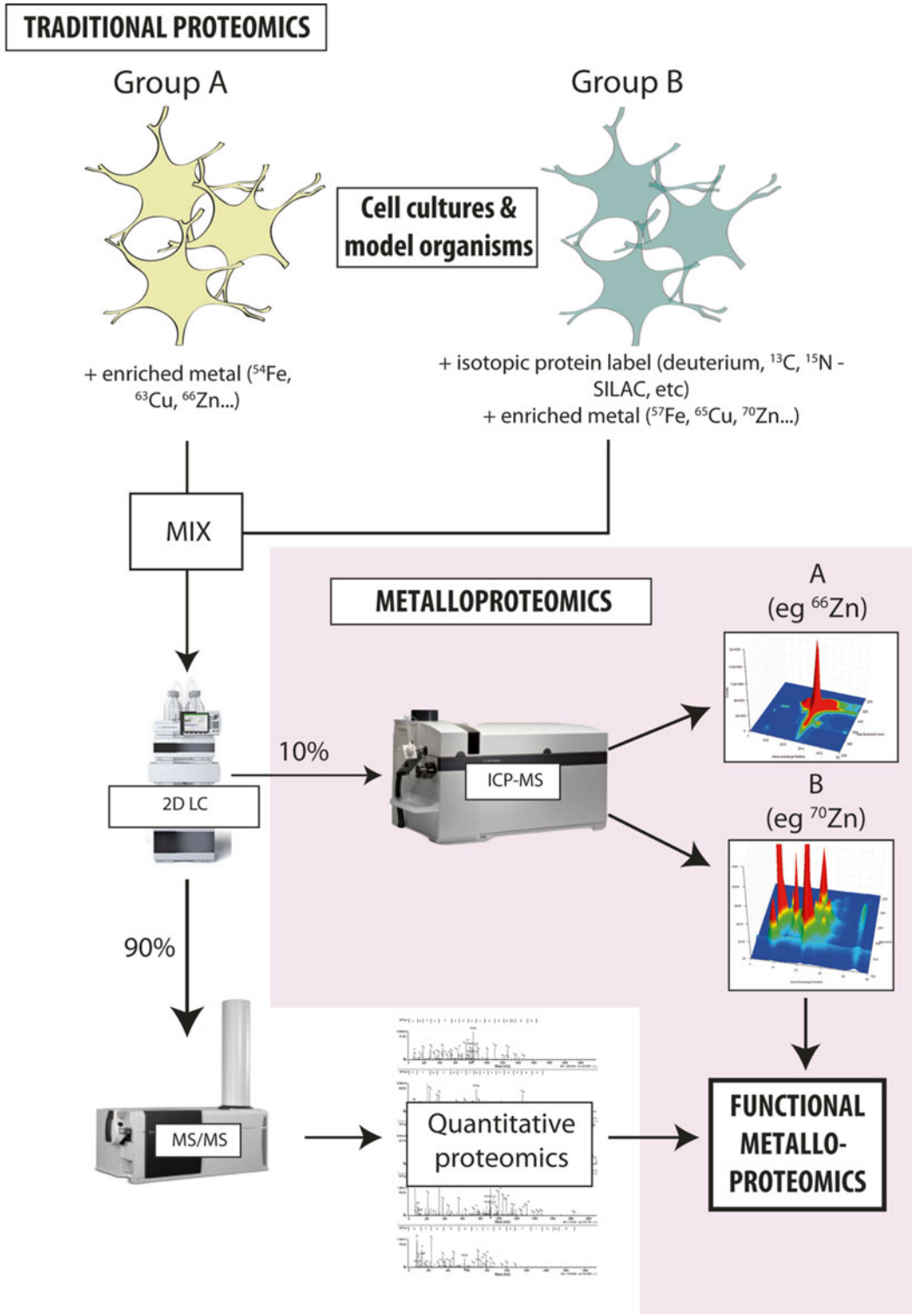


Fig. 1 Proposed workflow for functional metalloproteomics using stable isotope labeling, modeled on existing workflows for traditional proteomics. Cell cultures or model organisms are divided into two experimental groups, which are exposed to isotopically enriched metals. The second group is also labeled with enriched amino acids (such as stable isotope labeling by amino acids in cell culture, or SILAC). Soluble protein fractions

important in AD research, not only from the perspective of Aβ-centric studies, but all metalloproteins potentially associated with the disease state.

Relatively weak interactions between metals and their chaperoning proteins present a complex analytical challenge that has previously restricted rapid uptake of metalloproteomics by the wider systems biology community. Separation mechanisms typical of traditional proteomics employ denaturing conditions that disrupt the predominantly multi-dentate ligands between metal ions and amino acid residues that rely heavily upon protein conformation [47]. Reverse-phase liquid chromatography (RPLC), which demonstrates a superior number of plates compared to most native separation techniques, also has several major incompatibilities with online ICP-MS detection. Organic mobile phases used to separate biomolecules on the basis of polarity cannot be directly infused into an argon plasma without the use of an oxygen gradient system to compensate for plasma instability [48]. When appropriate measures are taken to reduce the carbon load on the plasma imparted by RPLC, the technique is of particular use in studying covalently bound heteroatoms, such as sulfur and phosphorus [49]. Miniaturized separation devices employing low flow rates have great potential to improve sensitivity and resolving power for heteroatom speciation by eliminating the need for the addition of oxygen to the plasma, though reliable interfacing of low flow eluent to the standard ICP-MS system is very much still in development. Consequently, the preferred native separation techniques applicable to ICP-MS include size exclusion chromatography (SEC), which has generally poor resolving power, and ion-exchange chromatography, which often required high salt buffers [50]. Even relatively inert buffers used for SEC, including Tris, Hepes and Mops-buffers appear to impart some degree of metal redistribution in biological samples [51]. Buffer components with little to no complexing capacity, such as NH_4NO_3 are thus preferable, with the added advantage that no post column modification (such as desalting) is necessary prior to hyphenation to an ICP-MS. Two-dimensional chromatography, where multimodal separations are performed in tandem, can dramatically increase the resolving power of a chromatographic experiment. Fractionation of protein species in the first dimension on the basis of size, isoelectric point or polarity, followed by a further separation based on an alternative criteria is all theoretically possible for ICP-MS detection, though the aforementioned metal-protein interactions are best preserved by a sequence of native chromatography, such as SEC and anion-exchange (AEX) [52]. Two-dimensional fractionation can also be used as a preparatory step for gel electrophoresis and peptide mass fingerprinting, though at the expense of metal-protein bond information.

4 Application of Metalloproteomics to Alzheimer's Disease Research

Considering the wealth of information available regarding the association of metals and Aβ, relatively little literature exists applying metalloproteomic techniques to the direct study of this interaction. ICP-MS' place in AD research has previously been as a detector for bulk assay of total metal content in systems investigating in vitro interactions between Aβ and metal species and potential therapeutics intervening in this relationship [53]. Aβ's relatively small size, combined with minor difference in mass between the proposed toxic $A\beta_{1-42}$ and shorter peptide lengths present a difficult analytical challenge for current LC-ICP-MS protocols.

However, more tractable targets for metalloproteomic study focused instead on possible reactive proteins with metal cofactors may bear more fruit. One recurring use of LC-ICP-MS technology has been the application to the study of metallothioneins (MT), which are cysteine-rich, low mass (approx. 6 kDa) proteins with antioxidant properties and are involved in group 11 and 12 metal regulation [54]. It is perhaps not coincidental that a protein that readily binds both Zn(II) and Cu(I) may thus be involved in potential Aβ-metal interactions downstream of the event. Four distinct isoforms of MT are known to exist in mammals, of which MT-1-MT-3 have been found in the central nervous system. MT-3 has been of significant interest since it was first discovered to be depleted in the human AD brain [55], a finding that has since brought about much debate. The metal-thiolate clusters forming protein-metal bonds in MTs have been suggested to be robust enough to survive reverse phase LC separation [56–58], though separation in non-native conditions cannot ensure that any metal-protein interaction observed is exclusively physiological. Off-line ICP-MS analysis of CSF fractions separated by SEC has been shown capable of detecting endogenous MT isoforms retaining their metal content [59]. Again, very little literature is available describing the analysis of MTs in AD. Prange et al. [60] reported one of the few examples, obtaining impressive resolution by using native capillary zone electrophoresis hyphenated to ICP-MS to compare MT levels in human parietal, occipital and temporal regions from AD patients and age-matched controls. Decreased expression of MT-1 and MT-3 was observed in both the temporal and occipital regions in AD brains, though the small sample size should be considered for interpretation. The same group used dithiothreitol (DTT) to stabilize MTs in brain homogenates, reporting that DTT prevents MT oxidation during sample preparation and assists in retaining MT-metal bonds, a reaction that was suggested to be reversible. Interestingly, it was reported that although both AD and control brains appeared to contain similar MT levels, addition of DTT restored metal levels to the AD

samples, suggesting that MTs are more susceptible to oxidative damage in AD [61]. This poses an interesting question regarding the interpretation of changes to MT expression in AD: are variations in total metal levels truly indicative of a deficiency or overexpression of a protein, or instead a symptom of protein confirmation changes that impair metal binding? Further, this also raises important issues regarding post mortem modification of metal-protein associations, and highlights the importance that native sample preparation and separation techniques in ensuring physiological features of disease are retained.

Heteroatom speciation, where the multi-element detection capabilities of ICP-MS are extended to non-metals such as sulfur and phosphorus has proved useful in emerging AD research, due mostly to its compatibility with higher resolution, yet denaturing separation mechanisms. Solid sampling, via laser ablation (LA) extends the versatility of ICP-MS as an in situ analytical technique of SDS-PAGE separation. Early examples of phosphorus quantification in 2D-PAGE separations of cytosols from human AD were more speculative than informative [62, 63], but did provide an interesting application of LA-ICP-MS to protein phosphorylation analysis. A similar approach employing 1D PAGE was used to study the uptake of copper and zinc isotopic tracers by tau protein, with a noteworthy deviation from the natural abundance ratio of $^{63}Cu/^{65}Cu$ observed corresponding to specific tau isoforms, confirmed by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) of tryptic digests taken at each band [64]. The authors noted that no metal-containing peptide was identified in the FTICR mass spectra, though it is important to note that tryptic digestion of the protein bands most likely stripped any metal from the tau isoforms.

As noted previously, several significant limitations stand in front of the widespread uptake of metalloproteomics by the systems biology discipline. However, these shortcomings are not insurmountable, and a great deal of advanced analytical research is helping to overcome these problems. Primarily, the major limitation remains a perceived lack of resolving power in native chromatographic conditions. This does not preclude techniques such as SEC-ICP-MS from worthwhile use in the AD research community; rather this technique provides a rapid and highly sensitive method for studying the metal-binding properties in well-defined proteins that have been associated with AD pathogenesis, such as hemoglobin [65]. Additionally, the application of two-dimensional separations to metalloprotein assay (such as that shown in Fig. 1) will greatly improve the resolution of the native chromatography used and exploit the high sensitivity of ICP-MS for probing low abundant metalloproteins. 2D separations using reverse-phase dimensions, used in tandem with electrospray MS have already been used to study trace species in biological matrices [66, 67],

and there is little preventing these methods being applied to brain homogenates. In fact, Barnett et al. [68] has utilized an impressive combination of chromatographic fractionation techniques, including immobilized metal affinity chromatography to study the metalloproteome of a marine cyanobacterium. This example, as well as the others outlined in the review from Barnett and co-workers [52] demonstrates that many of the technical problems associated with trace metal speciation are gradually being overcome, and this pace will only increase as the capabilities of ICP-MS within the systems biology laboratory are recognized. Miniaturized separation devices that currently feature predominately in the proteomics marketplace [69] will eventually be interfaced with ICP-MS frontends, allowing for rapid and reproducible separations on microfluidic columns. Design and application of low flow total consumption nebulizers will also allow use of capillary and nano-flow chromatography.

References

1. Mattson MP (2004) Pathways towards and away from Alzheimer's disease. Nature 430: 631–639
2. Tanzi RE, Bird ED, Latt SA, Neve RL (1987) The amyloid beta protein gene is not duplicated in brains from patients with Alzheimer's disease. Science 238:666–669
3. Clark CM, Ewbank D, Lee VM-Y, Trojanowski JQ (1998) Molecular pathology of Alzheimer's disease: neuronal cytoskeletal abnormalities. In: Growdon JH, Rossor MN (eds) The dementias Vol 19 of Blue books of practical neurology. Butterworth–Heinemann, Boston, pp. 285–304
4. Migliore L, Coppede F (2009) Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases. Mutat Res 667:82–97
5. Armstrong RA (2013) What causes Alzheimer's disease? Folia Neuropathol 51:169–188
6. Crapper DR, Krishnan SS, Dalton AJ (1973) Brain aluminum distribution in Alzheimer's disease and experimental neurofibrillary degeneration. Science 180:511–513
7. Frausto da Silva JJR, Williams RJP (2001) The biological chemistry of the elements. Oxford University Press, Oxford
8. Guengerich FP (2009) Thematic minireview series: metals in biology. J Biol Chem 284: 18557
9. Eskici G, Axelsen PH (2012) Copper and oxidative stress in the pathogenesis of Alzheimer's disease. Biochemistry 51:6289–6311
10. Duce JA, Bush AI (2010) Biological metals and Alzheimer's disease: implications for therapeutics and diagnostics. Prog Neurobiol 92:1–18
11. Hung YH, Bush AI, Cherny RA (2010) Copper in the brain and Alzheimer's disease. J Biol Inorg Chem 15:61–76
12. Que EL, Domaille DW, Chang CJ (2008) Metals in neurobiology: probing their chemistry and biology with molecular imaging. Chem Rev 108:1517–1549
13. Kepp KP (2012) Bioinorganic chemistry of Alzheimer's disease. Chem Rev 112: 5193–5239
14. Sensi SL, Paoletti P, Bush AI, Sekler I (2009) Zinc in the physiology and pathology of the CNS. Nat Rev Neurosci 10:780–791
15. Squitti R (2012) Metals in Alzheimer's disease: a systemic perspective. Front Biosci 17: 451–472
16. Adlard PA, Bush AI (2006) Metals and Alzheimer's disease. J Alzheimers Dis 10: 145–163
17. Collingwood J, Dobson J (2006) Mapping and characterization of iron compounds in Alzheimer's tissue. J Alzheimers Dis 10: 215–222
18. Deibel MA, Ehmann WD, Markesbery WR (1996) Copper, iron, and zinc imbalances in severely degenerated brain regions in Alzheimer's disease: possible relation to oxidative stress. J Neurol Sci 143:137–142
19. Magaki S, Raghavan R, Mueller C et al (2007) Iron, copper, and iron regulatory protein 2 in Alzheimer's disease and related dementias. Neurosci Lett 418:72–76
20. Religa D, Strozyk D, Cherny RA et al (2006) Elevated cortical zinc in Alzheimer disease. Neurology 67:69–75

21. Lovell MA, Robertson JD, Teesdale WJ et al (1998) Copper, iron and zinc in Alzheimer's disease senile plaques. J Neurol Sci 158: 47–52
22. Miller LM, Wang Q, Telivala TP et al (2006) Synchrotron-based infrared and X-ray imaging shows focalized accumulation of Cu and Zn co-localized with beta-amyloid deposits in Alzheimer's disease. J Struct Biol 155:30–37
23. Bush AI, Tanzi RE (2008) Therapeutics for Alzheimer's disease based on the metal hypothesis. Neurotherapeutics 5:421–432
24. Valensin D, Mancini FM, Luczkowski M et al (2004) Identification of a novel high affinity copper binding site in the APP(145–155) fragment of amyloid precursor protein. Dalton Trans 2004:16–22
25. Simons A, Ruppert T, Schmidt C et al (2002) Evidence for a copper-binding superfamily of the amyloid precursor protein. Biochemistry 41:9310–9320
26. Hesse L, Beher D, Masters CL, Multhaup G (1994) The beta A4 amyloid precursor protein binding to copper. FEBS Lett 349:109–116
27. Barnham KJ, McKinstry WJ, Multhaup G et al (2003) Structure of the Alzheimer's disease amyloid precursor protein copper binding domain. A regulator of neuronal copper homeostasis. J Biol Chem 278:17401–17407
28. Atwood CS, Scarpa RC, Huang X et al (2000) Characterization of copper interactions with Alzheimer amyloid beta peptides: identification of an attomolar-affinity copper binding site on amyloid beta1-42. J Neurochem 75: 1219–1233
29. Armendariz AD, Gonzalez M, Loguinov AV, Vulpe CD (2004) Gene expression profiling in chronic copper overload reveals upregulation of Prnp and App. Physiol Genomics 20:45–54
30. Bellingham SA, Lahiri DK, Maloney B et al (2004) Copper depletion down-regulates expression of the Alzheimer's disease amyloid-beta precursor protein gene. J Biol Chem 279:20378–20386
31. Lin R, Chen X, Li W et al (2008) Exposure to metal ions regulates mRNA levels of APP and BACE1 in PC12 cells: blockage by curcumin. Neurosci Lett 440:344–347
32. Angeletti B, Waldron KJ, Freeman KB et al (2005) BACE1 cytoplasmic domain interacts with the copper chaperone for superoxide dismutase-1 and binds copper. J Biol Chem 280:17930–17937
33. Das HK, Baez ML (2008) ADR1 interacts with a down-stream positive element to activate PS1 transcription. Front Biosci 13: 3439–3447
34. Hoke DE, Tan JL, Ilaya NT et al (2005) In vitro gamma-secretase cleavage of the Alzheimer's amyloid precursor protein correlates to a subset of presenilin complexes and is inhibited by zinc. FEBS J 272:5544–5557
35. Allinson TM, Parkin ET, Turner AJ, Hooper NM (2003) ADAMs family members as amyloid precursor protein alpha-secretases. J Neurosci Res 74:342–352
36. Maynard CJ, Cappai R, Volitakis I et al (2002) Overexpression of Alzheimer's disease amyloid-beta opposes the age-dependent elevations of brain copper and iron. J Biol Chem 277:44670–44676
37. Phinney AL, Drisaldi B, Schmidt SD et al (2003) In vivo reduction of amyloid-beta by a mutant copper transporter. Proc Natl Acad Sci U S A 100:14193–14198
38. Bellingham SA, Ciccotosto GD, Needham BE et al (2004) Gene knockout of amyloid precursor protein and amyloid precursor-like protein-2 increases cellular copper levels in primary mouse cortical neurons and embryonic fibroblasts. J Neurochem 91:423–428
39. White AR, Reyes R, Mercer JF et al (1999) Copper levels are increased in the cerebral cortex and liver of APP and APLP2 knockout mice. Brain Res 842:439–444
40. Evans GA (2000) Designer science and the "omic" revolution. Nat Biotechnol 18:127
41. Weinstein JN (1998) Fishing expeditions. Science 282:627
42. Haraguchi H (2004) Metallomics as integrated biometal science. J Anal Atom Spectrom 19:5–14
43. Szpunar J (2004) Metallomics: a new frontier in analytical chemistry. Anal Bioanal Chem 378:54–56
44. Dudev T, Lim C (2003) Principles governing Mg, Ca, and Zn binding and selectivity in proteins. Chem Rev 103:773–788
45. Cvetkovic A, Menon AL, Thorgersen MP et al (2010) Microbial metalloproteomes are largely uncharacterized. Nature 466:779–782
46. Lothian A, Hare DJ, Grimm R et al (2013) Metalloproteomics: principles, challenges and applications to neurodegeneration. Front Aging Neurosci 5:35
47. Yamashita MM, Wesson L, Eisenman G, Eisenberg D (1990) Where metal ions bind in proteins. Proc Natl Acad Sci U S A 87: 5648–5652
48. Meermann B, Kießhauer M (2011) Development of an oxygen-gradient system to overcome plasma instabilities during HPLC/ICP-MS measurements using gradient elution. J Anal Atom Spectrom 26:2069–2075

49. Szpunar J (2005) Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst 130:442–465
50. Szpunar J (2000) Bio-inorganic speciation analysis by hyphenated techniques. Analyst 125:963–988
51. Jahromi EZ, White W, Wu Q et al (2010) Remarkable effect of mobile phase buffer on the SEC-ICP-AES derived Cu, Fe and Zn-metalloproteome pattern of rabbit blood plasma. Metallomics 2:460–468
52. Barnett JP, Scanlan DJ, Blindauer CA (2012) Protein fractionation and detection for metalloproteomics: challenges and approaches. Anal Bioanal Chem 402:3311–3322
53. Choi J-S, Braymer JJ, Nanga RP et al (2010) Design of small molecules that target metal-Aβ species and regulate metal-induced Aβ aggregation and neurotoxicity. Proc Natl Acad Sci U S A 107:21990–21995
54. Stillman MJ (1995) Metallothioneins. Coord Chem Rev 144:461–511
55. Uchida Y, Takio K, Titani K et al (1991) The growth inhibitory factor that is deficient in the Alzheimer's disease brain is a 68 amino acid metallothionein-like protein. Neuron 7: 337–347
56. Ferrarello C, Fernández de la Campa MR, Sanz-Medel A (2002) Multielement trace-element speciation in metal-biomolecules by chromatography coupled with ICP-MS. Anal Bioanal Chem 373:412–421
57. Lobinski R, Chassaigne H, Szpunar J (1998) Analysis for metallothioneins using coupled techniques. Talanta 46:271–289
58. Chassaigne H, Lobinski R (1998) Characterization of horse kidney metallothionein isoforms by electrospray MS and reversed-phase HPLC-electrospray MS. Analyst 123: 2125–2130
59. Gellein K, Roos PM, Evje L et al (2007) Separation of proteins including metallothionein in cerebrospinal fluid by size exclusion HPLC and determination of trace elements by HR-ICP-MS. Brain Res 1174:136–142
60. Prange A, Schaumlöffel D, Brätter P et al (2001) Species analysis of metallothionein isoforms in human brain cytosols by use of capillary electrophoresis hyphenated to inductively coupled plasma-sector field mass spectrometry. Fresenius J Anal Chem 371: 764–774
61. Richarz A-N, Bratter P (2002) Speciation analysis of trace elements in the brains of individuals with Alzheimer's disease with special emphasis on metallothioneins. Anal Bioanal Chem 372:412–417
62. Becker JS, Zoriy M, Przybylski M, Becker JS (2007) High resolution mass spectrometric brain proteomics by MALDI-FTICR-MS combined with determination of P, S, Cu, Zn and Fe by LA-ICP-MS. Int J Mass Spectrom 261:68–73
63. Becker JS, Zoriy M, Becker JS et al (2004) Determination of phosphorus and metals in human brain proteins after isolation by gel electrophoresis by laser ablation inductively coupled plasma source mass spectrometry. J Anal Atom Spectrom 19:149–152
64. Becker JS, Zoriy M, Przybylski M, Sabine Becker J (2007) Study of formation of Cu- and Zn-containing tau protein using isotopically-enriched tracers by LA-ICP-MS and MALDI-FTICR-MS. J Anal Atom Spectrom 22:63–68
65. Chuang J-Y, Lee C-W, Shih Y-H et al (2012) Interactions between amyloid-β and hemoglobin: implications for amyloid plaque formation in Alzheimer's disease. PLoS One 7:e33120
66. Garcia-Sartal C, Taebunpakul S, Stokes E et al (2012) Two-dimensional HPLC coupled to ICP-MS and electrospray ionisation (ESI)-MS/MS for investigating the bioavailability in vitro of arsenic species from edible seaweed. Anal Bioanal Chem 402:3359–3369
67. Połatajko A, Banaś B, Encinar JR, Szpunar J (2005) Investigation of the recovery of selenomethionine from selenized yeast by two-dimensional LC–ICP MS. Anal Bioanal Chem 381:844–849
68. Barnett JP, Scanlan DJ, Blindauer CA (2012) Fractionation and identification of metalloproteins from a marine cyanobacterium. Anal Bioanal Chem 402:3371–3377
69. Yin H, Killeen K (2007) The fundamental aspects and applications of Agilent HPLC-Chip. J Sep Sci 30:1427–1434

Chapter 23

Redox Proteomics in Human Biofluids: Sample Preparation, Separation and Immunochemical Tagging for Analysis of Protein Oxidation

Fabio Di Domenico, Marzia Perluigi, and D. Allan Butterfield

Abstract

Proteomics offers the simultaneous detection of a large number of proteins in a single experiment and can provide important information regarding crucial aspects of specific proteins, particularly post-translational modifications (PTMs). Investigations of oxidative PTMs are currently performed using focused redox proteomics techniques, which rely on gel electrophoresis separations of intact proteins with the final detection of oxidative PTMs being performed by mass spectrometry (MS) analysis. The application of this technique to human biofluids is being subject of increasing investigation and is expected to provide new insights on the oxidative status of the peripheral proteome in neurological diseases such as Alzheimer's disease, towards purposes of early diagnosis and prognosis. This chapter describes all the experimental steps to perform redox proteomics analysis of cerebrospinal fluid and plasma/serum samples.

Key words Redox proteomics, Oxidative stress, Protein oxidation, Cerebrospinal fluid, Plasma, Serum, Diagnosis, Prognosis, Neurological disorders, Alzheimer's disease

1 Introduction

In the last decade, much effort has been given to the development of new sophisticated proteomics platforms for the screening and assessment of protein changes, both qualitative and quantitative, which may correlate with disease pathogenesis and progression. Different proteomics approaches have been proposed, among which redox proteomics is specifically dedicated to analyse irreversible oxidative modifications of proteins caused by oxidative stress, a well-established condition associated to many diseases [1, 2]. This experimental approach offers a wide spectrum of information that may shed light on the mechanisms underlying a complex disease, allow the identification of putative disease-associated markers and targets for specific therapeutic interventions. To achieve this goal, growing research is focused on establishing a direct link between tissue specific damage and systemic alteration,

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_23,

as well as in identifying biochemical markers that can be measured in body fluids such as cerebrospinal fluid (CSF), plasma or serum.

Cerebrospinal fluid is the superior biofluid to investigate the neurological status of a patient. Moreover, its relative availability makes possible to conduct longitudinal studies, molecular analyses of changes in CSF during the course of a disease. Indeed, CSF contains direct, valuable biochemical information that can support effective clinical management of brain disorders. The difficulty of using CSF however, resides in the complications and side effects the patients might encounter during sample collection by lumbar puncture.

Compared with CSF, the analysis of blood-related fluids requires less invasive procedures thereby representing an ideal strategy for diagnosis, prognosis and monitoring of therapeutic protocols, especially for large-scale studies with continued, repeated measurements. However, the profiling of the human plasma proteome is a difficult task due to the complexity of its protein composition. The most abundant protein fraction (albumin, immunoglobulins (Igs) among others), accounts for approximately 85 % of the total protein content, whilst other proteins, components of the low-abundance fraction, are present in a wide dynamic range of protein concentrations, of more than ten orders of magnitude [3].

To overcome such complexity and result in reliable and reproducible results, the analysis of the plasma proteome requires a multidimensional approach. First, in order to increase the number of reliable, detectable proteins by proteomic screening, the depletion of the most abundant plasma proteins can be set as a pre-requisite step. Since this step can also affect the detection of the low-abundance fraction that might interact with the high abundant proteins [4], the strategy can be optimized according to different biochemical and biophysical features of specific proteins, including molecular weight, hydrophobicity and isoelectric point. Among different techniques routinely utilized, the most common strategy today relies on antibody-based retention of a chosen set of the most abundant proteins. This pre-fractionation strategy has been validated by several studies comparing the various depletion protocols currently available. To date, best results, improved proteomics separation and quantitation are achieved after removal of high-abundant proteins [5].

Numerous studies have reported the identification of oxidized proteins in several neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson disease (PD), and amyotrophic lateral sclerosis (ALS), among others. Redox proteomics analysis of post-mortem brain from AD and mild cognitive impairment (MCI) subjects, have demonstrated the presence of oxidative modifications

of several proteins involved in energy metabolism, antioxidant defense, proteasome function, neuronal communication, cytoskeletal integrity [1]. These results have revealed the deregulation of selective intracellular pathways at the brain level and how these deregulations possibly translate into clinical symptoms.

From here, the next big challenge is represented by the application of redox proteomics to biological fluids (CSF, blood), for the identification of peripheral oxidized proteins potentially involved in the pathogenesis and progression of the disease. The redox proteomics approach combines two-dimensional gel electrophoresis with immunochemical detection of oxidized proteins followed by mass spectrometry for protein identification (Fig. 1). Proteins containing reactive carbonyl groups/3-NT/HNE are detected by 2D-western blot analysis using specific antibodies that recognize each of the above protein modifications [3]. The mass spectrometric analysis uses two different approaches that include peptide mass fingerprinting (PMF) using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry, and sequence tags using nano-electrospray ionization tandem mass spectrometry (nano-ESI-MS/MS). The identification of a protein is performed according to well curated protein databases [1, 3, 6].

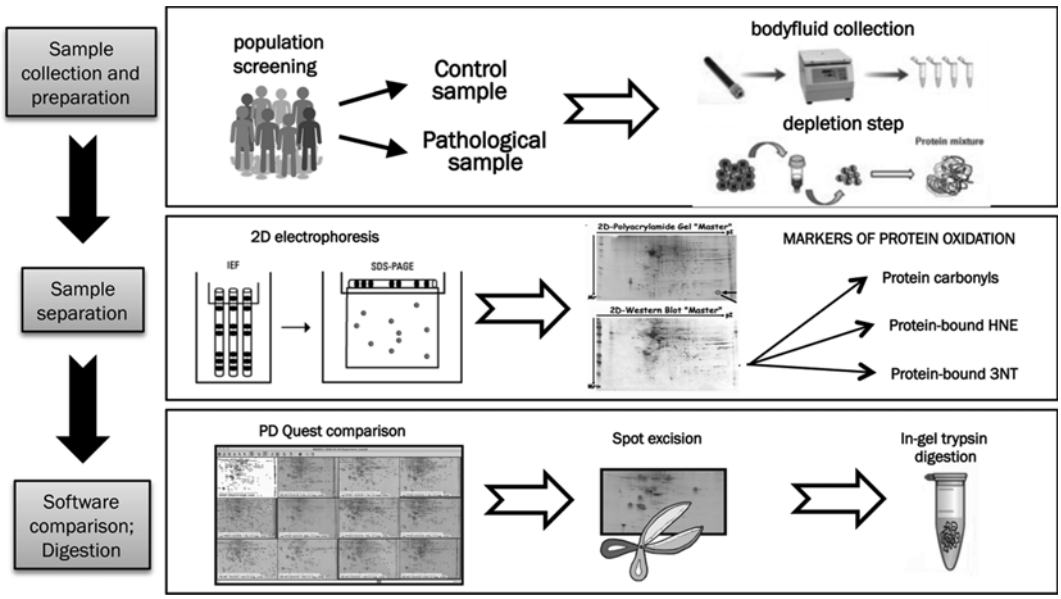


Fig. 1 Schematic overview of redox proteomics techniques applied to human biological fluids. Peptides obtained after digestion and clean-up steps are submitted to mass spectrometry search engines and databases for protein identification [1, 3, 6]

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain an electrical sensitivity of 18 MΩ cm at 25 °C).

2.1 Depletion Kits

1. ProteoPrep Blue albumin and IgGs depletion kit (Sigma-Aldrich).
2. Human 14 Multiple Affinity Removal System (MARS) spin cartridges for the depletion of high-abundant proteins from human proteomic samples (Agilent Technologies) (*see* **Note 1**).

2.2 Sample Preparation and Iso-electrofocusing (IEF)

1. 0 % acetone (dimethyl ketone).
2. Rehydration buffer (RB): 8 M urea, 2 M thiourea, 2 % CHAPS, 20 mM DTT, 0.2 % biolytes, bromophenol blue in water.
3. Equilibration buffer (EB): 6 M urea, 30 % glycerol, 1 % SDS, 50 mM Tris–HCl, pH 6.8.
   - EB with DTT: Weigh 200 mg of DTT and add 10 mL of EB.
   - EB with iodoacetamide (IA): Weight 250 mg of IA. Dissolve in 10 mL of EB.
4. Immobilized pH gradient (IPG) strip, "Ready strip", for IEF: 11 cm, pH 4–7 or 3–11, linear (Bio-Rad).
5. Protean IEF Cell apparatus (Bio-Rad).
6. Mineral oil.

2.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Criterion™ Tris–HCl Precast Gels: 12 % Criterion IPG +1 well, or Any kD™ (Bio-Rad) (*see* **Note 2**).
2. Agarose (0.5 %).
3. Running buffer, 1× Tris-glycine SDS: 25 mM Tris base, 190 mM glycine, 0.1 % SDS, pH 8.3.
4. Fixing solution: 40 % methanol, 10 % acid acetic in water.
5. Gel staining solutions: Coomassie Blue R-250, 40 % methanol, 1 % acetic acid, or Sypro Ruby protein gel stain (Bio-Rad) (*see* **Note 3**).
6. Nitrocellulose membrane 0.2 μm; blotting paper.
7. Transfer buffer, 1× Tris-glycine: 25 mM Tris base, 190 mM glycine, 20 % methanol, pH 8.3.

2.4 Immunochemical Detection of Oxidized Proteins

1. Tris-buffered saline tween (TBS)-T washing buffer (1×): 25 mM Tris base, 150 mM NaCl, 2 mM KCl, pH 7.4. Mix 900 mL of this TBS solution with 100 mL of Tween 20 to make 1 L of TBS-T.
2. Blocking solution: 3 % bovine serum albumin (BSA) in TBS-T.

3. Primary antibodies: anti-2,4-dinitrophenylhydrazone (DNPH), anti-DNPH (Millipore); antiprotein bound-4-hydroxynonenal (HNE), anti-HNE (alpha diagnostic international); anti-3-nitrotyrosine (3NT), anti-3NT (Millipore).
4. Post derivatization reagents: 20 % methanol washing buffer; 2 N HCl solution; 10 mM DNPH solution; 50 % methanol washing buffer.
5. Secondary antibodies: Anti-mouse or anti-rabbit IgG alkaline phosphatase (AP) conjugated (Sigma-Aldrich).
6. Colorimetric reagents for membrane staining: (A) 50 mg/mL 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) stock solution in 100 % dimethylformamide (DMF), for colorimetric detection of alkaline phosphatase activity. Store at −20 °C; (B) 100 mg/mL Nitroblue Tetrazolium Chloride (NBT) stock solution in 70 % DMF, detecting alkaline phosphatase activity. Store at −20 °C; (C) Alkaline phosphatase (ALP) buffer: 0.1 M Tris, 0.1 M NaCl, 5 mM $MgCl_2$, pH 9.5. Store at 4 °C.

2.5 Image Analyses Equipment and Tools

1. GS 800 densitometer (Bio-Rad) or equivalent, working in reflectance and transmittance modes for the acquisition of gel images stained with Coomassie blue and membranes developed with alkaline phosphatase stains (e.g. molecular imagers such as Chemidoc XP (Bio-Rad) or Typhoon FLA 7000/9500 (GE Healthcare)), and laser-based or UV imaging systems to acquire gel images developed with Sypro Ruby stain (*see* **Note 3**).
2. PD-Quest 2D-analysis software (Bio-Rad) or equivalent for gel and blot quantitative analyses (*see* **Note 4**).

2.6 Trypsin Digestion

1. 0.1 M NH_4HCO_3 solution: 0.079 g of NH_4HCO_3 in 10 mL of water.
2. DTT solution: 0.00308 g of DTT in 1 mL of 0.1 M NH_4HCO_3.
3. 100 % acetonitrile (AcN).
4. Iodoacetamide (IA) solution: 0.010175 g of IA in 1 mL of 0.1 M NH_4HCO_3.
5. Trypsin stock solution: 1 μg/μL in 50 mM acetic acid.
6. Trypsin ready-to-use solution: 20 μg/mL in 40 mM NH_4HCO_3, 10 % acetonitrile.
7. SpeedVac vacuum concentrator.
8. Fume hood.

2.7 Peptide Extraction and Clean-up

1. Buffer A: 5 % acetonitrile, 0.1 % formic acid.
2. Buffer B: 95 %, acetonitrile, 0.1 % formic acid, 0.001 M NH_4CO_3.
3. Buffer C: 100 % acetonitrile.

4. Buffer D: 50 % acetonitrile, 0.1 % formic acid.
5. ZipTip C18 pipette tips (Millipore) with a bed of chromatography media fixed at its end for concentrating and purifying samples (*see* **Note 5**).
6. Sonication bath.

3 Methods

When performing redox proteomics each sample is analyzed twice to obtain the protein expression profile (gel staining method) and the protein oxidation profile (membrane staining protocol) (Fig. 1). Oxidation values are then normalized to expression values to obtain specific protein oxidation values per amount of protein.

3.1 Sample Preparation

Carry out all procedures at room temperature unless otherwise specified.

3.1.1 Cerebrospinal Fluid (CSF) Sample (See Note 6)

1. From 200 μL of CSF sample, precipitate protein with 600 μL of acetone overnight at −20 °C (*see* **Note 7**).
2. Spin-dry the sample in a microfuge at 4 °C. 5,000 rpm (1400 × *g*) for 10 min.
3. Discard the supernatant and air dry the pellet for 15 min.
4. Re-suspend the pellet in 200 μL of RB buffer and stir the sample for at least 2 h to solubilize the sample.
5. Sonicate the sample until its complete solubilization (*see* **Note 8**).

3.1.2 Plasma/Serum Sample

1. Deplete high abundant proteins from the sample using the selected depletion kit (*see* **Note 1**).
2. Mix the sample (150–200 μg of protein) with RB buffer (up to 200 μL). Stir the sample for a minimum of 2 h (*see* **Note 9**).
3. Sonicate the sample until its complete solubilization (*see* **Note 8**).

3.2 Isoelectrofocusing (IEF). Separation by Isoelectric Point (pI)

1. Pipette the sample along one of the lanes of the IEF tray (*see* **Note 10**).
2. Apply the strip with the gel facing the sample.
3. Leave gel strip for 45 min at room temperature to let the gel absorb the sample (*see* **Note 11**).
4. Add mineral oil on the strip lane (*see* **Note 12**).
5. Start the active rehydration program on the protean IEF cell apparatus for 17 h.
6. Start the IEF running program: 300 V for 2 h linearly (linear gradient); 500 V for 2 h linearly; 1,000 V for 2 h linearly, 8,000 V for 8 h linearly, and 8,000 V for 10 h at rapid gradient.

7. At the end of the program the strip can be stored at –80 °C until second dimension separation (SDS-PAGE) is performed (*see* **Note 13**).

3.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE). Separation by Size

1. Incubate the gel strip in equilibrium buffer with DTT (EB with DTT) for 15 min. Wash with running buffer. Incubate again the gel strip, this time with EB buffer with IA for 15 min.
2. Place the gel strip in the precast Criterion Tris–HCl gel to perform the second dimension electrophoresis.
3. Use 0.5 % agarose to fix/block the strip in the plastic support of the precast gel.
4. Run the gel into the electrophoresis apparatus at 200 V for 60 min (*see* **Note 14**).
5. For gel staining: Fix the proteins into the gel with fixing solution for 45 min. Stain the gel with Comassie Blue for 90 min, or Sypro Ruby stain for 24 h, then rinse with water (*see* **Note 4**).

3.4 Transfer of Proteins onto Nitrocellulose Membranes

1. For blot analysis, proteins from gels are transferred onto nitrocellulose membranes (*see* **Note 15**) using a semi-dry method at 45 mA per gel for 2 h (*see* **Note 16**).
2. Incubate the nitrocellulose membrane with 3 % BSA in TBS-T for at least 90 min.

3.5 Post-derivatization of Membranes for Protein Carbonyl Detection (See Note 17)

1. Put the membrane in a solution of 20 % methanol washing buffer at room temperature for 5 min.
2. Incubate membrane in a solution of 2 N HCl for 5 min.
3. Incubate membrane in 1 mM DNPH solution for 10 min (*see* **Note 18**).
4. Wash membrane with 2 N HCl for three times for 5 min (*see* **Note 19**).
5. Wash in 50 % methanol washing buffer five times for 5 min (*see* **Note 19**).
6. Wash with washing blot for three times for 5 min and block with BSA solution (*see* **Note 19**).

3.6 Immunochemical Detection of Oxidized Proteins

1. Add primary antibody to the membrane (*see* **Note 20**).
   - For protein carbonyls analysis add anti-DNPH (Millipore) to the post-derivatized membrane.
   - For protein-bound HNE analysis add anti-HNE antibody (alpha diagnostic international).
   - For protein-bound 3-NT analysis add anti-3NT antibody (Millipore).

2. Wash three times (10 min each) and add secondary antibody AP conjugated (*see* **Note 21**).
3. Prepare fresh BCIP/NBT solution adding 37 μL of BCIP stock solution and 50 μL of NBT stock solution to 10 mL of ALP buffer.
4. Incubate the membrane with fresh BCIP/NBT solution until staining development (*see* **Note 22**). Rinse with water to stop the staining reaction.

3.7 Image Analyses

1. Scan the images of gels and membranes and save in the appropriate format (*see* **Note 23**).
2. Load the images on the PD-quest software and perform these steps in the following order:
   - Matching among all gels. Identify a master gel and make the appropriate selection of the faintest and the smallest spots, and a large representative section of the image containing spots, streaks and background gradations to make corrections/filtering for background noise. Molecular mass and isoelectric point of the major spots can be automatically determined by bilinear interpolation between most relevant features on each image; PD-Quest software performs the auto-matching, and images are manually edited to confirm proper spot detection and matching.
   - Matching among all blots. Follow the instructions above.
   - Matching between the master gel and the master blot. Match manually the master gel and the master blot according with the auto-matching performed by the software (*see* **Note 24**).
3. After the matching steps, spot quantitation can be performed using the image software, normalizing the value of the oxidation (from membrane analysis) to the value of protein expression (from gel analysis) for every single matched spot (*see* **Note 25**).
4. Perform statistical analyses to identify the spots with significantly different patterns of oxidation among the groups of comparison (*see* **Note 26**).

3.8 Protein Digestion with Trypsin (See Note 27)

1. Cut the spot from the gel using a tip or a scalpel and transfer the gel piece into a 1.5 mL tube (*see* **Note 28**).
2. Add 10–20 μL of 0.1 M NH_4HCO_3 and incubate the sample at room temperature for 15 min in the hood.
3. Add 15–30 μL of acetonitrile and incubate at room temperature for 15 min.
4. Remove NH_4HCO_3/acetonitrile with pipette and let gel dry for 30 min.

5. Add 20–30 μL of DTT per tube and incubate at 56 °C for 45 min.
6. Remove DTT and add 20–30 μL of IA. Incubate at room temperature for 15 min.
7. Remove IA, add 150 μL of 0.05 M NH_4HCO_3 and incubate at room temperature for 15 min.
8. Add 200 μL of acetonitrile and let it sit at room temperature for 15 min.
9. Remove NH_4HCO_3/acetonitrile and let the gel piece to dry for 30 min with air or using a SpeedVac vacuum concentrator for 5 min at room temperature (*see* **Note 29**).
10. Add 10–15 μL of ready-to-use trypsin solution and incubate at 37 °C in a microfuge at 230 rpm for 16–18 h.
11. Store the gel spot at −20 °C before peptides extraction and clean-up steps.

3.9 Peptides Extraction

1. Remove digest solution (0.6 mL) and put it in a new tube.
2. Add 20 μL of Buffer A to gel piece in old tube (*see* **Note 30**).
3. Sonicate in bath for 15 min.
4. Add 30 μL of Buffer B and sonicate for 15 min.
5. Combine total solution with supernatant digest solution in **step 3**.
6. Using SpeedVac, concentrate sample to a volume of ~10 μL (*see* **Note 31**).

3.10 Clean-up, Concentration and Purification of Sample Using ZipTip Pipette Tips (See Note 5)

1. Aspirate 10 μL of Buffer C in ZipTip pippette tip (ZipTip) and empty to waste. Repeat this step five times.
2. Equilibrate the ZipTip with 10 μL of Buffer A by aspirating and emptying to waste. Repeat this step five times.
3. Draw up sample from above and push gently. Repeat this step ten times very slowly (*see* **Note 32**).
4. Wash sample on ZipTip with Buffer A by drawing up 10 μL of Buffer A and aspirating several times. Repeat this process three times.
5. Remove all liquid from inside/outside of ZipTip.
6. Draw up 10 μL of Buffer D with ZipTip and elute into a new 1.5 mL tube. Draw up eluent several times and aspirate in ZipTip to ensure all sample is removed from column.
7. Throw ZipTip away and store sample at −80 °C until mass spectrometry analysis.

At this point the gel spot is ready for mass spectrometry analysis to identify e.g. a specific protein increasingly oxidized in the pathological sample compared to the control group (*see* **Note 33**).

4 Notes

1. According to the objective of the proteome fraction of the study, the user can consider applying different depletion strategies to the sample. Our laboratory employed two different kits: ProteoPrep Blue albumin and IgGs depletion kit to deplete the two most abundant proteins (about 85 % of the total proteome) and obtained a clear and clean image of serum/plasma and amniotic fluid proteomes, with detection of proteins normally hidden by albumin or IgGs abundance [6–9]; MARS 14 depletion column to deplete 14 of the most abundant proteins (albumin, IgGs, IgMs, IgAs, haptoglobin, transferrin, alpha1-antitrypsin, alpha2-macroglobulin, complement C3, alpha1-acid glycoprotein, apolipoprotein AI, transthyretin, apolipoprotein AII, apolipoprotein B100) removed ca. 94 % of the total plasma proteome allowing analysis of the remaining 6 % of low abundant proteins [10]. Other commercially available depletion strategies are: Aurum Affi-Gels Blue mini kit (Bio-Rad), Vivapures anti-HSA/IgG kit (Sartorius Stedim Biotech), Qproteome albumin/IgGdepletion kit (Qiagen), MARC (human 6) and MARS Hu-7 kit (Agilent Technologies), Seppros MIXED12-LC20 column (GenWay Biotech), ProteoPreps 20 plasma immunodepletion kit (Sigma-Aldrich) and Amicon Ultra-4 filters (Millipore) [4, 5, 11].
2. The polyacrylamide percentage (w/v) of the gel should be decided according to the proteome fraction to be analyzed and the type of electrophoresis separation needed. The 12 % (v/w) or Any kD™ polyacrylamide gels (Bio-Rad) allows a good migration of proteins ranging from 6.5 to 200 kDa.
3. The Sypro stain is an ultrasensitive, luminescent stain that allows the linear quantitation range of over three orders of magnitude having a detection limit of 0.25–1 ng of proteins. Its Ex/Em ratio is 280, 450/610 nm and requires a laser-based or UV imaging system.
4. PD-Quest software is one of most commonly used software for proteomics image analysis. Alternative options are: Image master 2D (GE Healthcare), Delta 2D (Decodon), SameSpot (Nonlinear Dynamics), REDFIN (Ludesi) and similar.
5. ZipTip C18 pipette tips (Millipore) contain a bed of chromatography media fixed at its end for concentrating and purifying samples. Alternatively, SupelcoTip C18 Pipette tips (Sigma-Aldrich) or equivalent can be used.
6. Performing depletion strategies on CSF samples is not always recommended due to their low protein concentrations, which may result in final protein levels precluding the application of the redox proteomics technique. Thus, for the CSF sample

preparation we excluded the depletion step and performed a protein precipitation step with acetone to concentrate the protein in a lower working volume.

7. A higher volume of CSF can be mixed with acetone to obtain an increased amount of precipitated protein, always maintaining the 1:3 ratio CSF/acetone.
8. The solubilizing step is crucial to obtain a good 2D migration pattern. Thus, before starting the IEF step, check the sample and, in case of presence of proteins aggregates the solubilizing time can be extended, and an additional sonication step can be performed.
9. In this step, if the sample volume is higher than 50 μL the user should precipitate the protein with 15 % (w/v) trichloroacetic acid (TCA) for 30 min at 4 °C and re-suspend the protein pellet directly in 200 μL of RB.
10. Distribute the sample along the entire length of the lane to obtain a better absorption by the gel strip.
11. Alternatively, a 45 min passive rehydration program can be set on the protean IEF cell apparatus.
12. This step is crucial to obtain reproducible results because mineral oil allows the entire sample to be in contact with the gel strip for its complete absorption.
13. To obtain a slightly better 2D image quality, the strip should be run for a second dimension after the IEF step.
14. Gels with different polyacrylamide percentage need different running time; thus, always check the migration line of the samples.
15. Use nitrocellulose membranes at 0.2 μm pore size (which is better than 0.45 μm for transfer efficiency) and not polyvinylidene fluoride (PVDF) membranes. In our experience, the latter do not allow the post-derivatization step for protein carbonyls analysis to be performed. In addition, PVDF membranes are in general more delicate to handle for image analysis.
16. The transfer method can be semi-dry or wet according with the user background and equipment (*see* Protein Blotting Guide. Bio-Rad bulletin: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_2895.pdf).
17. The post-derivatization step needs to be performed only if the user wants to analyze protein carbonyls as a marker of protein oxidation in biofluids [12]. Otherwise skip this step and proceed to Subheading 3.4, **step 1**.
18. DNPH solution can be used at 0.05 mM instead of 0.1 mM, but with a lower developing efficiency.

19. For these three steps the user needs to be careful about the time. Five minutes time has to be strictly applied. The volume of solutions needs to cover the membrane completely.
20. Choose the primary antibody according to the selected protein oxidative marker for the experimental analyses. Check the datasheet of the antibody of choice for the concentration and timing of incubation.
21. Incubation time is usually 90 min for standard secondary antibodies AP conjugated. However, always check the antibody datasheet for confirmation.
22. According to the antibody used, oxidative marker analyzed, type of sample and amount of sample, the time of stain development ranges from a few minutes to several hours (sometimes overnight incubation is preferred).
23. Check what file format is preferred by the image analysis software employed (*see* **Note 4**), usually .tiff is broadly accepted.
24. User experience in matching the images is crucial to obtain reliable data. Thus, perform several tests before the final matching.
25. The specific protein oxidation value for each matched spot can be calculated as the ratio of the oxidation value (membrane analysis) to the expression value (gel analysis) [12].
26. Use software packages for advanced statistical analyses executing *T*-test and ANOVA statistical analyses.
27. Always wear gloves and, if possible, perform protein digestion steps under a hood to avoid sample contamination. Keratin from skin is a routine contaminant in the absence of a negative air pressure in a hood.
28. User must be careful on cutting only the spot of interest, to avoid false identifications by mass spectrometry analysis. It is always suggested to cut the same spot from different gels for confirmation analysis on protein identification.
29. If available, always use the SpeedVac to dry the gel pieces.
30. The volume used should be more than two times higher than the volume necessary to submerge the gel.
31. Check sample. Avoid drying to completion.
32. After SpeedVac step (*see* Subheading 3.6, **step 6**), if volume is less than 10 μL add buffer A to get a volume to 10 μL.
33. Mass spectrometry analyses of peptides obtained after digestion and clean-up steps are submitted to MS search engines and databases (e.g. SEQUEST, http://fields.scripps.edu/sequest/; Swiss-Prot, http://www.ebi.ac.uk/swissprot/) with specific guidelines and criteria for protein identification [12].

References

1. Butterfield DA, Perluigi M, Reed T et al (2012) Redox proteomics in selected neurodegenerative disorders: from its infancy to future applications. Antioxid Redox Signal 17:1610–1655
2. Shacter E (2000) Quantification and significance of protein oxidation in biological samples. Drug Metab Rev 32:307–326
3. Di Domenico F, Coccia R, Butterfield DA, Perluigi M (2011) Circulating biomarkers of protein oxidation for Alzheimer disease: expectations within limits. Biochim Biophys Acta 1814:1785–1795
4. Bandow JE (2010) Comparison of protein enrichment strategies for proteome analysis of plasma. Proteomics 10:1416–1425
5. Polaskova V, Kapur A, Khan A et al (2010) High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. Electrophoresis 31:471–482
6. Cocciolo A, Di Domenico F, Coccia R et al (2012) Decreased expression and increased oxidation of plasma haptoglobin in Alzheimer disease: insights from redox proteomics. Free Radic Biol Med 53:1868–1876
7. Spadaccio C, Di Domenico F, Perluigi M et al (2012) Serum proteomics in patients with diagnosis of abdominal aortic aneurysm. Cardiovasc Pathol 21:283–290
8. Perluigi M, di Domenico F, Fiorini A et al (2011) Oxidative stress occurs early in Down syndrome pregnancy: a redox proteomics analysis of amniotic fluid. Proteomics Clin Appl 5:167–178
9. Perluigi M, Di Domenico F, Cini C et al (2009) Proteomic analysis for the study of amniotic fluid protein composition. J Prenat Med 3:39–41
10. Fiorini A, Koudriavtseva T, Bucaj E et al (2013) Involvement of oxidative stress in occurrence of relapses in multiple sclerosis: the spectrum of oxidatively modified serum proteins detected by proteomics and redox proteomics analysis. PLoS One 8:e65184
11. Millioni R, Tolin S, Puricelli L et al (2011) High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. PLoS One 6:e19603
12. Di Domenico F, Coccia R, Cocciolo A et al (2013) Impairment of proteostasis network in Down syndrome prior to the development of Alzheimer's disease neuropathology: redox proteomics analysis of human brain. Biochim Biophys Acta 1832:1249–1259

Chapter 24

Advanced Shotgun Lipidomics for Characterization of Altered Lipid Patterns in Neurodegenerative Diseases and Brain Injury

Miao Wang and Xianlin Han

Abstract

Multi-dimensional mass spectrometry-based shotgun lipidomics (MDMS-SL) is a powerful technology platform among current lipidomics practices due to its high efficiency, sensitivity, and reproducibility, as well as its broad coverage. This platform has been widely used to determine the altered lipid profiles induced by diseases, injury, genetic manipulations, drug treatments, and aging, among others. Herein, we summarize the principles underlying this platform and present a protocol for analysis of many of the lipid classes and subclasses covered by MDMS-SL directly from lipid extracts of brain samples. We believe that this protocol can aid researchers in the field to determine altered lipid patterns in neurodegenerative diseases and brain injury.

Key words Alzheimer's disease, Brain injury, Lipidome, Metabolome, Metabolomics, Mass spectrometry, Multi-dimensional mass spectrometry (MS)-based shotgun lipidomics, MDMS-SL, Neurodegeneration, Shotgun lipidomics

1 Introduction

Lipidomics, defined as the large-scale study of cellular lipids, is a rapidly expanding research field [1–3]. Numerous new discoveries and advances have been made in recent years [3–12]. Its essential roles in identifying the biochemical mechanisms of lipid metabolism, investigating the functions of an individual gene of interest, identifying novel biomarkers, and evaluating drug efficacy, among others, are becoming increasingly appreciated. One important task in lipidomics is the high-throughput identification and quantitation of individual lipid molecular species in each cellular lipidome.

One of the major new developments in current lipidomics is the multi-dimensional mass spectrometry (MS)-based shotgun lipidomics (MDMS-SL) method [4, 13, 14]. The principle underlying the direct infusion-based MDMS-SL technology is to maximally exploit the unique chemical and physical properties of lipid classes

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_24,

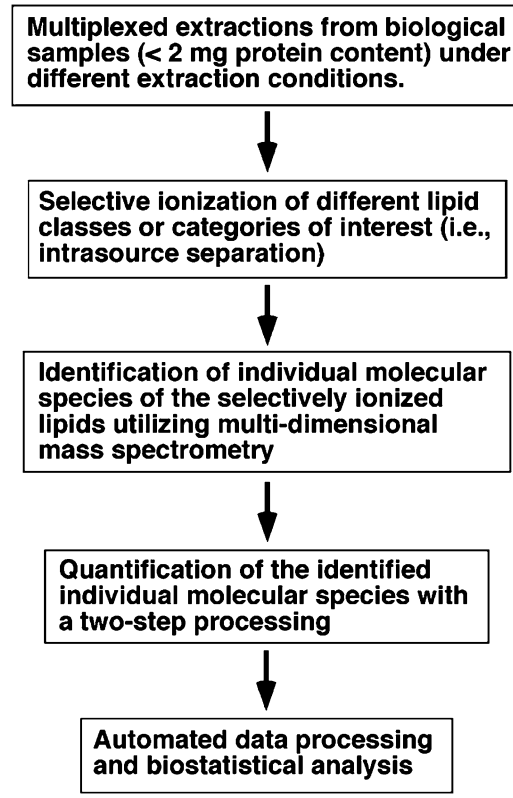


Fig. 1 Workflow of multi-dimensional mass spectrometry-based shotgun lipidomics

in combination with the latest advances in lipid MS analysis, thereby achieving maximal separation and ionization, and minimal ion suppression. This principle and the comparison of its differences with other approaches have been extensively described in our review article [14]. The workflow of this platform is schematically illustrated in Fig. 1.

In brief, the lipids of a biological sample (cells, tissue or biologic fluid) containing less than 2 mg of protein can be extracted by solvent(s) under acidic, basic, and/or neutral conditions (i.e. multiplexed extractions). The distinct solubility of the different lipid classes in various solvents and pHs can be exploited in this critical step to maximize separation and enrich the lipid class(es) of interest. For example, many lipid classes (e.g., sphingosine-1-phosphate, lysophosphatidic acid (LPA), acylcarnitine, etc.) can be efficiently extracted under acidic conditions [4] or recovered from aqueous phase (Han, unpublished data). Gangliosides and acyl-CoA compounds are highly soluble in polar solvents and are partitioned into the aqueous phase during chloroform extraction [15–17]. Thus, these lipid classes can be reverse-extracted using

butanol or other solvents under acidic conditions. Moreover, highly hydrophobic lipid classes (e.g. cholesterol and its esters, triacylglycerol (TAG), non-esterified fatty acids (NEFA) and others) can be extracted and enriched with hexane. Fluorenylmethoxylcarbonyl (Fmoc) chloride can be added to quickly tag the amine-containing lipids and increase the sensitivity for analysis of these lipids through neutral-loss scanning of the tagged Fmoc moiety [18]. Base hydrolysis of all ester-linked glycerolipids can be exploited for isolation and enrichment of sphingoid-backbone containing lipids [19]. In contrast, vinyl ether-linked lipid species (i.e. plasmalogens) are labile under acidic conditions. This chemical instability can be used to unambiguously identify the presence of plasmalogen species with a comparison of the mass spectra acquired before and after acid treatment [20].

At the MS technological level, the electrospray ion source behaves like an electrophoretic cell and can selectively separate different charged moieties under high electrical potential (typically ~4 kV) [21, 22]. Since different lipid classes have different electrical properties, largely depending on the nature of their polar head groups [1, 4], the electrospray ion source can be used to resolve lipid classes in a crude lipid extract based upon the intrinsic electrical properties of each lipid class, by "intrasource separation of lipids" [4, 13, 23]. In shotgun lipidomics, the differential acidic or basic properties of lipid classes in a solution at a specific pH are exploited to selectively ionize different lipid classes in the positive- or negative-ion modes and to achieve a maximal ionization sensitivity [24]. Thus, the lipid classes containing phosphate (e.g., anionic phospholipids, ethanolamine glycerophospholipid (PE), acyl-CoA, and sphingosine-1-phosphate), sulfate (e.g., sulfatide), and carboxylate (e.g., gangliosides and NEFA) can be selectively ionized in the negative-ion mode, for some classes under basic conditions (i.e., in the presence of NH_4OH or LiOH in a concentration of 50 % of PE concentration), whereas lipid classes containing amine (e.g. acylcarnitine) can be readily ionized in the positive-ion mode under acidic conditions [4]. Molecular species of other lipid classes can be ionized as either alkaline or anion (e.g. chloride, acetate or formate) adducts in the positive- or negative-ion mode respectively, as discussed in detail in previous studies [4].

At this point, finding a sensitive and unique fragment specific to a class or a group of lipids of interest, readily detectable after collision induced-dissociation in mass spectrometry is the third key step for successfully identifying, profiling and quantifying individual lipid species in a class or group. Either neutral-loss scanning (NLS) or precursor-ion scanning (PIS) at the mass or m/z of the fragment of interest, can be performed to "isolate" a given class or a group of lipids respectively, from which individual lipid fragments, molecular species, can be identified in a multi-dimensional array analysis [13, 14]. Each of these fragments represents a building

block of the class or group of lipids. All of the building blocks of each lipid class together constitute an "additional dimension" to the molecular ions present in the survey scan. Here the molecular ions constitute the first dimension, whereas the building blocks constitute the second dimension [7, 13]. For example, three moieties linked to the hydroxyl groups of glycerol can be recognized as three individual building blocks, and if each building block is identified, then each individual glycerol-derived lipid molecular species in a given sample can be determined [13].

As the last step, quantitation by shotgun lipidomics is performed in a two-step procedure [13, 25, 26]. First, the abundant and non-overlapping molecular species of a class are quantified by comparing the ion peak intensity of each individual identified molecular species to that of the pre-selected internal standard of the class after ^{13}C de-isotoping [4, 27] from a survey scan. Next, some or all of these determined molecular species of the class (plus the pre-selected internal standard) are used as standards to determine the content of other low-abundance or overlapping molecular species using one or multiple NLS and/or PIS scans which are specific to the building blocks (e.g. headgroup) of the lipid class of interest (*see* above). Multiple standards are necessary in this second step since the fragmentation kinetics of different molecular species may be different [28, 29]. Here, it should be pointed out that such an approach by using tandem MS/MS spectrum along with at least two internal standards for quantitation has been broadly employed in the field [29–32]. Through this second step, the linear dynamic range can be dramatically extended by eliminating background noise, and by filtering the overlapping molecular species through a multi-dimensional mass spectrometric approach [4].

At its current stage, advanced shotgun lipidomics, including lipid class-selective intrasource ionization and subsequent multi-dimensional MS analyses enables us to fingerprint and quantify individual molecular species of most major and many minor lipid classes in cellular lipidomes, which collectively represent >95 % of the total lipid mass (composed of hundreds to thousands of molecular species), directly from their chloroform extracts after multiplexed sample preparation. These classes of lipids include, choline glycerophospholipid (PC), PE, phosphatidylinositol (PI), phosphatidylglycerol (PG), phosphatidylserine (PS), phosphatidic acid (PA), sphingomyelin (SM), monohexosylceramide (i.e., galactosylceramide and/or glucosylceramide, HexCer), sulfatide, NEFA, TAG, lysoPC, lysoPE, lysoPA, acylcarnitine, cholesterol and cholesteryl esters, and ceramide (Cer) (including dihydroceramide). Special methodologies for cardiolipin (CL) [33], 4-hydroxyalkenal [34], sphingosine-1-phosphate [35], sulfatide [36], and sphingosine, psychosine, and lysoSM [19] have also been developed based on their chemical properties.

In this chapter, the protocol for identification and quantitation of several representative lipid classes for research of neurodegenerative diseases and brain injury is described. Although we believe that the MDMS-SL technology platform is powerful for comprehensive analysis of the majority of lipid classes present in cellular lipidomes and the described protocol is readily applicable to other studies, the specific characteristics of lipidomic research in neurodegenerative diseases should be recognized and solutions to address these concerns provided. To our knowledge, at the metabolomics level, up to date only MDMS-SL has thoroughly addressed these concerns and provides a criterion towards direct, truly reliable standardization of sampling procedures. Thus, since the major biological materials in neurodegeneration/brain injury research are brain tissues, one of the concerns is the homogeneity/heterogeneity (varying percentage) of different cell populations in the sampled tissues. For example, neurons are enriched in gray matter whereas oligodendrocytes are mainly present in white matter. Differences in the ratio of co-existing gray to white matter may represent an unpredictable variable which may overshadow real differences between samples from disease states relative to controls. Comprehensive MDMS-SL analysis of human brain samples demonstrates the presence of very distinct lipid profiles of PE molecular species in gray and white matter samples (Fig. 2). More specifically, electrospray ionization mass spectrometry (ESI-MS) analysis of lipid extracts of cortex gray matter from post-mortem

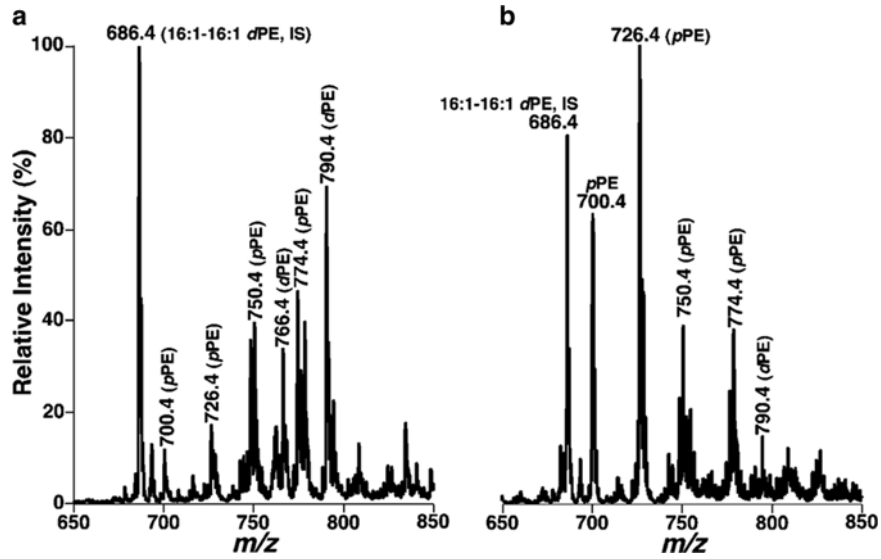


Fig. 2 Distinct profiles of ethanolamine glycerophospholipid molecular species in lipid extracts of cognitively normal human occipital gray (panel **a**) and white matter (panel **b**). Plasmenylethanolamine and phosphatidylethanolamine are abbreviated as "pPE" and "dPE", respectively. "IS" denotes internal standard. Reproduced from ref. [51] with permission from Elsevier B.V., Copyright 2010

subjects has demonstrated multiple predominant deprotonated ion peaks corresponding to PE species (Fig. 2a) in which over 80 % of PE molecular species (mol %), and 55–60 % of plasmalogen PE (pPE) species contain polyunsaturated fatty acyl chains at the *sn*-2 position [37]. In contrast, ESI-MS analysis of lipid extracts of white matter from different brain regions has revealed the presence of one predominant peak at *m/z* 726.4 containing monounsaturated acyl chain (18:1–18:1 pPE) which represents over 85 % (mol %) of the total pPE (Fig. 2b). Accordingly, the distinct PE molecular species profiles between brain gray and white matters provide an important criterion to distinguish gray and white matters but, more importantly, to determine the degree of cross-contamination from co-existing gray and white matters. The degree of the cross-contamination can be accurately determined based upon the peak intensity ratios of ions at *m/z* 726.4 (18:1–18:1 pPE) and 790.4 (18:0–22:6 PE). Standardization of sampling with minimum contamination should constitute a priority, followed by determination and characterization of specific disease state patterns based on entire profiles (Fig. 2).

2 Materials

2.1 Equipment (See Note 1)

1. Nano-ESI source device (TriVersa NanoMate, Advion Bioscience Ltd., Ithaca, NY).
2. Mass spectrometers (Thermo TSQ VANTAGE, San Jose, CA; ABSciex 4800 MALDI TOF/TOF Analyzer, Framingham, MA).

2.2 Reagents and Solutions

1. Commonly used reagents and solutions:
   - Phosphate buffered saline (PBS) (1×): 8 g NaCl, 1.44 g of Na_2HPO_4, 0.24 g KH_2PO_4, 0.2 g KCl, pH, 7.4 in 1 L of ultrapure water.
   - Solvent A (extraction solvent): chloroform/methanol solution (1:1, v/v).
2. Individual lipid internal standards stock solutions (1 mg/mL) in solvent A, including:
   - 1,2-Dimyristoleoyl-*sn*-glycero-3-phosphocholine (di14:1 PC) solution.
   - 1,2-Dipalmitoleoyl-*sn*-glycero-3-phosphoethanolamine (di16:1 PE) solution.
   - 1,2-Dipentadecanoyl-*sn*-glycero-3-phosphoglycerol (sodium salt) (di15:0 PG) solution.
   - 1,2-Dimyristoyl-*sn*-glycero-3-phosphoserine (sodium salt) (di14:0 PS) solution.

- 1,2-Dimyristoyl-*sn*-glycero-3-phosphate (sodium salt) (di14:0 PA) solution;
- 1,1′,2,2′-Tetramyristoyl cardiolipin (T14:0 CL) solution.
- 1-Heptadecanoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (17:0 lysoPC) solution.
- *N*-Lauroryl sphingomyelin (N12:0 SM) solution.
- *N*-Heptadecanoyl ceramide (N17:0 Cer) solution.
- *N*-Lauroyl sulfatide (N12:0 sulfatide) solution.
- (the above purchased from Avanti Polar Lipids, Inc., Alabaster, AL, except noted).
- 7,7,8,8-d_4-Palmitic acid (d_4-16:0 NEFA) solution (Cambridge Isotope Laboratories, Andover, MA).
- Triheptadecenoin (T17:1 TAG) solution (Nu Chek, Inc., Elysian, MN).
- *N*-Pentadecanoyl galactosylceramide (N15:0 GalCer) solution (Matreya, Inc., Pleasant Gap, PA).

3. Internal standards mixture:
   - This includes, di14:1 PC, di16:1 PE, di15:0 PG, di14:0 PS, di14:0 PA, T14:0 CL, d_4-16:0 NEFA, 17:0 lysoPC, T17:1 TAG, N12:0 SM, N17:0 Cer, N12:0 sulfatide, N15:0 GalCer, etc. The amount of each individual lipid species is based on the abundance of the corresponding lipid class in the sample (the molecular species of internal standards are selected because they represent <0.1 % of the endogenous cellular lipid mass levels, predetermined by ESI-MS lipid analysis).
4. 50 mM lithium chloride (LiCl) solution.
5. 10 mM lithium chloride (LiCl) solution.
6. 1 M lithium methoxide (LiOMe) solution in methanol.
7. 0.4 % acetic acid solution.
8. 80 times diluted lithium hydroxide (LiOH) saturated in methanol.
9. (1:2:4, v/v/v) chloroform/methanol/isopropanol solution.

3 Methods

3.1 Cellular Lipid Extraction and Sample Preparation

Overall, tissue samples are homogenized in ten times diluted phosphate buffered saline (PBS). An appropriate amount of internal standard premixture is spiked into the homogenates or other biofluid samples. After a modified Bligh and Dyer extraction procedure using standard methods as previously reported [38], the generated lipid extract is analyzed by ESI-MS and ESI-MS/MS.

1. Keep the brain sample (~25 mg) in a 1.5 mL Eppendorf tube. Add 300 μL of ten times diluted PBS in the tube. The samples are separately homogenized for 1 min by using a disposable soft tissue homogenizer with an up-and-down dabbing motion (*see* **Note 2**). Pipette an aliquot of 25 μL for determination of protein content.
2. Protein assay is carried out by using a 96-well microplate and following the manufacturer's instruction (e.g. BioRad or equivalent) using bovine serum albumin (BSA) as standard.
3. Accurately transfer individual homogenate of the tissue samples in **step 1** to a disposable culture borosilicate glass tube (16 × 100 mm) and record the transferred volume. Add a certain amount of the premixture of internal standards based on the protein content of the transferred homogenate (*see* **Note 3**).
4. For liquid samples (e.g., plasma, serum, or cerebrospinal fluid), for research on neurodegeneration and/or brain injury studies, the individual sample is accurately transferred into a disposable culture borosilicate glass tube (16 × 100 mm) and the transferred volume of the sample is recorded. Add a certain amount of the premixture of internal standards based on the volume of the transferred sample (*see* **Notes 3** and **4**).
5. Prepare extraction solvent (solvent A) and 10 and 50 mM lithium chloride solutions.
6. Add 4 mL of extraction solvent to the glass tube for extraction (*see* **step 3** or **4**), and an appropriate volume of 50 mM LiCl to bring the aqueous phase to a final volume of 2 mL. Cap the tubes and vortex them for 20 s. The samples are then centrifuged at 2,700 × *g* for 10 min.
7. Collect the bottom layer to a new borosilicate glass tube (*see* **Note 5**). Add 2 mL chloroform to individual glass tubes with the residual top layer. Cap the tubes and vortex them for 20 s. The samples are again centrifuged at 2,700 × *g* for 10 min.
8. Collect the bottom layer and combine it with the one collected in **step 7** (*see* **Note 5**). Evaporate the combined bottom layer under a stream of nitrogen with a nitrogen-evaporator until totally dried.
9. Resuspend individual residue in **step 8** with 4 mL of solvent A, and add 2 mL of 10 mM LiCl. Cap the tubes and vortex them for 20 s. The samples are centrifuged at 2,700 × *g* for 10 min. Repeat **steps 7** and **8**.
10. Resuspend the individual lipid extract residue from **step 9** with solvent A in a volume of 200 μL/mg protein or 1 mL/mL original liquid samples. The lipid extracts are flushed with nitrogen, capped, and stored at −20 °C for MS analysis.

11. A quarter of each individual lipid extract (**step 10**) is transferred to a conic centrifuge glass test tube and the solvent is evaporated under the stream of nitrogen. A small volume (50 μL) of ice-cold 1 M lithium methoxide (LiOMe) solution in methanol is added to the test tube at 0 °C. The reaction mixture is vortexed for 15 s, stood in an ice bath for 1 h, and quenched with 2 mL of 0.4 % acetic acid solution. The pH of the quenched reaction solution should be adjusted to (4–5) by addition of acetic acid if necessary. The aqueous phase is washed with hexane (2 mL, three times) and discarded. The lipids in the aqueous phase are extracted by the modified Bligh and Dyer method as described in **step 9**. The combined chloroform phase is dried under a stream of nitrogen. Each individual extract is reconstituted in 100 μL of solvent A, flushed with nitrogen, capped, and stored at −20 °C for the analysis of the sphingolipidome [19].

3.2 Mass Spectrometric Analysis

Lipid classes present in the samples, with or without hydrolysis with lithium methoxide, are analyzed in three different ionization modes: negative-ion ESI, negative-ion ESI plus lithium hydroxide, and positive-ion ESI plus lithium hydroxide. Negative-ion ESI-MS analysis of brain PE species in the presence of lithium hydroxide (LiOH) is first performed to determine the cell populations in an individual sample as previously described [37, 39]. In case the criteria for the purity of cell population are not met [37, 39], lipid extraction and sample preparation (Subheading 3.1) have to be repeated.

1. Dilute each lipid extract solution to a concentration lower than 50 μM of total lipids with chloroform/methanol/isopropanol (1/2/4), with or without 2–5 % LiOH solution in a chemically-resistant 96-well microplate (*see* **Note 6**).
2. Set the ionization voltage of the nanospray ionization source at 1.15 kV in the positive-ion mode, −1.15 kV in the negative-ion mode, and gas pressure at 0.55 psi. Nanospray ionization for each sample is performed by a customized sequence subroutine operated under the Chipsoft software (*see* **Note 7**).
3. For mass spectrometric analysis, collect 2-min duration of signal averaging in the profile mode for each survey MS scan (*see* **Note 8**). For tandem mass spectrometric analysis, set collision gas pressure at 1.0 mTorr, varying the collision energy with the class of lipids, and collect a 5-min period of signal averaging in the profile mode for each tandem MS spectrum, including PIS and NLS scans (which are sensitive and specific to the lipid class or category of lipid of interest as shown in Table 1). All mass spectra are automatically acquired by a customized sequence subroutine operated under Xcalibur software. For analysis of sulfatide species an alternative method

Table 1
Summary of the specific scans in each lipid class used to identify and quantify individual molecular species

Lipid class (ref)	Ion format	Scans for class specific prescreen	Scans for identification of acyl chain and/or regioisomers	Preliminary scans for the second step quantitation
PC [43]	$[M+Li]^+$	NLS189.1, −35 eV	NLS(59.0+FA), −40 eV	NLS183.1, −35 eV for polyunsaturated acyl chain containing species NLS59.0, −24 eV for plasmalogen species NLS189.1, −35 eV for all the other species
lysoPC [43]	$[M+Na]^+$	NLS59.0, −22 eV NLS205.0, −34 eV	PIS104.1, −34 eV PIS147.1, −34 eV	NLS59.0, −22 eV NLS205.0, −34 eV
PE, lysoPE [18]	$[M-H]^-$ $[M-H+Fmoc]^-$ ($[M+C_{15}H_9O_2]^-$)	PIS196.1, 50 eV for $[M-H]^-$ NLS222.2, 30 eV	PIS(FA-H), 30 eV	NLS222.2,30eVfor$[M-H+Fmoc]^-$
PI, lysoPI [23]	$[M-H]^-$	PIS241.1, 45 eV	PIS(FA-H), 47 eV	PIS241.1, 45 eV
PS, lysoPS [23]	$[M-H]^-$	NLS87.1, 24 eV	PIS(FA-H), 30 eV	NLS87.1, 24 eV
PG, PA, lysoPG, lysoPA [23]	$[M-H]^-$	PIS153.1, 35 eV	PIS(FA-H), 30 eV	PIS153.1, 35 eV
CL, mono-lysoCL [33]	$[M-2H]^{2-}$	Full MS at high resolution	PIS(FA-H) at high resolution, 25 eV; NLS(FA-H2O) at high resolution, 22 eV	
TAG [27]	$[M+Li]^+$		NLS(FA), −35 eV	

Sphingomyelin [43]	$[M+Li]^+$	NLS213.2, −50 eV	NLS (neutral fragments from sphingoid backbone)	NLS213.2, −50 eV
Ceramide [44]	$[M-H]^-$	NLS (neutral fragments from sphingoid backbone), (e.g., NLS256.2, 32 eV for *d*18:1 non-hydroxyl species)	NLS (neutral fragments from sphingoid backbone), (e.g., NLS256.2, 32 eV for *d*18:1 non-hydroxyl species)	NLS (neutral fragments from sphingoid backbone), (e.g., NLS256.2, 32 eV for *d*18:1 non-hydroxyl species)
Hexosyl ceramide [45, 46]	$[M+Li]^+$	NLS162.2, −50 eV	NLS (neutral fragments from sphingoid backbone)	NLS162.2, −50 eV
Sulfatide [47]	$[M-H]^-$	PIS 97.1, 65 eV	NLS (neutral fragments from sphingoid backbone)	PIS97.1, 65 eV
Sphingoid base-1-phosphate [35]	$[M-H]^-$	PIS79.1, 24 eV		PIS79.1, 24 eV
Sphingoid base [19]	$[M+H]^+$	NLS48.0, −18 eV		NLS48.0, 18 eV
Psychosine [48]	$[M+H]^+$	NLS180.0, −24 eV		NLS180.0, −24 eV
Cholesterol [49]	[Cholesteryl methoxyacetate+MeOH+Li]$^+$	PIS97.1, −22 eV		PIS97.1, −22 eV
Acyl carnitine [50]	$[M+H]^+$	PIS85.1, −30 eV	PIS85.1, −30 eV for all species; PIS145.1, −30 eV for hydroxyl species	PIS85.1, −30 eV
Acyl-CoA [16]	$[M-H]^-$, $[M-2H]^{2-}$ $[M-3H]^{3-}$	PIS134.0, 30 eV	PIS134.0, 30 eV	PIS134.0, 30 eV
4-Hydroxyalkenal [34]	$[M+carnosine+H]^+$	NLS71.2, −28 eV NLS117.2, −26 eV		NLS71.2, −28 eV NLS117.2, −26 eV

NLS and PIS stand for neutral-loss scan and precursor-ion scan, respectively. FA and (FA-H) denote free fatty acid and fatty acyl carboxylate anion, respectively. The abbreviations of phospholipid classes are given in the main text

is matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI TOF/TOF) mass spectrometry (*see* **Note 9**).

3.3 MS Data Analysis

Data processing of MS spectra including ion peak selection, data transfer, baseline correction, peak intensity comparison and quantification is conducted by a self-programmed Microsoft Excel macros software [40]. The principles of the macros are summarized as follows.

3.3.1 Establishment of the Database of Lipid Classes and Individual Molecular Species

1. The building block concept regarding lipid molecular structures is extensively employed to build the database of the program [40]. On the basis of the differences of these building blocks, the majority of lipid classes present in the mammalian cellular lipidomes are classified into five categories, including glycerophospholipids, glycerolipids, sphingolipids, sterols, and metabolites (*see* **Note 10**).
2. All lipid classes are defined by backbones and building blocks. For example, choline glycerophospholipid (PC) is one class of lipids in the glycerophospholipid category. It has glycerol as its backbone with three building blocks connecting the three hydroxy groups. The phosphocholine head group at the *sn*-3 position specifies the class. The oxygen atom of glycerol at the *sn*-1 position is connected to an aliphatic chain through an ester, ether, or vinyl ether bond, which defines the phosphatidyl-, plasmanyl-, and plasmenyl-subclasses of PC, according to the IUPAC nomenclature (*see* **Note 11**). The oxygen atom at the *sn*-2 position is connected to the other aliphatic chain through an ester bond.
3. The building blocks of fatty acyl chain vary with the number of carbon atoms and the number of double bonds, as well as the location of the double bonds in the aliphatic chains. The variations of the carbon atom number and double bond number in the aliphatic chains composite of the entire lipid class, such as PC, in the database of lipid classes for MDMS-SL analysis (*see* **Note 12**).

3.3.2 Automated Data Processing to Identify and Quantify Individual Lipid Species of a Specific Class

1. Tabular raw data from the mass spectra is transferred by self-programmed software directly from the Xcalibur platform.
2. The baseline levels of the raw data from the mass spectra are determined based on the existence of an accelerated intensity change from noise to signal [41]. The precisely determined baseline level is deduced from the specific raw data.
3. An ion peak list of the molecular species in a lipid class of interest present in the lipid extract is generated by matching the m/z values of the detected ion peaks after baseline correction

in the specific scan (i.e., PIS or NLS, Table 1) with those of the candidate species in the established database of the lipid class of interest. This peak list represents all the detectable species of the specific class, including isomeric species, and provides information about the total number of carbon atoms and the total number of double bonds of the aliphatic chain(s) from the lipid database of the program.

4. Identification of acyl chain moieties is achieved by loading all PIS or NLS data, with specific acyl chain information. The combination of the paired aliphatic chains is determined by the restriction of the total number of carbon atoms and the total number of double bonds present in the acyl chains identified for each individual species.
5. Before direct quantification of the class of lipid molecular species of interest, two ^{13}C isotope effects need to be considered [4, 27]. The first type comes from the carbon number difference between a given molecular species and the selected internal standard. The second effect can occur because of the overlapping of the ion peak of the species of interest ($m/z = M$) with the ^{13}C isotope peak of other species containing an additional double bond ($m/z = M-2$) (*see* **Note 13**).
6. Quantification of the identified individual molecular species is performed in a two-step procedure [40]. First, an algorithm determines whether overlapping or low-abundance peaks in the peak list of interest exists. The first quantification step is performed for the abundant non-overlapping peaks, by direct ratiometric comparison to the ion peak intensity of the selected internal standard of the respective class in the survey MS scan, after baseline correction and removal of ^{13}C isotope effects [40].
7. The non-overlapping abundant species plus the exogenously added internal standard are the candidate standards to be used for the second quantification step. In this step, the corrected ion peak intensities of the overlapping and/or low-abundance species from the class-specific PIS or NLS (Table 1) are used, to be quantified by ratiometric comparison to the ion-peak intensities of the candidate standards (*see* **Note 14**).

4 Notes

1. The nanospray source is controlled by Chipsoft 8.3.1 software. All MS or tandem MS analyses are operated under the Xcalibur software. Additional equipment and supplies needed include analytical balance (readability 0.01 mg), multi-sample biopulverizer (12 wells, capacity 10–100 mg per well), cryogenic vials (2.0 mL), Branson digital sonifier 450, vortex shaker and

mixer, razor blade or scissors, tissue tearor, 1.5-mL Eppendorf tubes, 1.5 mL polypropylene pestles (disposable soft tissue homogenizer) with handheld pellet pestle motor, disposable culture borosilicate glass tubes (16×100 mm), 5.75″ disposable borosilicate glass Pasteur pipettes, Drummond pipette-aids, table top centrifuge, analytical nitrogen evaporator, 96-well microplates (transparent for protein assay and chemical resistance for preparing lipid samples for direct infusion).

2. During sonification or homogenization, samples are kept in an ice bath to keep them cold.
3. The internal standards mixture includes, di14:1 PC, di16:1 PE, di15:0 PG, di14:0 PS, di14:0 PA, T14:0 CL, d_4-16:0 NEFA, 17:0 lysoPC, T17:1 TAG, N12:0 SM, N17:0 Cer, N12:0 sulfatide, N15:0 GalCer, etc. The stock solution of individual internal standard is prepared in either solvent A or pure chloroform with a concentration approximately 1 mg/mL. The amount of each individual lipid species in the premixture is prepared based on the abundance of the corresponding lipid class in the samples. The molecular species of internal standards are selected because they represent <0.1 % of the endogenous cellular lipid mass levels as predetermined by ESI-MS lipid analysis.
4. Alternatively, liquid samples can also be normalized to their protein contents. In this case, an aliquot of the liquid sample is used to determine the protein content prior to addition of the internal standard premixture.
5. In order to avoid contamination from the top layer (aqueous phase) to the bottom layer, the glass Pasteur pipette can be inserted slowly with minimum air pressure (to prevent top layer liquid entering the pipette), until the pipette reaches the bottom layer. After carefully collecting the bottom layer and taking the pipette out, remove the aqueous contaminant outside of the tip by gently swirling the tip on the edge of the glass tube and quickly transfer the bottom layer to a clean glass tube.
6. The total lipid concentration of a lipid extract can be estimated on the basis of the protein content, or in the range of concentrations from samples obtained in previous studies [38]. This knowledge is useful for estimation of the concentrations of total lipids to prevent lipid aggregation during analysis. The lithium hydroxide is made of 200-time dilution of a saturated methanol solution.
7. Since sample ionization and spectra collection are operated with two separate software programs (i.e., ChipSoft and Xcalibur, respectively), the ionization polarity and time controlled by the ChipSoft should be matched to those of the mass spectrometer. The mass spectrometer will be triggered to start collecting spectra with the start of the nanospray.

8. For the triple-quadrupole mass spectrometer, the first and third quadrupoles are used as independent mass analyzer with a mass resolution of 0.7 Th, and the second quadrupole serves as a collision cell for tandem mass spectrometry. For the analysis of cardiolipin, the mass resolution is set at 0.3 Th to detect its doubly-charged ions [33].
9. MALDI-MS using 9-aminoacridine as matrix can be used for selective desorption/ionization of sulfatide species over other examined anionic lipids present in lipid extracts of biological samples [36]. The structure of individual sulfatide species can be elucidated through product ion analysis by MALDI TOF/TOF-MS.
10. Databases of lipid classes and individual molecular species. MDMS-SL allows to analyze lipid molecular species of a class of interest in a non-targeted approach. Therefore, the database should be as broad and flexible as possible. The initial database covered all possible natural lipid molecular species, whereas the flexibility allows to modify or expand it with new lipid species as necessary [40] (*see* **Note 12**).
11. To date, the plasmanyl and plasmenyl subclasses have been identified only in choline, ethanolamine, and serine glycerophospholipids in mammalian lipidomes [42].
12. The molecular species that are included in our database are approximately 6,500 glycerophospholipid species, 3,200 glycerolipid species, 26,000 sphingolipid species, 100 sterol lipids, and 410 metabolites [40]. Therefore, a total of over 36,000 molecular species, not counting regioisomers, oxidized lipids or other covalently modified entities, are included in the initial construction of the database. Moreover, by modifying the general chemical formulas, the databases can easily be extended to cover any new species or subclasses in each lipid class detectable in the future, e.g. with new mass spectrometers with improved higher sensitivity, or when new, unusual lipid profiles are detected in biological samples.
13. The isotope effects from other atoms, such as hydrogen, nitrogen or phosphorus, are usually neglected due to extremely low abundance of the isotope or no significant difference between the species and the selected internal standard.
14. An algorithm based on two variables (i.e., the differences in the number of total carbon atoms and the number of total double bonds present in fatty acyl chains of each individual species from those of the selected standards) with multivariate least-square regression to determine the correction factors for each individual molecular species for the second-step quantification was generated and applied [40]. With this second step, the linear dynamic range of quantification is extended dramatically

to quantify overlapping and/or low-abundance species with one or more MS/MS scans to reduce background noise, increase signal/noise ratios of low-abundance species, and filter the overlapping molecules with class-specific PIS or NLS.

Acknowledgements

This work was partially supported by National Institute of General Medical Sciences Grant R01 GM105724 and intramural institutional research funds. Special thanks are expressed to Ms. Imee Tiu for her editorial assistance.

References

1. Han X, Gross RW (2003) Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. J Lipid Res 44:1071–1079
2. Lagarde M, Geloen A, Record M et al (2003) Lipidomics is emerging. Biochim Biophys Acta 1634:61
3. Wenk MR (2005) The emerging field of lipidomics. Nat Rev Drug Discov 4:594–610
4. Han X, Gross RW (2005) Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of the cellular lipidomes directly from crude extracts of biological samples. Mass Spectrom Rev 24:367–412
5. Walker JM, Krey JF, Chen JS et al (2005) Targeted lipidomics: fatty acid amides and pain modulation. Prostaglandins Other Lipid Mediat 77:35–45
6. Serhan CN (2005) Mediator lipidomics. Prostaglandins Other Lipid Mediat 77:4–14
7. Han X (2007) Neurolipidomics: challenges and developments. Front Biosci 12:2601–2615
8. Ivanova PT, Milne SB, Forrester JS, Brown HA (2004) LIPID arrays: new tools in the understanding of membrane dynamics and lipid signaling. Mol Interv 4:86–96
9. Welti R, Shah J, Li W et al (2007) Plant lipidomics: discerning biological function by profiling plant complex lipids using mass spectrometry. Front Biosci 12:2494–2506
10. Schiller J, Suss R, Fuchs B et al (2007) MALDI-TOF MS in lipidomics. Front Biosci 12: 2568–2579
11. Albert CJ, Anbukumar DS, Monda JK et al (2007) Myocardial lipidomics. Developments in myocardial nuclear lipidomics. Front Biosci 12:2750–2760
12. Ejsing CS, Duchoslav E, Sampaio J et al (2006) Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. Anal Chem 78:6202–6214
13. Han X, Gross RW (2005) Shotgun lipidomics: multi-dimensional mass spectrometric analysis of cellular lipidomes. Expert Rev Proteomics 2:253–264
14. Han X, Yang K, Gross RW (2012) Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. Mass Spectrom Rev 31:134–178
15. Tsui ZC, Chen QR, Thomas MJ et al (2005) A method for profiling gangliosides in animal tissues using electrospray ionization-tandem mass spectrometry. Anal Biochem 341:251–258
16. Kalderon B, Sheena V, Shachrur S et al (2002) Modulation by nutrients and drugs of liver acyl-CoAs analyzed by mass spectrometry. J Lipid Res 43:1125–1132
17. Golovko MY, Murphy EJ (2004) An improved method for tissue long-chain acyl-CoA extraction and analysis. J Lipid Res 45:1777–1782
18. Han X, Yang K, Cheng H et al (2005) Shotgun lipidomics of phosphoethanolamine-containing lipids in biological samples after one-step in situ derivatization. J Lipid Res 46:1548–1560
19. Jiang X, Cheng H, Yang K et al (2007) Alkaline methanolysis of lipid extracts extends shotgun lipidomics analyses to the low abundance regime of cellular sphingolipids. Anal Biochem 371:135–145
20. Yang K, Zhao Z, Gross RW, Han X (2007) Shotgun lipidomics identifies a paired rule for the presence of isomeric ether phospholipid molecular species. PLoS One 2:e1368
21. Ikonomou MG, Blades AT, Kebarle P (1991) Electrospray-ion spray: a comparison of mechanisms and performance. Anal Chem 63: 1989–1998

22. Gaskell SJ (1997) Electrospray: principles and practice. J Mass Spectrom 32:677–688
23. Han X, Yang J, Cheng H et al (2004) Towards fingerprinting cellular lipidomes directly from biological samples by two-dimensional electrospray ionization mass spectrometry. Anal Biochem 330:317–331
24. Han X, Yang K, Yang J et al (2006) Factors influencing the electrospray intrasource separation and selective ionization of glycerophospholipids. J Am Soc Mass Spectrom 17:264–274
25. Han X, Cheng H, Mancuso DJ, Gross RW (2004) Caloric restriction results in phospholipid depletion, membrane remodeling and triacylglycerol accumulation in murine myocardium. Biochemistry 43:15584–15594
26. Schwudke D, Oegema J, Burton L et al (2006) Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. Anal Chem 78:585–595
27. Han X, Gross RW (2001) Quantitative analysis and molecular species fingerprinting of triacylglyceride molecular species directly from lipid extracts of biological samples by electrospray ionization tandem mass spectrometry. Anal Biochem 295:88–100
28. Han X, Gross RW (1995) Structural determination of picomole amounts of phospholipids via electrospray ionization tandem mass spectrometry. J Am Soc Mass Spectrom 6:1202–1210
29. Brugger B, Erben G, Sandhoff R et al (1997) Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. Proc Natl Acad Sci U S A 94: 2339–2344
30. Lehmann WD, Koester M, Erben G, Keppler D (1997) Characterization and quantification of rat bile phosphatidylcholine by electrospray-tandem mass spectrometry. Anal Biochem 246:102–110
31. Welti R, Li W, Li M et al (2002) Profiling membrane lipids in plant stress responses. Role of phospholipase Da in freezing-induced lipid changes in Arabidopsis. J Biol Chem 277: 31994–32002
32. Welti R, Wang X (2004) Lipid species profiling: a high-throughput approach to identify lipid compositional changes and determine the function of genes involved in lipid metabolism and signaling. Curr Opin Plant Biol 7:337–344
33. Han X, Yang K, Yang J et al (2006) Shotgun lipidomics of cardiolipin molecular species in lipid extracts of biological samples. J Lipid Res 47:864–879
34. Wang M, Fang H, Han X (2012) Shotgun lipidomics analysis of 4-hydroxyalkenal species directly from lipid extracts after one-step in situ derivatization. Anal Chem 84:4580–4586
35. Jiang X, Han X (2006) Characterization and direct quantitation of sphingoid base-1-phosphates from lipid extracts: a shotgun lipidomics approach. J Lipid Res 47:1865–1873
36. Cheng H, Sun G, Yang K et al (2010) Selective desorption/ionization of sulfatides by MALDI-MS facilitated using 9-aminoacridine as matrix. J Lipid Res 51:1599–1609
37. Han X, Holtzman DM, McKeel DW Jr (2001) Plasmalogen deficiency in early Alzheimer's disease subjects and in animal models: molecular characterization using electrospray ionization mass spectrometry. J Neurochem 77: 1168–1180
38. Christie WW, Han X (2010) Lipid analysis: isolation, separation, identification and lipidomic analysis, 4th edn. The Oily Press, Bridgwater, UK
39. Cheng H, Wang M, Li J-L et al (2013) Specific changes of sulfatide levels in individuals with pre-clinical Alzheimer's disease: an early event in disease pathogenesis. J Neurochem 127: 733–738
40. Yang K, Cheng H, Gross RW, Han X (2009) Automated lipid identification and quantification by multi-dimensional mass spectrometry-based shotgun lipidomics. Anal Chem 81: 4356–4368
41. Yang K, Fang X, Gross RW, Han X (2011) A practical approach for determination of mass spectral baselines. J Am Soc Mass Spectrom 22:2090–2099
42. Vance DE, Vance JE (2008) Biochemistry of lipids, lipoproteins and membranes, 5th edn. Elsevier Science B.V., Amsterdam, The Netherlands
43. Yang K, Zhao Z, Gross RW, Han X (2009) Systematic analysis of choline-containing phospholipids using multi-dimensional mass spectrometry-based shotgun lipidomics. J Chromatogr B 877:2924–2936
44. Han X (2002) Characterization and direct quantitation of ceramide molecular species from lipid extracts of biological samples by electrospray ionization tandem mass spectrometry. Anal Biochem 302:199–212
45. Han X, Cheng H (2005) Characterization and direct quantitation of cerebroside molecular species from lipid extracts by shotgun lipidomics. J Lipid Res 46:163–175
46. Hsu FF, Turk J (2001) Structural determination of glycosphingolipids as lithiated adducts by electrospray ionization mass spectrometry using low-energy collisional-activated dissociation on a triple stage quadrupole instrument. J Am Soc Mass Spectrom 12:61–79

47. Hsu F-F, Bohrer A, Turk J (1998) Electrospray ionization tandem mass spectrometric analysis of sulfatide. Determination of fragmentation patterns and characterization of molecular species expressed in brain and in pancreatic islets. Biochim Biophys Acta 1392: 202–216
48. Jiang X, Yang K, Han X (2009) Direct quantitation of psychosine from alkaline-treated lipid extracts with a semi-synthetic internal standard. J Lipid Res 50:162–172
49. Cheng H, Jiang X, Han X (2007) Alterations in lipid homeostasis of mouse dorsal root ganglia induced by apolipoprotein E deficiency: a shotgun lipidomics study. J Neurochem 101: 57–76
50. Su X, Han X, Mancuso DJ et al (2005) Accumulation of long-chain acylcarnitine and 3-hydroxy acylcarnitine molecular species in diabetic myocardium: identification of alterations in mitochondrial fatty acid processing in diabetic myocardium by shotgun lipidomics. Biochemistry 44:5234–5245
51. Han X (2010) Multi-dimensional mass spectrometry-based shotgun lipidomics and the altered lipids at the mild cognitive impairment stage of Alzheimer's disease. Biochim Biophys Acta 1801:774–783

Chapter 25

AlzPathway, an Updated Map of Curated Signaling Pathways: Towards Deciphering Alzheimer's Disease Pathogenesis

Soichi Ogishima, Satoshi Mizuno, Masataka Kikuchi, Akinori Miyashita, Ryozo Kuwano, Hiroshi Tanaka, and Jun Nakaya

Abstract

Alzheimer's disease (AD) is a complex neurodegenerative disorder in which loss of neurons and synaptic function causes dementia in the elderly. To clarify AD pathogenesis and develop drugs for AD, thousands of studies have elucidated signaling pathways involved. However, knowledge of AD signaling pathways has not been compiled as a pathway map. In this chapter, we introduce the manual construction of a pathway map in AD which we call "AlzPathway", that comprehensively catalogs signaling pathways in the field of AD. We have collected and manually curated over 100 review articles related to AD, and have built the AD pathway map. AlzPathway is currently composed of thousands of molecules and reactions in neurons, brain blood barrier, presynaptic, postsynaptic, astrocyte, and microglial cells, with their cellular localizations. AlzPathway provides a systems-biology platform of comprehensive AD signaling and related pathways which is expected to contribute to clarification of AD pathogenesis and AD drug development.

Key words Alzheimer's disease, Systems biology, Signaling pathway, Pathway map, Manual curation, AlzPathway, Drug discovery

1 Introduction

Alzheimer's disease (AD) is a complex neurodegenerative disorder neuropathologically characterized by extracellular plaques of amyloid-beta (Aβ) peptide and intra-neuronal accumulation of neurofibrillary tangles (NFTs) [1]. AD causes dementia of the Alzheimer type in the elderly, with the number of patients increasing rapidly, which is becoming a serious social issue in the aging society. To address this issue, clarification of the pathogenic mechanisms of AD and development of AD drugs are urgently needed.

Genetic association studies with identification of putative AD susceptibility genes have been performed, and information collected in a publicly available database (AlzGene; http://www.alzgene.org/) [2].

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_25,

Efforts to clarify pathogenic signaling proteins and their signaling pathways in AD are also subject of continuous investigation. These are essential to understand two core pathological hallmarks of AD, amyloid plaques and neurofibrillary tangles (NFT) accumulation, with their underlying origin and exact role yet to be revealed. Several AD pathways associated with these two hallmarks have been studied in separate articles. However, they have not been properly compiled yet.

In this chapter, we introduce a manual construction of a pathway map in AD called "AlzPathway" that comprehensively catalogs signaling pathways in the field of AD [3]. We have collected and manually curated over 100 review articles related to AD, and manually elaborated an AD pathway map. AlzPathway is currently composed of thousands of molecules and reactions in neuron, brain blood barrier, presynaptic, postsynaptic, astrocyte, and microglial cells, with their cellular localizations.

Next generation high-throughput technologies (e.g. next generation sequencing (NGS), RNA-Sequencing (RNA-Seq), proteomics, metabolomics and others) are advancing rapidly, producing massive data contributing to identify e.g. pathogenic gene mutations, aberrant mRNA expression profiles, and aberrant protein interactions. AlzPathway allows not only to evaluate candidate risk genes listed by whole genome sequencing (WGS), but also to analyze 'omics data including e.g. RNA-Seq expression data to reveal patterns and pathways involved in pathogenesis of AD. AlzPathway provides a systems-biology platform of comprehensive AD signaling, which is expected to contribute to clarification of AD pathogenesis and AD drug development.

2 Materials

Over 100 carefully selected review articles involved in AD searched by PubMed were collected, to be manually curated, and pathogenic signaling proteins and their signaling pathways compiled as an AD pathway map (*see* **Note 1**).

3 Methods

According to our guideline of manual construction of AlzPathway (*see* **Note 1**), first, we collected selected AD review articles and conducted manual curation.

3.1 Collection of Review Articles and Manual Curation

We manually curated collected 123 review articles, and compiled pathogenic signaling proteins and their signaling pathways as an AD pathway map by using CellDesigner (http://www.celldesigner.org/) [4] (Fig. 1) (*see* **Note 2**). Molecules are distinguished, including

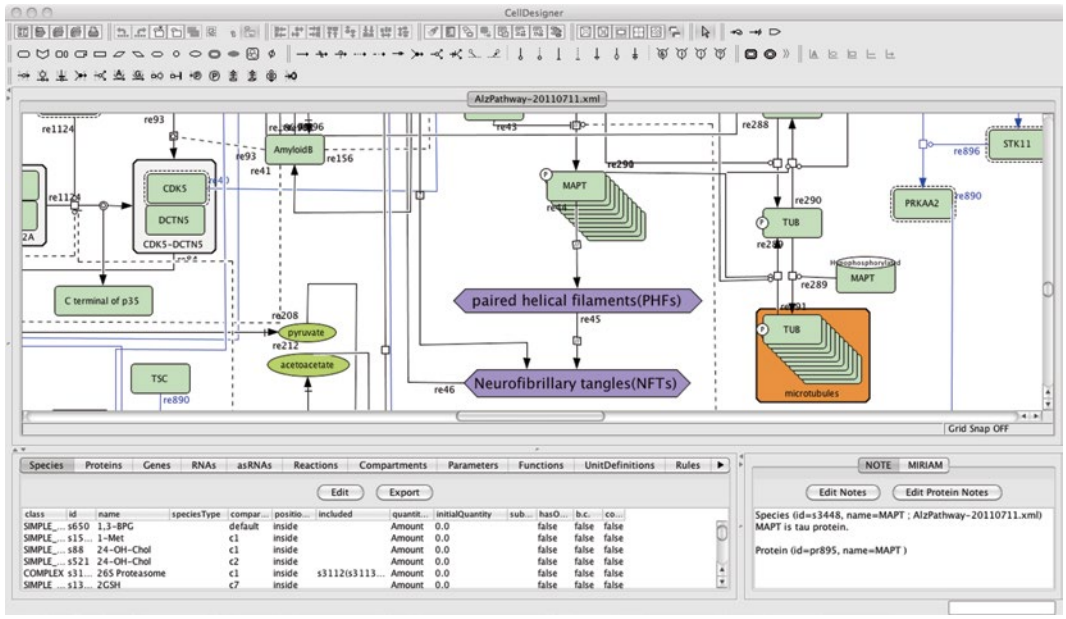


Fig. 1 Manual curation of collected review articles and compilation of AD pathogenic signaling proteins and signaling pathways using CellDesigner. 123 selected AD review articles were collected, manually curated and compiled as pathogenic signaling proteins and their signaling pathways. Visualization using CellDesigner

the following types: proteins, complexes, simple molecules, genes, RNAs, ions, degraded products and phenotypes. Reactions include the following categories: state transition, transcription, translation, heterodimer association, dissociation, transport, unknown transition, and omitted transition. Evidences/links to articles should be described as PubMed IDs using the MIRIAM scheme [5] for all reactions. Cellular types should be distinguished including the following: neuron, astrocyte, and microglial cells. Cellular compartments should include: brain blood barrier, presynaptic, postsynaptic, and cellular localizations. From here, we created a pathway model for AD by adding the molecules on a canvas, creating the reactions, with distinguishing molecule types, reaction categories, cellular types, and cellular compartments. We also added the notes and MIRIAMs to the molecules and the reactions, and tidying up the layout.

3.2 Update of AlzPathway

We have been incorporating new data and updating AlzPathway since the first release of the database [3]. As an example, in 2013 an association between the rs75932628 single nucleotide polymorphism (SNP) in the TREM2 gene and Alzheimer's disease was reported in persons of European ancestry [6, 7], a strong association comparable to those found for apolipoprotein E (APOE) gene variants. According to this new finding, three new species and six new reactions of the TREM2 gene and its related signaling molecules and relations were added to AlzPathway.

3.3 Web Service of AlzPathway

AlzPathway is available as a Systems Biology Markup Language (SBML) map for CellDesigner, i.e. compliant with the SBML language [8] for file exchange between different applications, and as a high resolution image map at http://alzpathway.org/.

AlzPathway is also available as a web service, online map implemented using Payao [9] (Fig. 2). Payao is a community-based, collaborative web service to enable a community to work on the same gene-regulatory and biochemical pathway model simultaneously, insert tags to the model, exchange comments, record discussions and update models. Payao will allow AD researchers not only to browse reactions and their references in PubMed ID but also to comment, correct and update AlzPathway in a community-wide collaboration.

3.4 Overview of AlzPathway

An overview of AlzPathway is shown in Fig. 3. The AlzPathway map consists of 1,347 species, 1,070 reactions, and 129 phenotypes. The molecules are classified as follows: 650 proteins, 232 complexes, 223 simple molecules, 32 genes, 36 RNAs, 24 ions, and 21 degraded products. The reactions are classified as 401 state transitions, 22 transcriptions, 30 translations, 172 heterodimer associations, 49 dissociations, 87 transports, 20 unknown transitions, and 228 omitted transitions. The map consists of the AD hallmark pathways and canonical pathways. The AD hallmark pathways are amyloid β cleavage, amyloid β degradation, APOE-cholesterol pathway and NFT accumulation, which are major pathological

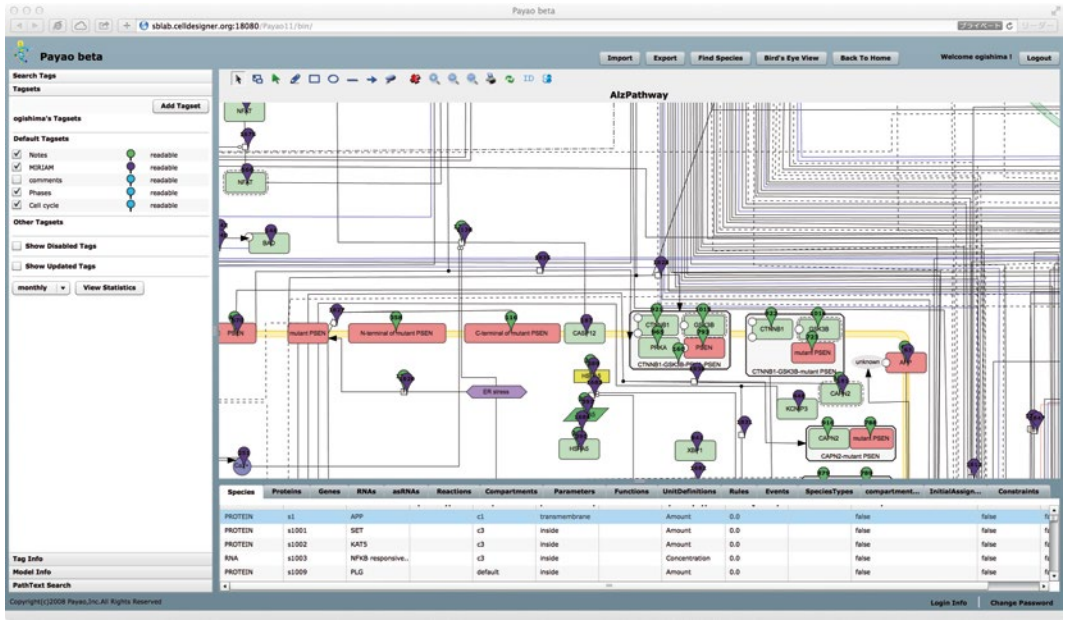


Fig. 2 Community curation of AlzPathway using Payao. Payao system provides a community-based, collaborative web service (online map) platform for pathway manual curation

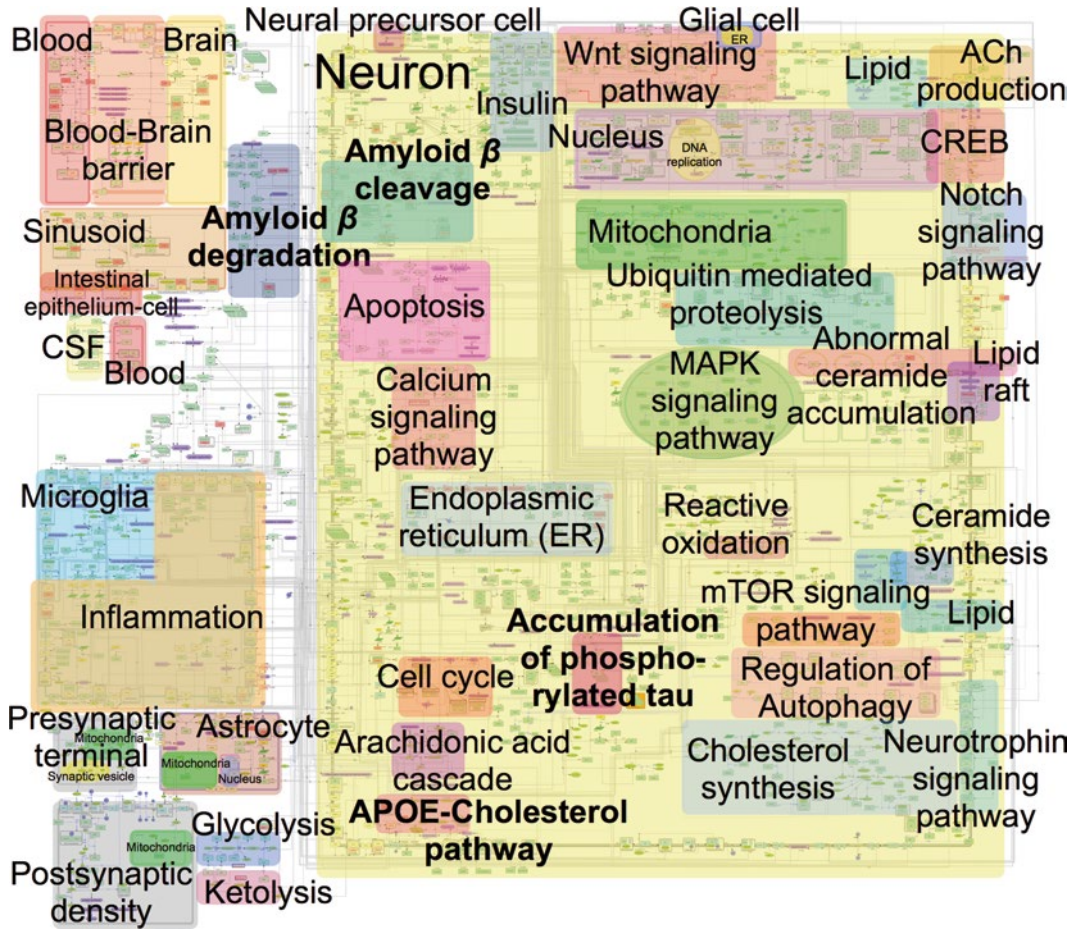


Fig. 3 Overview of AlzPathway overlaid with canonical pathway annotations AlzPathway consists of 1,347 molecules, 1,070 reactions, and 129 phenotypes. AlzPathway is available at the http://alzpathway.org website

pathways of AD. The canonical pathways are acetylcholine production, cholesterol synthesis, Wnt signaling pathway, Notch signaling pathway, Ubiquitin mediated proteolysis, apoptosis, calcium signaling pathway, ER stress, MAPK signaling pathway, abnormal ceramide accumulation, ceramide synthesis, reactive oxidation process, regulation of autophagy, neurotrophin signaling pathway, cell cycle, arachidonic acid cascade, mTOR signaling pathway, lipid pathway, lipid raft, inflammation pathway, insulin pathway, and CREB pathway. Manual elaborations of comprehensive maps have been made before, e.g. for epidermal growth factor receptor (EGF) signaling, toll-like receptor signaling network, RB/E2F signaling and mTOR signaling pathways [10–13], which are individual signaling pathways, but not compilations of pathways/maps involved in a particular disease. Therefore, AlzPathway is the first comprehensive map of a particular disease manually constructed,

which catalogs not only intra- but also inter- and extracellular signaling pathways among neurons, glial cells, microglia, presynaptic cells, postsynaptic cells, astrocytes, and the blood–brain barrier. The brain and spinal cord are made up of various regions and cells, including neurons and glial cells. To clarify pathogenic mechanisms of AD, complex signaling pathways among neurons, glial cells, microglia, presynaptic and postsynaptic cells, astrocytes, and the blood–brain barrier should be elucidated.

3.5 AlzPathway Applications

AlzPathway provides most relevant reported pathogenic signaling proteins and their complex relations (*see* above). Using AlzPathway, we could explore e.g. relationships between two core pathological hallmarks of AD, amyloid plaques and neurofibrillary tangles (NFT) accumulation, and find key molecules in complex signaling pathways. To explore key molecules, network analysis is efficient. We converted the AlzPathway in SBGN PD notation (Systems Biology Graphical Notation-Process Description) notation [14], to a binary-relation notation as a simple interaction format (SIF) file, which can be opened using Cytoscape [15] (the SBGN PD notation [14] is a precise notation for describing pathways but is unsuitable for network analyses). We then calculated edge betweenness centrality, which is defined as the number of the shortest paths that go through an edge in a graph or network V [16]. The edge betweenness centrality is formulated as follows:

3.5.1 Key Molecules Discovery

$$C_b(e) = \sum_{u,w \in V, u \neq w} \frac{\sigma_{uw}(e)}{\sigma_{uw}}$$

where $\sigma_{uw}(e)$ denotes the total number of shortest path between u and w that pass through edge e, and σ_{uw} denotes the total number of shortest paths between u and w. According to their centralities, high centrality relations were obtained and are highlighted in Fig. 4. Highlighted binary relations found were e.g. amyloid plagues formation (amyloid β accumulation) and NFT accumulation (hyperphosphorylated tau accumulation), two AD hallmark pathways. The γ-secretase mediates e.g. amyloid β 1–40 production, which aggregates to form oligomeric amyloid β (amyloid β accumulation) crucial for AD progression. Together with this, microtubule-associated protein tau (encoded by the MAPT gene) is phosphorylated by mutant presenilins (PSEN) and APC-AXIN-GSK3B-CTNNB1, which leads to tau hyperphosphorylation and aggregates to result in NFTs accumulation.

3.5.2 Pathway-Based Drug Discovery

As described above, high centrality relations can be highlighted, and their molecular components, e.g. amyloid β, γ-secretase, APP, APOE, and MAPT, are considered key molecules in AD pathogenesis. A key molecule might be a drug target. A drug targeting a key molecule can affect (e.g. inhibit) its functioning, which might be

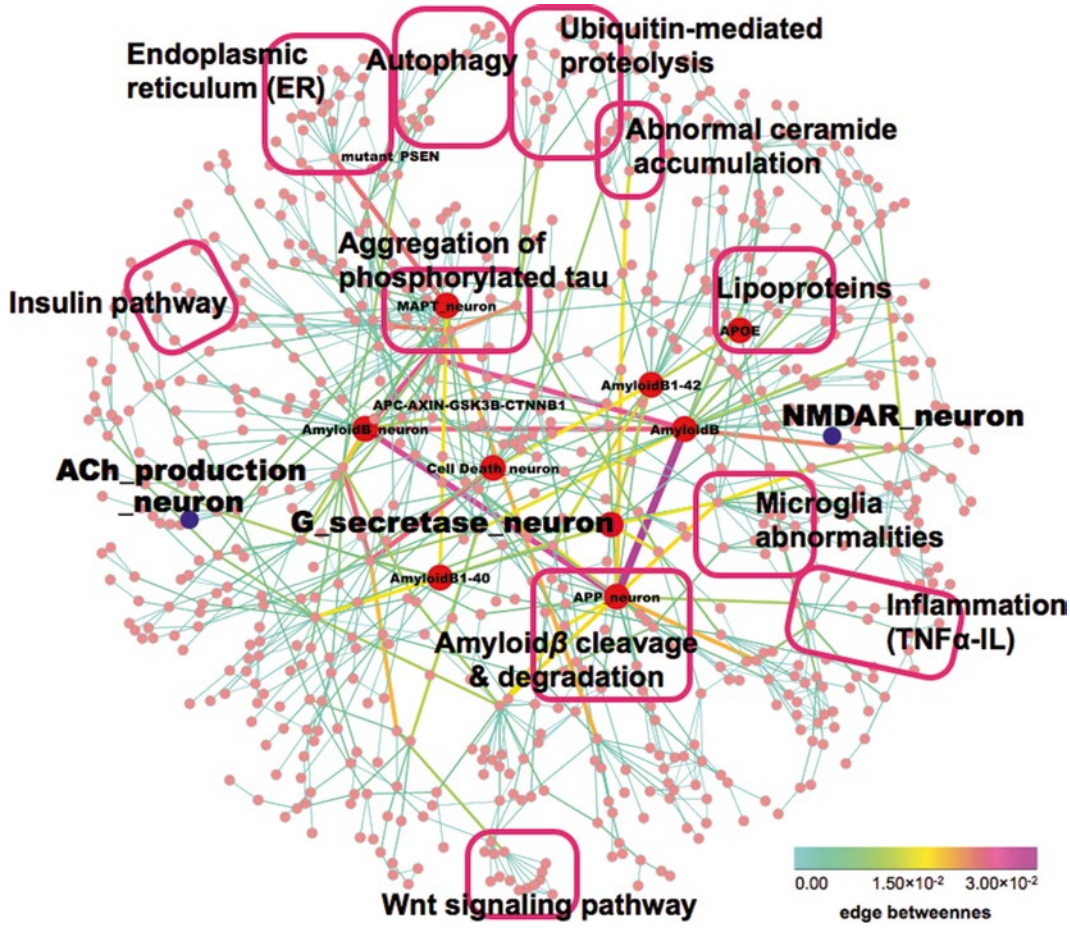


Fig. 4 High centrality relations of AlzPathway and key molecules and pathways. Overview of AlzPathway in binary-relation notation. High centrality relations and key molecules and pathways

investigated as a candidate drug for AD. At the same time, a drug WWtargeting a key molecule might develop significant side effects by affecting off-targets. Could a drug targeting a key molecule be a curative drug for AD?

Currently approved drugs by the U.S. Food and Drug Administration (FDA) are only able to treat, palliate AD symptoms. These are for example, tacrine, rivastigmine, galantamine, donepezil, and memanatine. Tacrine, rivastigmine, galantamine and donepezil are cholinesterase inhibitors, and memanatine is an *N*-methyl-D-aspartic acid (NMDA) receptor antagonist. Interestingly, according to the AlzPathway map, cholinesterase and NMDA receptor are peripheral, not key molecules (Fig. 4), no major interacting pathway appears involved and no interacting compensatory pathways are observed either (Fig. 4). This may implies that these drugs might not cause significant side effects due to off-target effects (focusing on the

AD map only). For these drugs, it can be predicted to have relatively specific effects, due to the low centrality of their target molecules in the AD signaling network. These drugs are dementia-suppressing, palliative drugs, not AD curative drugs. Drugs targeting key molecules with minimum side effects might be investigated as candidate curative drugs.

Semagacestat, a γ-secretase inhibitor targeting a key molecule, γ-secretase, was expected to be a promising drug for AD. Semagacestat had been developed for AD treatment, and was in Phase III. However, it was found to increase the risk of skin cancer, with significant side effects due to its off-targets. Semagacestat targets not only γ-secretase but also peripheral Notch signaling pathway, which heightens a risk of skin cancer compared to placebo. In fact, in the AlzPathway, γ-secretase is as a key molecule showing high centrality, which, if inhibited, could affect unintended downstream molecules and pathways (Fig. 4). Also γ-secretase has a clear relationship with Notch signaling. Thus, AlzPathway collects and provides comprehensive knowledge of AD pathways, and is able to show the possibility of significant side effects according to the structure/topology and relations of the AD signaling pathways.

3.5.3 Cross-Pathway Analysis Between Neurodegenerative Diseases

The rs75932628 SNP in the TREM2 gene was reported to be very strongly associated with Alzheimer's disease, and association comparable to those found for apolipoprotein E (APOE) gene variants. Therefore, recently the TREM2 and its related signaling molecules and reactions were added to the AlzPathway as the latest key molecules and reactions. Interestingly, the TREM2 gene is also reported to be associated with another neurodegenerative disease, Parkinson's disease (PD), and a PD map is available as a PD pathway map. (http://minerva.uni.lu/pd_map) [17]. We could explore the possibility that AlzPathway and PD maps may have common pathogenic signaling molecules and reactions. We conducted cross-pathway analysis between AlzPathway and PD maps. The TREM2 gene could not appear as a common pathogenic signaling molecule because the TREM2 gene has not been compiled in the PD map yet. However, we found several common pathogenic signaling molecules and reactions including amyloid β precursor protein (APP) and tau (MAPT gene). Cross-pathway analysis may clarify pathogenic signaling molecules and reactions common between Alzheimer's disease and other neurodegenerative diseases at different stages of the disease.

AlzPathway is the first comprehensive map of intra, inter and extra cellular signaling pathways networks of a particular disease, towards deciphering pathogenesis of AD and to assist in the developing of AD drugs. AlzPathway is currently composed of 1,347 molecules, 1,070 reactions, and 129 phenotypes in neuron, brain

blood barrier, presynaptic, postsynaptic, astrocyte, and microglial cells and their cellular localizations. We are planning to update AlzPathway with support of natural language processing (NLP). AlzPathway is freely available and can be updated by the AD research community. AlzPathway provides a comprehensive resource to the AD community towards deeper insights into AD pathogenesis and identification of novel therapeutic targets.

4 Notes

1. The guideline for manual construction of AlzPathway can be summarized as follows: (1) Collection of review articles searched in PubMed. To manually elaborate AlzPathway, manual curation of ca. 100,000 AD research articles published after 2000 was needed, which was discarded. Instead of this, careful selection and manual curation of review articles with current state and understanding of AD pathogenic signaling proteins and signaling pathways was performed; (2) Manual curation of collected review articles. Manually curate AD review articles, and compile AD pathogenic signaling proteins and their signaling pathways using CellDesigner. Molecules should be distinguished, including the following types: proteins, complexes, simple molecules, genes, RNAs, ions, degraded products and phenotypes. Reactions should be also distinguished, including the following categories: state transition, transcription, translation, heterodimer association, dissociation, transport, unknown transition, and omitted transition. Evidences/links to articles should be described as PubMed IDs using the MIRIAM scheme for all reactions. Cellular types should be distinguished including the following: neuron, astrocyte, and microglial cells. Cellular compartments should include: brain blood barrier, presynaptic, postsynaptic, and cellular localizations.
2. CellDesigner is a structured diagram editor for drawing gene-regulatory, biochemical and signaling networks. Intuitive user-interface allows to draw a diagram in rich graphical notation. Notation is compliant with the PD (Process Description) of SBGN (Systems Biology Graphical Notation) [14].

Acknowledgements

This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Alzheimer's Society (2013) What is Alzheimer's disease? Alzheimers.org.uk. Last updated Mar 2012
2. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23
3. Mizuno S, Iijima R, Ogishima S et al (2012) AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst Biol 6:52
4. Funahashi A, Matsuoka Y, Jouraku A et al (2008) Cell Designer 3.5: a versatile modeling tool for biochemical networks. Proc IEEE 96:1254–1265
5. Laibe C, Le Novère N (2007) MIRIAM resources: tools to generate and resolve robust cross-references in systems biology. BMC Syst Biol 1:58
6. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127
7. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116
8. Hucka M, Finney A, Sauro HM et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531
9. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H (2010) Payao: a community platform for SBML pathway model curation. Bioinformatics 26:1381–1383
10. Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol 1:2005.0010
11. Oda K, Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. Mol Syst Biol 2:2006.0015
12. Calzone L, Gelay A, Zinovyev A et al (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol 4:173
13. Caron E, Ghosh S, Matsuoka Y et al (2010) A comprehensive map of the mTOR signaling network. Mol Syst Biol 6:453
14. Le Novère N, Hucka M, Mi H et al (2009) The systems biology graphical notation. Nat Biotechnol 27:735–741
15. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
16. Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821–7826
17. Fujita KA, Ostaszewski M, Matsuoka Y et al (2013) Integrating pathways of Parkinson's disease in a molecular interaction map. Mol Neurobiol 49:88–102

Part V

Computational Systems Biology, Network Biology: Next Generation Computational and Integrative Network Biology Approaches for the Study of Modules, Network Dynamics, and Their Interplay in Complex Diseases

Chapter 26

A Computational Network Biology Approach to Uncover Novel Genes Related to Alzheimer's Disease

Andreas Zanzoni

Abstract

Recent advances in the fields of genetics and genomics have enabled the identification of numerous Alzheimer's disease (AD) candidate genes, although for many of them the role in AD pathophysiology has not been uncovered yet. Concomitantly, network biology studies have shown a strong link between protein network connectivity and disease. In this chapter I describe a computational approach that, by combining local and global network analysis strategies, allows the formulation of novel hypotheses on the molecular mechanisms involved in AD and prioritizes candidate genes for further functional studies.

Key words Alzheimer's disease, Network biology, Bioinformatics, Protein interaction networks, Systems biology, Functional modules

1 Introduction

1.1 Protein Interaction Networks and Disease

There is solid evidence indicating that disease-related gene products do not usually have isolated cellular functions but they rather work together in the same macromolecular complex, sub-network or pathway through their interactions [1]. This points to a strong link between protein connectivity and disease [2].

The increasing availability of human interaction data from large-scale experiments [3–5] and literature mining [6–8] enabled the investigation of the properties of disease genes from a global perspective. Indeed, it has been suggested that disease genes have a tendency to encode proteins that have a larger number of interaction partners than non disease-related proteins and that they preferentially interact with one another [9, 10]. Concomitantly, different research groups focused their efforts on charting the interaction network of specific disorders, either using experimental [11–13] or computational approaches [14, 15], or a combination thereof [16], to find out putative causative genes or novel modifier genes of the disease phenotypes.

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_26,

Following these observations, we have recently mapped the most complete interaction network (i.e. interactome) associated to Alzheimer's disease (AD) [17]. Its analysis suggested a putative role for the programmed cell death protein 4 (PDCD4) as a neuronal death regulator and pinpointed the Evolutionarily Conserved Signaling Intermediate in Toll pathway protein (ECSIT) as a potential molecular link between oxidative stress, inflammation, and mitochondrial dysfunction in AD [18].

1.2 Overview of the Protocol

In this chapter, I illustrate a computational network-based approach to identify novel AD-related genes complemented by some theoretical and practical advice that takes into account the recent advance in the field of computational network biology (*see* Fig. 1). It relies on the construction of an AD interactome starting from a set of seed proteins implicated in the disease. The protein interaction data is gathered from public available repositories. Once the interactome is built, candidate gene products are mapped onto the network that is further enriched with orthogonal information

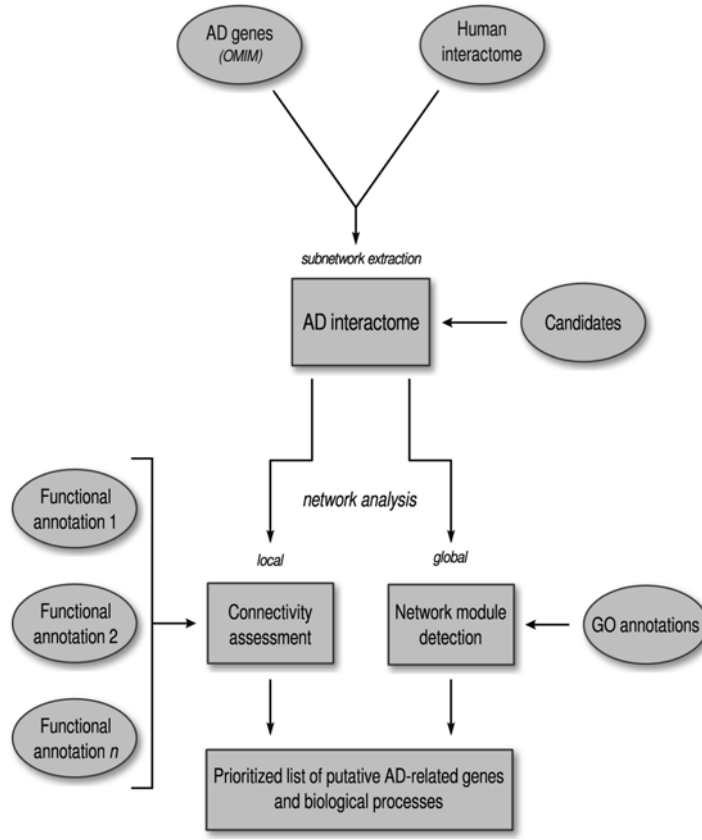


Fig. 1 Flow diagram of the computational network biology approach to uncover potential AD-related genes

such as gene expression profiling and Gene Ontology annotations. Subsequently, the analysis of the AD interactome local and global properties, including its network structure, facilitates the formulation of novel hypotheses on the molecular mechanisms involved in AD and prioritizes potential novel AD-related proteins for further functional studies (*see* **Note 1**).

2 Materials

2.1 Alzheimer's Disease Genes

The starting point is the selection of genes that are causally implicated in or contribute to the susceptibility to AD. Such information is available in the Online Mendelian Inheritance in Man database (OMIM) [19], a comprehensive compendium of human genes and disease phenotype relationships. Currently, the OMIM database lists 14 causative/susceptibility genes associated to the AD phenotype (*see* Table 1).

2.2 Candidate Genes

Gene lists generated by genetic or physical mapping experiments of human disease loci, or by large-scale association studies [20, 21], represent interesting sets of candidate genes that might be involved in AD mechanisms (*see* **Note 2**).

Table 1
Causative/susceptibility genes implicated in AD according to the OMIM database (as of September 2013)

Gene symbol	Name	OMIM ID
A2M	Alpha-2-macroglobulin	103950
ACE	Angiotensin converting enzyme	106180
APBB2	Amyloid beta A4 precursor protein-binding, family B, member 2	602710
APOE	Apolipoprotein E	107741
APP	Amyloid beta A4 protein	104760
BLMH	Bleomycin hydrolase	602403
HFE	Hereditary hemochromatosis protein	613609
MPO	Myeloperoxidase	606989
NOS3	Nitric oxide synthase	163729
PAXIP1	PAX interacting protein 1	608254
PLAU	Urokinase-type plasminogen activator	191840
PSEN1	Presenilin 1	104311
PSEN2	Presenilin 2	600759
SORL1	Sortilin-related receptor	602005

2.3 Protein Interaction Data

Several public databases gather experimentally verified protein physical interaction data [22] such as the BioGRID [6], IntAct [7] and MINT [8]. These resources provide data in standardized formats, such as the HUPO PSI-MI format [23], to make interaction information easily accessible to the scientific community. Usually, interaction datasets consist of a list of protein identifier pairs (*see* **Note 3**) along with the experimental technique(s) used for the interaction detection as well as the reference of the corresponding article(s) describing the given association (*see* **Note 4**).

2.4 Network Analysis Tools

Interaction data can be represented graphically as a network in which proteins correspond to *nodes* and interactions between protein pairs to *edges*. This representation enables several analyses based on graph theory principles that define the role of each protein and interaction in the network. For instance, one can measure network characteristics such as the *node degree* and the *clustering coefficient* (*see* **Note 5**) using specific functions developed for the most widely used scientific programming languages such as Java, Python and R or use the Cytoscape software (*see* **Note 6**). Another common practice is to study the global structure of the network by detecting clusters or modules, defined as groups of highly interconnected proteins that usually tend to be functionally related. In recent years, many algorithms have been developed for network cluster identification including MCODE [24], RNSC [25], CFinder [26], MCL [27], ClusterONE [28] and OCG [29]. Most of them are available as stand-alone programs or as plugins in Cytoscape (*see* **Note 7**).

2.5 Functional Annotations

1. Gene Ontology (GO) represents the state-of-the-art resource for functional information [30]. It consists of three structured controlled vocabularies (ontologies) that describe genes, and their products, in terms of biological processes, molecular functions and cellular components. Genes are associated with GO terms according to specific source reference(s) (e.g. a scientific paper) upon which the association is based. Every annotation reports an evidence code that indicates how the annotation is supported (*see* **Note 8**). GO annotations for human gene products are constantly updated and regularly released either by the NCBI Entrez [31] website or the UniprotKB Gene Ontology Annotation resource [32].
2. Gene expression profiling represents a valuable means for identifying the set of genes and, thus, the biological process or pathways that are dys-regulated in a given disease phenotype. In the case of AD, several datasets are available, such as the transcriptomics of the hippocampus [33, 34], the entorhinal cortex [35] or the neocortex [36] of post-mortem patient brain tissues compared to non-AD control samples (*see* **Note 9**).

3 Methods

3.1 AD Interactome Building

The first step to construct the AD interactome starting from the set of causative genes (named *AD seeds*) taken from the OMIM database, is to gather from public databases all the proteins identified as direct interactors of AD seeds (*see* **Note 10**). Subsequently, this initial AD network is further extended to the next level by including the interactors of the AD seed interactors and the interconnections among them, thus obtaining the AD interactome (*see* **Note 11**).

Once the AD interactome is built, the set of candidates can be identified on the network. It might be that no interaction information is available for some candidates. To tackle this problem, different strategies can be chosen. For instance, interaction identification experiments can be performed, using the candidates as baits, to discover new interactors or to test the ability of candidates to bind to AD seeds. These novel interactions would then be included in the AD interactome. If this option is not feasible, a valuable alternative can be the integration of protein interactions inferred by homology [37] from other organisms (*see* **Note 12**).

3.2 Functional Analysis of Candidates Based on Their Local Connectivity

Once the candidates are mapped onto the AD interactome, they can be firstly ranked according to their distance from the AD seeds. As stated in Subheading 1, the proteins associated with a given disease tend to interact with each other. Candidates interacting with one or more seeds are likely to be more informative than those that are not direct interactors and are located far away in the network. Such type of inference is also known as the *guilty-by-association* method [38]. However, given the incompleteness of the human interactome [39], the prioritization of candidates based solely on their proximity to seeds might be insufficient. In order to cope with this limitation, it is recommended to perform a functional analysis of the candidates using Gene Ontology annotations. Candidates involved, for example, in biological processes related to AD (i.e. such as beta-amyloid plaque formation, oxidative stress or neuronal death) or having similar functions and cellular localizations as AD seeds, can be considered as potential AD players. However, more intriguingly, such analysis can also pinpoint interesting candidates annotated with certain unexpected functions or subcellular localizations that are not yet associated to known AD genes, opening the avenue for new hypotheses.

The analysis based on GO annotations can be coupled with an additional layer of functional scrutiny by checking whether one or more candidate genes are related to the AD phenotype based on gene expression data. Indeed, the significant up- or down-regulation of candidate genes interacting with AD seeds can provide further clues on their involvement in the AD pathophysiology.

3.3 Global Analysis: The Modular Structure of the AD Interactome

The investigation of the global structure of the AD interactome can be useful to gain a deeper comprehension of the role of candidate genes in AD. The prevalent approach is to detect putative functional modules in a two-step procedure that consists in the identification of network modules, using one of the tools described in Section 2.4, and their subsequent annotation with functional information such as Gene Ontology.

1. The choice of the appropriate algorithm for the detection of network modules depends on different aspects [40, 41]. For instance, the topology of the network can affect the algorithm performance. Indeed, some algorithms work better on dense networks, whereas others are more suitable for sparse ones, thus influencing the number of modules that can be detected. Secondly, whereas some algorithms assign exclusively proteins to one module (e.g.: MCL and RNSC), others identify overlapping modules, meaning that a protein can belong to one or more modules (e.g. CFinder, ClusterONE, MCODE and OCG). Furthermore, most of them can take into account interaction weights in detecting modules (*see* **Note 13**). Finally, these algorithms require the setting of one or more running parameters.
2. Afterwards, the detected modules are annotated with Gene Ontology terms and their functional coherence is assessed using an homogeneity criterion [9], defined as the maximum number of proteins in the same module that have the same GO annotation(s) (*see* **Note 14**). The homogeneity computation may be affected by the structure of the network and by the granularity achieved running the module detection algorithm of choice. Therefore it is recommended to tune the parameters to maximize the functional coherence of the network modules detected (*see* **Note 15**). Alternatively, modules can be annotated by performing an enrichment analysis to identify over-represented Gene Ontology terms that can be used as functional labels (*see* **Note 16**).
3. The result will be a list of functionally homogenous/enriched network modules that can be further ranked depending on their functional relevance to AD or by assessing their statistical significance based on the content in either AD seeds or candidate genes, or both of them. This rationale can guide the selection of candidate genes for further functional validation and trigger many hypotheses on the molecular mechanisms implicated in AD.

3.4 Concluding Remarks

In this chapter I have described a general computational network-based approach that represents an ideal complement to genome-wide association studies and next generation sequencing techniques in the hunt for novel genes associated to human disorders, such as

Alzheimer's disease. Indeed, network biology strategies provide a global perspective to investigate the molecular mechanism underlying human diseases beyond the individual genes. Nevertheless, the functional clues arising from computational network analyses will need to be experimentally validated.

4 Notes

1. The described protocol can be implemented in any generic/scientific programming language such as Java, Python or R. Several libraries have been developed to perform network and functional analyses. For instance, the JUNG framework (for Java, http://jung.sourceforge.net/), NetworkX (for Python, http://networkx.github.io/) and igraph (both for Python and R, http://igraph.sourceforge.net) are good options for network generation and analyses, whereas Bioconductor [42] represents the state-of-the-art framework for biological analysis (for R http://www.bioconductor.org).
2. It is possible to integrate the list of candidate genes coming from in-house experiments with other candidate genes gathered from public available resources. For instance, the OMIM database reports several chromosomal regions for which the association with the AD phenotype was confirmed but that lack supporting evidence for a direct association of specific genes. An alternative source of candidates genes is the AlzGene database [43], which collects and analyzes the results of published genetic association studies carried out on AD phenotypes, including genome-wide association studies and other large-scale experiments. The most recent release of the AlzGene database stores information on more than 1,300 association studies and about 700 human genes. AlzGene applies strict criteria for including studies in its collection and scores genes according to the results of a meta-analysis based on established guidelines [44].
3. Protein interaction databases generally use UniprotKB [45] accession number as human protein identifiers. However, other database identifiers may be present such as NCBI Entrez [31] or Ensembl [46] gene accession numbers. One can unify different identifiers by using the Ensembl BioMart tool [47] or the UniprotKB ID mapping page (http://www.uniprot.org/?tab=mapping).
4. The number of experimentally verified physical interactions between human proteins is well above 130,000 and they were detected by distinct techniques (i.e. yeast two-hybrid, affinity purification/mass spectrometry, co-localization, etc.) in different conditions (e.g. in vivo or in vitro) and experimental settings

(high-throughput versus low-throughput). This draft human interactome captures biochemically possible and diverse associations (i.e. binary/multimeric, transient/obligate, etc.) with different degree of reliability. For this reason, a few scoring systems were developed to assess the experimental confidence of protein interaction data [48–50]. These scoring systems, usually ranging between 0 and 1, depend on several parameters such as the technique used for the interaction detection assays, the experimental setting, the number of times that an interaction was observed and the number of distinct scientific articles describing a given interaction. Although it can be useful to consider confidence scores when gathering protein interaction data, setting a stringent threshold would reduce the coverage of the interaction space producing a sparser network.

5. The *node degree* is the number of connections of the node with others in the network [51]. The *clustering coefficient* measures the extent of nodes in a graph that cluster together [52]. For a given node, it corresponds to the number of edges between its adjacent nodes, divided by the total number of possible edges between them.
6. Researchers that are not proficient in programming may take advantage of Cytoscape (http://www.cytoscape.org/), a freely-available software platform for the visualization and the analysis of complex networks [53]. Cytoscape provides many tools, also known as *plugins* or *apps*, to perform several types of analyses [54]. For instance, the topological properties of a network can be computed and analyzed by the *Network Analyzer* plugin.
7. In Cytoscape, the user can install additional tools by selecting the plugins/app manager option in the corresponding menu. To date, there are 20 plugins (e.g. ClusterViz, ClusterMaker [55] or Clust&See [56]) that provide easy-to-use module detection algorithms.
8. The GO annotation evidence codes fall into five categories: experimental, computational analysis, author statements, curational statements and inferred from electronic annotation. The latter type of annotation does not undergo manual curation. It is important to notice that evidence codes do not represent an annotation quality assessment, meaning that an experimental support of a GO annotation can be as good as an electronically inferred annotation. However, in the case of functional analysis of protein interaction network, it is preferable to discard annotations with the *IPI* (*Infer from Physical Interaction*) evidence code to avoid circularity problems.
9. Generally, the list of significantly up- and down-regulated genes is available in the supplementary materials of the corresponding publication, as in [36]. However, if this is not the

case, one can gather this information from gene expression profiling repositories such as the Gene Expression Omnibus (GEO) database [57]. Indeed, GEO provides a web-based application, called *GEO2R*, that allows the comparison of two or more groups of samples in order to identify the 'top 250' genes that are differentially expressed across experimental conditions.

10. As explained in **Note 3**, public databases store molecular interactions of different types, meaning that they are defined as *binary* or *co-complex* (i.e. multimeric) associations according to the technique used for the interaction detection [3]. On one hand, discovery methods such as yeast two-hybrid identify interactions between pairs of proteins that are likely to be direct and not mediated by others (i.e. binary). On the other hand, techniques such as affinity purification coupled to mass-spectrometry spot groups of interacting proteins for which it is not possible to define precisely who is interacting with whom.
11. If one is not familiar with scripting languages and automatic data extraction and integration from interaction databases, she/he may use one of the web tools dedicated to sub-network extraction from interactomes such as Mentha [58] or any of the data importer plugins (e.g. the PSICQUIC client [59]) provided by Cytoscape.
12. This approach is based on the following assumption: given two interacting proteins A and B in an organism and A' and B' their respective orthologs in another organism, then A' and B' are likely to interact with one another. The interaction A'–B' is a potentially conserved interaction across two organisms and is defined as an *interolog*. Inferred interaction data is available in public databases such as HomoMINT [60] or I2D [61], previously known as OPHID. Both resources provide interaction data in the standard PSI-MI format that can be easily integrated with the human interaction data described in Subheading 2.
13. Interaction networks can be weighted using an experimental confidence score as described in **Note 3**.
14. For the homogeneity computation, it is important to have enough GO annotation coverage of the network modules. For this reason, it is advisable to require that at least 50 % of the proteins that are present in the module to be annotated with at least one GO term. In addition, it is desirable to assess the statistical significance of every homogeneous module, comparing, for instance, its functional homogeneity to the mean value of a reference distribution obtained by computing the functional homogeneity for 1,000 randomly generated sets having the same module size. For sake of consistency, the proteins for the randomization should be picked from the human interactome used for the generation of the AD interactome.

15. In our original work [17] we detected the functional modules in the AD network using the MCL algorithm. Since the granularity of the clustering depends on one parameter, the inflation coefficient, *I*, we ran MCL on our AD network exploring a wide range of *I* (from 0.1 to 10.0 by steps of 0.1). We then choose the value of *I* that maximizes the number of functionally homogenous modules.
16. Several tools are available to perform functional enrichment analysis. For example, the BiNGO [62] and ClueGO [63] plugins in Cytoscape or the Bioconductor package GOstats [64] in R.

Acknowledgments

The author acknowledges the "Plan Cancer 2009-2013, Biologie des systèmes" funded by the French government for current support. The author would also like to thank Christine Brun and Daniela Ruffell for critically reading this chapter.

References

1. Oti M, Brunner HG (2007) The modular nature of genetic diseases. Clin Genet 71:1–11
2. Zanzoni A, Soler-Lopez M, Aloy P (2009) A network medicine approach to human disease. FEBS Lett 583:1759–1765
3. Rual J-F, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437:1173–1178
4. Stelzl U, Worm U, Lalowski M et al (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122:957–968
5. Ewing RM, Chu P, Elisma F et al (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol 3:89
6. Stark C, Breitkreutz B-J, Chatr-Aryamontri A et al (2011) The BioGRID interaction database: 2011 update. Nucleic Acids Res 39: D698–D704
7. Kerrien S, Aranda B, Breuza L et al (2012) The IntAct molecular interaction database in 2012. Nucleic Acids Res 40:D841–D846
8. Ceol A, Chatr Aryamontri A, Licata L et al (2010) MINT, the molecular interaction database: 2009 update. Nucleic Acids Res 38:D532–D539
9. Goh K-I, Cusick ME, Valle D et al (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690
10. Feldman I, Rzhetsky A, Vitkup D (2008) Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci U S A 105:4323–4328
11. Lim J, Hao T, Shaw C et al (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 125:801–814
12. Kaltenbach LS, Romero E, Becklin RR et al (2007) Huntingtin interacting proteins are genetic modifiers of neurodegeneration. PLoS Genet 3:e82
13. Camargo LM, Collura V, Rain J-C et al (2007) Disrupted in achizophrenia 1 interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. Mol Psychiatry 12:74–86
14. Chen JY, Shen C, Sivachenko AY (2006) Mining Alzheimer disease relevant proteins from integrated protein interactome data. Pac Symp Biocomput: 367–378
15. Liu B, Jiang T, Ma S et al (2006) Exploring candidate genes for human brain diseases from a brain-specific gene network. Biochem Biophys Res Commun 349:1308–1314
16. Pujana MA, Han J-DJ, Starita LM et al (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 39:1338–1349
17. Soler-Lopez M, Zanzoni A, Lluis R et al (2011) Interactome mapping suggests new mechanistic

details underlying Alzheimer's disease. Genome Res 21:364–376
18. Soler-López M, Badiola N, Zanzoni A, Aloy P (2012) Towards Alzheimer's root cause: ECSIT as an integrating hub between oxidative stress, inflammation and mitochondrial dysfunction: hypothetical role of the adapter protein ECSIT in familial and sporadic Alzheimer's disease pathogenesis. Bioessays 34:532–541
19. McKusick VA (2007) Mendelian inheritance in man and its online version, OMIM. Am J Hum Genet 80:588–604
20. Laird NM, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 7:385–394
21. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108
22. Orchard S (2012) Molecular interaction databases. Proteomics 12:1656–1662
23. Kerrien S, Orchard S, Montecchi-Palazzi L et al (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. BMC Biol 5:44
24. Bader GD, Hogue CWV (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2
25. King AD, Przulj N, Jurisica I (2004) Protein complex prediction via cost-based clustering. Bioinformatics 20:3013–3020
26. Adamcsek B, Palla G, Farkas IJ et al (2006) CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics 22:1021–1023
27. Van Dongen S, Abreu-Goodger C (2012) Using MCL to extract clusters from networks. Methods Mol Biol 804:281–295
28. Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 9:471–472
29. Becker E, Robisson B, Chapple CE et al (2012) Multifunctional proteins revealed by overlapping clustering in protein interaction network. Bioinformatics 28:84–90
30. The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res 38:D331–D335
31. NCBI Resource Coordinators (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41:D8–D20
32. Barrell D, Dimmer E, Huntley RP et al (2009) The GOA database in 2009—an integrated Gene Ontology Annotation resource. Nucleic Acids Res 37:D396–D403
33. Blalock EM, Geddes JW, Chen KC et al (2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci U S A 101:2173–2178
34. Blalock EM, Buechel HM, Popovic J et al (2011) Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. J Chem Neuroanat 42: 118–126
35. Dunckley T, Beach TG, Ramsey KE et al (2006) Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. Neurobiol Aging 27:1359–1371
36. Tan MG, Chua W-T, Esiri MM et al (2010) Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. J Neurosci Res 88:1157–1169
37. Matthews LR, Vaglio P, Reboul J et al (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res 11:2120–2126
38. Oliver S (2000) Guilt-by-association goes global. Nature 403:601–603
39. Venkatesan K, Rual J-F, Vazquez A et al (2009) An empirical framework for binary interactome mapping. Nat Methods 6:83–90
40. Brohée S, van Helden J (2006) Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics 7:488
41. Song J, Singh M (2009) How and when should interactome-derived clusters be used to predict functional modules and protein function? Bioinformatics 25:3143–3150
42. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80
43. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23
44. Ioannidis JPA, Boffetta P, Little J et al (2008) Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol 37:120–132
45. The UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res 41: D43–D47
46. Flicek P, Ahmed I, Amode MR et al (2013) Ensembl 2013. Nucleic Acids Res 41:D48–D55
47. Kinsella RJ, Kähäri A, Haider S et al (2011) Ensembl BioMarts: a hub for data retrieval

across taxonomic space. Database (Oxford) 2011:bar030
48. Chatr-Aryamontri A, Ceol A, Licata L, Cesareni G (2008) Protein interactions: integration leads to belief. Trends Biochem Sci 33:241–242, author reply 242–243
49. Braun P, Tasan M, Dreze M et al (2009) An experimentally derived confidence score for binary protein-protein interactions. Nat Methods 6:91–97
50. Kamburov A, Stelzl U, Herwig R (2012) IntScore: a web tool for confidence scoring of biological interactions. Nucleic Acids Res 40:W140–W146
51. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45: 167–256. http://arxiv.org/abs/cond-mat/0303516/
52. Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440–442
53. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
54. Saito R, Smoot ME, Ono K et al (2012) A travel guide to Cytoscape plugins. Nat Methods 9:1069–1076
55. Morris JH, Apeltsin L, Newman AM et al (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics 12:436
56. Spinelli L, Gambette P, Chapple CE et al (2013) Clust&See: a Cytoscape plugin for the identification, visualization and manipulation of network clusters. Biosystems 113:91–95
57. Barrett T, Wilhite SE, Ledoux P et al (2013) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41: D991–D995
58. Calderone A, Castagnoli L, Cesareni G (2013) mentha: a resource for browsing integrated protein-interaction networks. Nat Methods 10:690–691
59. Aranda B, Blankenburg H, Kerrien S et al (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Methods 8:528–529
60. Persico M, Ceol A, Gavrila C et al (2005) HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. BMC Bioinformatics 6(Suppl 4):S21
61. Brown KR, Jurisica I (2005) Online predicted human interaction database. Bioinformatics 21:2076–2082
62. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448–3449
63. Bindea G, Mlecnik B, Hackl H et al (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25: 1091–1093
64. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258

Chapter 27

Network Approaches to the Understanding of Alzheimer's Disease: From Model Organisms to Humans

Justin Yerbury, Dan Bean, and Giorgio Favrin

Abstract

It is becoming increasingly evident that Alzheimer's disease cannot be considered as the outcome of a single pathway, but rather we should view it as a system, that is, a network of interactions between large numbers of different protein molecules. In the last few years, probably because of the inherent limitations of traditional methods and because of the great increase in availability of sequencing data, this type of approach is being used more and more. In the following, we will discuss what constitutes a "network approach," what are its pros and cons, a number of recent case studies and finally what are the future perspectives of this type of analysis.

Key words Alzheimer's disease, Network approaches, Physical and genetic interactions, Systems biology, Bioinformatics

1 Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia and is characterized by early and progressive episodic memory impairment along with other cognitive changes. It is associated with brain volume loss in regions such as the hippocampus, entorhinal cortex and amygdala and atrophy in the medial temporal lobe but ultimately much of the brain is affected [1]. The pathological hallmarks of AD are amyloid plaques and neurofibrillary tangles. Plaques and tangles have been shown to be composed primarily of amyloid-β (Aβ) peptide and hyper-phosphorylated tau, respectively. The early onset familial forms of the disease are due to inheritance of mutations in APP, PSEN1 and PSEN2, all of which affect the excision of the Aβ peptide from APP, providing strong evidence that the amyloid cascade is an important part of disease pathogenesis [2]. For this reason, the pathogenesis of AD is often simplified to single pathway built around the amyloid cascade. However, one must consider that mutations in genes that modify Aβ processing account for only 5 % of patients (termed Early Onset

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_27,

or Familial AD). The remaining 95 % of cases are termed Late-Onset alzheimer's Disease (LOAD), and in these cases there is not a clear effect on the processing of the Aβ peptide (Covered in more detail in "Why do we need a new approach to understand LOAD," below). In fact, there are many other aspects of AD pathology that may be vital to the understanding of the pathogenesis of the disease. AD is also associated with aberrant lipid metabolism [3], neuroinflammation such as microglia activation and astrogliosis [4], oxidative stress [5], and metal homeostasis [6]. The connections between and the relative role of all these pathways remains elusive mainly because the lack of good disease models [7].

2 Why Do We Need a New Approach to Understand LOAD?

In the last two decades work has focused on elucidating the mechanisms by which protein aggregation initiates a cascade of events that lead to cellular toxicity and ultimately to the death of neurons. These approaches have culminated in the development of a number of candidate therapies based upon blocking the production/aggregation of the Aβ peptide, however so far none have passed clinical trials. One possible explanation for this is that the underlying pathways responsible for early onset inherited AD are not responsible for LOAD; this is supported by the finding that production of Aβ does not increase in LOAD [8]. As current disease models are based on over production of Aβ it is probable these models will always fail to fully replicate the human disease, particularly LOAD. Indeed, while the formation of amyloid plaques in aged APP transgenic mice recapitulates amyloid pathology in human AD none of the current models currently reproduces all aspects of human AD, for instance neurofibrillary tangles, significant neurodegeneration and cerebral atrophy. Consequently, the mouse models may already inadequately represent the familial forms of AD they are based on. Given the lack of a representative model, it is unlikely that the complexity of LOAD can be fully appreciated today.

The most informative data concerning LOAD to date has come from genome wide association studies (GWAS). Understanding the role that genetic variability plays in the pathogenesis of LOAD has been a major focus of investigation for over a decade. The strongest genetic risk for LOAD is the *APOE* gene, which has become the benchmark against which new genetic risk factors are compared. In addition to approximately 20 genetic loci linked to the risk of LOAD [9], a recent meta-analysis of 74,046 individuals with LOAD identified 11 new susceptibility loci for AD. These loci highlight pathways such as endocytosis, the immune response and lipid processing as important in LOAD. Despite such large-scale studies, it is thought that much of the heritability of

LOAD is yet to be uncovered [10]. This missing heritability in complex diseases such as LOAD has been proposed to stem from large numbers of variants whose effects are small but are cumulative or from those rare variants (occurring in less than 5 % of the population) with potential for larger effects when found in combination with other mutations or environmental changes [11]. With the current techniques it is seemingly impossible then to find this missing variability or "dark matter" of heritability as the effect of any one perturbation may be too small to detect individually despite a causative role in disease pathology. It is only through their connections in the cellular network that the effect of several small perturbations emerges.

It is becoming increasingly evident that AD cannot be considered as the outcome of a single pathway (i.e. one cause leading to one effect). Instead a paradigm shift is needed in order to understand the complexity of disease pathogenesis in which many small perturbations together can cause the same clinical outcome as a single large perturbation due to interactions in the cellular network. One way forward is to interrogate LOAD as a "system"—a network of interactions between large numbers of different protein molecules. Indeed evidence is just beginning to emerge suggesting that analysis of the LOAD GWAS data with networks in mind can identify nodes and/ or sub-networks that are not detectable when examining the data for individual hits [12], thus shining light on some of the "dark matter."

In summary, AD is an extremely complex disorder, which is not confined to a single protein aggregation pathway. Indeed, AD pathology encompasses several diverse pathways and may involve tens or even hundreds of different genes. It is probable that this complexity hides the fact that AD (and other neurodegenerative diseases) is not a single well defined condition but a spectrum of conditions.

3 What Is a Network Approach?

Constructing biological networks allows us to visualize biological data, such as metabolic pathways and their interrelationships, and consider single elements such as genes or proteins as part of a bigger picture. Formally, networks are described using graph theory [13] in which a network consists of nodes connected by edges (Fig. 1). For example if the nodes are proteins then edges may represent a physical interaction between two proteins. This formalism can be applied to biological networks at many different levels, from individual reactions to whole populations. At the single cell level, we may consider an interaction network for signaling molecules, metabolic reactions, protein-protein interactions or genetic interactions. In the context of AD, generally the state of one

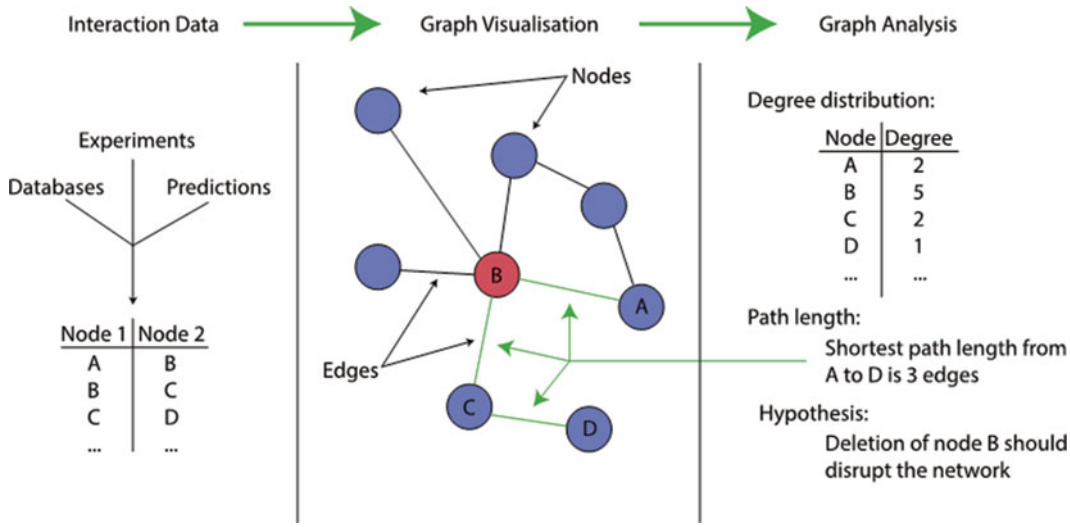


Fig. 1 An overview of a typical network analysis pipeline. Interaction data is collected from various sources (*left panel*) and used to construct a graph of the network (*central panel*). In addition to being a visualization tool, network graphs can be formally analyzed in a number of ways (examples in *right panel*), allowing quantitative comparisons of different networks

neuron is considered with the use of a diagram containing the set of all relevant proteins or genes connected by edges representing different types of interactions between them. One major advantage of using these formal networks is that their properties can be studied mathematically and/or computationally [13].

The precise regulation of the various cellular pathways is responsible for the correct and healthy functioning of an organism. In this view, a disease such as AD can be seen as the dysregulation a number of pathways, for example in Early Onset AD the perturbation of the system is at the level of APP metabolism but the affects cascade through the cellular network to involve many other downstream pathways, ultimately resulting in the loss of neurons. In the case of LOAD, we expect there to be a (potentially large) number of small perturbations that may be spread throughout the network. The phenotypic outcome (AD) is related to these perturbations through the interaction network in the cell. It is known that many different perturbations can cause the same (or a clinically very similar) outcome, and again the reason for this must reside in the connections between different pathways, or the occurrence of multiple perturbations in a given pathway, in the cell. A network-based approach allows important pathways to be identified even if their effect is small or indirect. Studying these underlying networks is therefore essential to our fundamental understanding of the pathology of this highly complex disease.

Experimental data from sources such as GWAS or microarray will yield a list of significantly altered genes. In order to study the

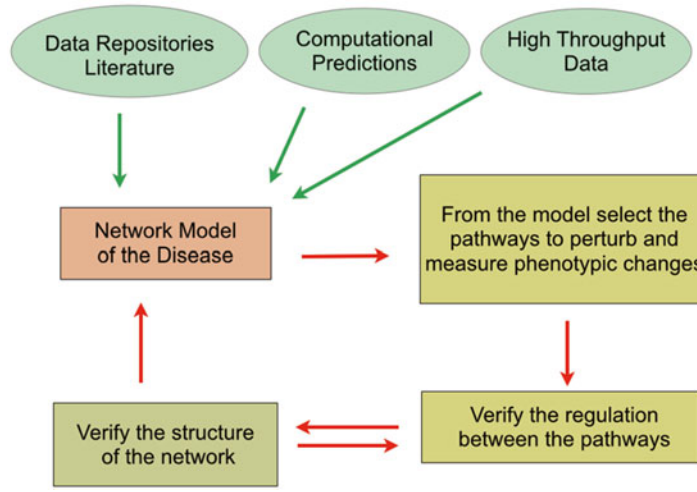


Fig. 2 Schematic representation of a network approach. For detailed information see text

relationships between these genes, including whether they belong to the same or parallel pathways, or their relationships to other genes that play important roles in these pathways, one can use a network approach which can be outlined in five simple steps which are part of an iterative procedure (Fig. 2).

(a) Build the network of interactions (Interactome) combining our initial gene (protein) list with data available from: (1) literature, (2) computational predictions and (3) high-throughput experiments.

(b) Generate data for normal and diseased samples, typically in high-throughput experiments.

(c) Determine the differences that exist between the normal (healthy) and disease networks. Formulate hypotheses of dysregulation that could, given the network structure, lead to the observed phenotype (e.g. protein aggregation and neuronal loss).

(d) Test these hypotheses.

(e) Update the network structure and restart the loop.

Given a large set of interaction data we can extract the subset of interactions that contain the genes (proteins) in our initial list [14]. Genetic or protein-protein interaction data can be obtained in three possible ways [15]: (1) Compilation or curation of existing data from literature, (2) Computational predictions based on available information, and (3) Direct measurements in high-throughput experiments. In the following sections we discuss each of these in turn.

3.1 Compilation or Curation of Existing Data from Literature

A number of repositories already exist for Interactome data (e.g. BioGRID, HPRD, IntAct [16–18] *see* Subheading 4.2 below). Such repositories are extremely valuable for constructing the network for the genes (or protein, organism, etc.) of interest, however an element of caution is necessary. It is easy to forget that these databases, though extremely large, remain incomplete. Furthermore the data they contain should be subject to the same precautions as any other experimental data.

A case in point is the high false positive rate for some methods used to find interactions [19] in high-throughput experiments, and many interactions may not have been validated. It is advised that users filter interaction data, based on the experimental procedure from which it was generated, to determine the likelihood of a given edge. Additionally different repositories may use different methods to place edges between the members of a protein complex (i.e. whether to add interactions between all pairs of proteins in the complex or not), resulting in different interaction networks. Finally, the data suffers from a "sociological bias," that is, some genes of interest have been study in more depth and therefore they appear to have a larger number of interactors [19]. It is essential to bear these caveats in mind when working with interaction data.

3.2 Computational Predictions Based on Available Information

The String database [20] alongside the experimental data provides a number of interactions which have been computationally predicted, using information available from many databases and different model organisms. For example if orthologous proteins A and B exist in mice and humans respectively, we can use the known interactions of A in mice to predict interactions of B in humans.

Additionally, the application of network inference techniques may be able to predict the structure of a given network based on the response to a perturbation [21].

3.3 Direct Measurements in High-Throughput Experiments

There are several techniques in widespread use to measure different types of interaction at high-throughput. For protein-protein (physical) interactions, these include techniques such as yeast two-hybrid screens [22], affinity capture MS [23] and tandem affinity purification tagging [24].

Common techniques for measuring genetic interactions include Synthetic Genetic Array (SGA) [25] technology, which is being used to systematically score all pairs of gene-gene interactions in baker's yeast [26]. Genetic interactions can be quantified in higher organisms, generally by constructing a single mutant of interest and then performing random mutagenesis e.g. with transposable elements.

Protein-DNA interactions may be relevant to disease biology for example by modifying the expression of specific genes or large sets of genes. There are several techniques such as Chromatin

Immunoprecipitation (ChIP) combined with DNA microarray technology (ChIP-chip) or with high-throughput sequencing (ChIP-Seq) that can be used to identify these interactions.

4 Differential Networks and Their Application to AD

The generation of high-throughput interaction data provides a snapshot of the system being studied. The data found in large repositories can be used to represent normal (healthy) cells. However when we apply network theory to AD, we want to understand what parts of the network are relevant to the disease. The concept of differential network biology [27] encompasses this goal with the following approach. Considering samples from two conditions, A (normal) and B (stressed), interaction data for both samples can be generated separately and networks directly compared. Intuitively, we would say that the interactions likely to be the most relevant to the stress phenotype are those that change the most between the two networks, whereas interactions that are constant are not likely to be important to this process. We can therefore construct a differential network using only those edges that are different between the two samples. It is important to realize that these edges will not necessarily be the strongest edges in either network, but those with the greatest relative change in strength between the networks. For physical networks (protein–protein or protein–DNA), differential interactions imply mechanistic changes that are a result of an organism's response to environmental conditions. For genetic networks (synthetic-lethal or epistasis), interactions reflect functional consequences of mutations, not direct physical mechanisms. Going forward, building a differential interaction map [28] will be the priority in order to understand the dysregulation of the pathways involved in AD.

4.1 What Network Approaches Are Currently Being Used to Study AD?

Systems biology approaches are finally becoming mainstream in the AD field [12, 29, 30]. Such studies have demonstrated that network-level changes are evident in interaction networks constructed for AD patients versus controls [12]. In particular, several groups are starting to construct network maps of important pathways, built either from data available from the literature or from experimental data sets [31].

In a case study, Zhang and co-workers [12] constructed connectivity networks using microarray data from 1,647 post mortem brain tissues from 376 patients with late onset AD and 173 non-demented healthy controls. These networks represent modules of the interactome that are remodeled in the diseased state. The authors then used a Bayesian model to infer probable regulators of these sub-networks. The highest scoring group of genes was found to involve immune/microglia genes. In particular TYROBP was

the highest scoring causal regulator network with other genes previously associated with AD such as TREM2, MS4A4A, MS4A6A and CD33. This prediction was then tested in a separate in vitro study where it was found that TYROBP is directly involved in amyloid-β turnover and neuronal damage. This study serves as an example of the power that an integrated systems approach brings to the challenge of understanding the causal changes involved in LOAD.

Recently [14] a network approach very similar to the one we have described previously has been used to analyze data from a series of whole exome sequencing (WES) experiments in patients with Hereditary Spastic Paraplegias (HSP). From the WES the authors identified 15 novel genes implicated in the disease. Using these candidate genes in addition to the genes previously known to be involved in the disease they constructed a network of protein-protein interactions by using data from publicly available databases (iREFINDEX [32], ConsensusPathDB [33], human interactome database (http://interactome.dfci.harvard.edu/H_sapiens/). This network allowed the authors to (1) better identify modules of HSP pathology and (2) identify new genes implicated in the disease from WES data on disease affected families not previously analyzed, thus demonstrating that this type of approach can indeed facilitate the identification of biological pathways and genes implicated in the disease. Importantly, the authors have also found a significant similarity between the HSP genes and genes involved in other neurodegenerative disorders such as AD, Parkinson's and Amyotrophic Lateral Sclerosis, indicating that there are a number of pathways which appear to be dysregulated in these type of disorders and that better understanding of these pathways in one disease will help the understanding of other diseases within this class.

4.2 Databases and Repositories

In the last few years a number of new databases and repositories have become available as well as new tools to mine their data. Interaction data (physical and genetic) is being deposited in a multitude of repositories such as BioGRID [16, 34], IntAct [18], HPRD [17], MINT [18] etc. As well as the tools provided to access each repository individually, a number of tools are being developed that either facilitate the retrieval of the data from multiple sources in parallel such as PSICQUIC [35] or tools that create a unifying index that allows the unequivocal identification of each interaction (IRefINDEX [32]) and in doing so facilitate the comparisons between the various databases.

Another set of important tools to mine a variety of databases are the InterMine [36] powered databases such a FlyMine [37], ModMine [38], YeastMine [39] and others. InterMine databases can be mined with powerful queries that can be customized completely by users. Importantly both the results and queries can be saved. Additionally, the InterMine API enables programmatic

access to this data from a number of programming languages. The necessary code can be automatically generated from either a preset or user-defined query designed online. Queries can span diverse datasets e.g. expression, interactions, citations, genomic features etc. and organisms (e.g. orthology relationships).

4.3 Software/ Libraries

The last few years have seen also an explosion in the number of the software tools available to visualize and analyze the biological data. One of the most widely used tools is certainly Cytoscape [40]. Cytoscape is a tool for both visualization and analysis of network data. External plugins [41] may be loaded, which allow the user to calculate or retrieve important properties of the network such as the gene ontology (GO) annotations (BINGO [42]) or the relative KEGG pathways (KEGG Parser [43]) etc. Lastly we should mention Bioconductor [44], a curated repository of open source R packages for bioinformatics.

4.4 Statistical Analysis of Networks

Having built a network of interactions it is often important to try to compare it with other networks or lists of genes (or proteins). Some example metrics for comparison are the calculation of: (1) the overlaps i.e. how many nodes are in common between the two networks); (2) Node degree (how many edges a node is connected to) distribution. This may highlight that a given network contains hub nodes (highly connected nodes) whereas another does not; and (3) Average path length. This is the average shortest distance (number of edges) between all pairs of nodes in a network, and can be used as a measure of the efficiency of information transfer through the network. For directed networks, it is important to bear in mind edge directionality for these metrics.

In order to calculate whether these differences are statistically significant we need to compare with random networks that satisfy both of the following criteria: (a) that the network is constructed from random lists of the same size as the one we used initially and (b) should contain a similar number of nodes to the network we want to compare them to. These conditions are equivalent to stating that the average connectivity of the nodes within the random networks should be the same as in our network.

It is of particular importance to enforce these conditions as networks of genes (proteins) associated with disease have a significantly higher degree of connectivity [19] and therefore are not as easily comparable to random ones.

4.5 Experimental Validation of Networks and Nodes

As the power of GWAS studies is limited by the sample size, it is becoming very difficult to discover new disease associated genes in disorders such as AD. For this reason network-based approaches are rapidly becoming an important tool for the discovery of novel disease modifiers mostly because they represent a rapid and cost effective way to produce new hypotheses, which then will have to

be tested experimentally. Typically the prediction from a disease network will be that a given gene, or set of genes, are potential modifiers of disease pathology. This hypothesis then needs to be tested either in an animal model, or more commonly in cell culture.

5 Network Approaches to AD: Looking to the Future

Since the identification of a link between mutations affecting Aβ processing and AD, the "amyloid cascade" has been central to the field of AD research. Clinically however, patients with mutations directly linked to Aβ processing are rare. Today it is becoming clear that each patient will carry a set of mutations that would individually may have no or very little toxic affect but that together cause the disease via their roles in the cellular network. Ultimately a precise understanding of the pathways involved in the disease will help the design of a more accurate and targeted drugs to combat the AD. With the increase in the availability of "omics" data is becoming easier to build interaction networks, which will help us to achieve a better and more precise definition of what AD is and which pathways become dysregulated in the disease as a function of time. Importantly as these pathways are present also in the healthy individual, we need to understand how they change and become dis-regulated in the diseased condition. We need in other words differential disease networks [28].

Several neurodegenerative diseases are now being approached as each being a spectrum of related disorders. As we enter the era of "personalized medicine," it is possible that an understanding of how an individuals set of genetic mutations would lead to pathological changes could allow treatment to be tailored accordingly.

Network based approaches to AD promise many advantages compared to more traditional ones, but the techniques are still largely in their infancy, due to both limitations on the availability of experimental data, and the ongoing development of analysis methods. We strongly believe that giving open access to the disease associated databases, will push the development of the field and will ultimately lead to more sophisticated and powerful methodologies for data mining.

References

1. Ballard C, Gauthier S, Corbett A et al (2011) Alzheimer's disease. Lancet 377:1019–1031
2. Karran E, Mercken M, De Strooper B (2011) The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. Nat Rev Drug Discov 10:698–712
3. Di Paolo G, Kim TW (2011) Linking lipids to Alzheimer's disease: cholesterol and beyond. Nat Rev Neurosci 12:284–296
4. Eikelenboom P, Bate C, Van Gool WA et al (2002) Neuroinflammation in Alzheimer's disease and prion disease. Glia 40:232–239

5. Praticò D (2008) Oxidative stress hypothesis in Alzheimer's disease: a reappraisal. Trends Pharmacol Sci 29:609–615
6. Zatta P, Drago D, Bolognin S, Sensi SL (2009) Alzheimer's disease, metal ions and metal homeostatic therapy. Trends Pharmacol Sci 30:346–355
7. Couzin-Frankel J (2013) When mice mislead. Science 342(922–923):925
8. Mawuenyega KG, Sigurdson W, Ovod V et al (2010) Decreased clearance of CNS beta-amyloid in Alzheimer's disease. Science 330: 1774
9. Fogh I, Ratti A, Gellera C et al (2014) A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. Human Mol Genet 23:2220–2231
10. So HC, Gui AH, Cherny SS, Sham PC (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 35:310–317
11. Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–753
12. Zhang B, Gaiteri C, Bodea LG et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease. Cell 153:707–720
13. Fisher J, Henzinger TA (2007) Executable cell biology. Nat Biotechnol 25:1239–1249
14. Novarino G, Fenstermaker AG, Zaki MS et al (2014) Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. Science 343:506–511
15. Vidal M, Cusick M, Barabási A-L (2011) Interactome networks and human disease. Cell 144:986–998
16. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S et al (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41:D816–D823
17. Keshava Prasad TS, Goel R et al (2009) Human protein reference database—2009 update. Nucleic Acids Res 37:D767–D772
18. Orchard S, Ammari M, Aranda B et al (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42:D358–D363
19. Goh K-I, Cusick M, Valle D et al (2007) The human disease network. Proc Natl Acad Aci U S A 104:8685–8690
20. Franceschini A, Szklarczyk D, Frankild S et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–D815
21. Albert R (2007) Network inference, analysis, and modeling in systems biology. Plant Cell 19:3327–3338
22. Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proc Natl Acad Sci U S A 97:382–7387
23. Turecek F (2002) Mass spectrometry in coupling with affinity capture-release and isotope-coded affinity tags for quantitative protein analysis. J Mass Spectrom 37:1–14
24. Gregan J, Riedel CG, Petronczki M et al (2007) Tandem affinity purification of functional TAP-tagged proteins from human cells. Nat Protoc 2:1145–1151
25. Tong AH, Evangelista M, Parsons AB et al (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368
26. Costanzo M, Baryshnikova A, Bellay J et al (2010) The genetic landscape of a cell. Science 327:425–431
27. Ideker T, Krogan NJ (2012) Differential network biology. Mol Syst Biol 8:565
28. Markowetz F (2010) How to understand the cell by breaking it: network analysis of gene perturbation screens. PLoS Comput Biol 6:e1000655
29. Griciuc A, Serrano-Pozo A, Parrado A et al (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 78:631–643
30. Gandy S, Heppner F (2013) Microglia as dynamic and essential components of the amyloid hypothesis. Neuron 78:575–577
31. Ogishima S, Mizuno S, Kikuchi M et al (2013) A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery. Clin Pharmacol Ther 93:399–401
32. Razick S, Magklaras G, Donaldson IM (2008) iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics 9:405
33. Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB—a database for integrating human functional interaction networks. Nucleic Acids Res 37:D623–D628
34. Stark C, Breitkreutz BJ, Reguly T et al (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34:D535–D539
35. Aranda B, Blankenburg H, Kerrien S et al (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Methods 8:528–529
36. Smith RN, Aleksic J, Butano D et al (2012) InterMine: a flexible data warehouse system for

the integration and analysis of heterogeneous biological data. Bioinformatics 28:3163–3165

37. Lyne R, Smith R, Rutherford K et al (2007) FlyMine: an integrated database for Drosophila and Anopheles genomics. Genome Biol 8:R129
38. Contrino S, Smith RN, Butano D et al (2012) modMine: flexible access to modENCODE data. Nucleic Acids Res 40:D1082–D1088
39. Balakrishnan R, Park J, Karra K et al (2012) YeastMine—an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) 2012:bar062
40. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
41. Saito R, Smoot ME, Ono K et al (2012) A travel guide to Cytoscape plugins. Nat Methods 9:1069–1076
42. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448–3449
43. Arakelyan A, Nersisyan L (2013) KEGGParser: parsing and editing KEGG pathway maps in Matlab. Bioinformatics 29:518–519
44. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Chapter 28

Characterization of Genetic Networks Associated with Alzheimer's Disease

Bin Zhang, Linh Tran, Valur Emilsson, and Jun Zhu

Abstract

At the molecular level, the genetics of complex disease such as Alzheimer's disease (AD) manifests itself as series of alterations in the molecular interactions in pathways and networks that define biological processes underlying the pathophysiological states of disease. While large-scale genome-wide association (GWA) studies of late-onset alzheimer's disease (LOAD) have uncovered prominent genomic regions linked to the disease, the cause for the vast majority of LOAD cases still remains unknown. Increasingly available large-scale genomic and genetic data related to LOAD has made it possible to comprehensively uncover the mechanisms causally lined to LOAD in a completely data-driven manner. Here we review the various aspects of systems/network biology approaches and methodology in constructing genetic networks associated with AD from large sampling of postmortem brain tissues. We describe in detail a multiscale network modeling approach (MNMA) that integrates interaction and causal gene networks to analyze large-scale DNA, gene expression and pathophysiological data from multiple post-mortem brain regions of LOAD patients as well non-demented normal controls. MNMA first employs weighted gene co-expression network analysis (WGCNA) to construct multi-tissue networks that simultaneously capture intra-tissue and inter-tissue gene–gene interactions and then quantifies the change in connectivity among highly co-expressed genes in LOAD with respect to the normal state. Co-expressed gene modules are then rank ordered by relevance to pathophysiological traits and enrichment of genes differentially expressed in LOAD. Causal regulatory relationships among the genes in each module are then determined by a Bayesian network inference framework that is used to formally integrate genetic and gene expression information. MNMA has uncovered a massive remodeling of network structures in LOAD and identified novel subnetworks and key regulators that are causally linked to LOAD. In the end, we will outline the challenges in systems/network approaches to LOAD.

Key words Alzheimer's disease, Late-onset alzheimer's disease, LOAD, Gene co-expression network, Bayesian network, Causal regulator, Multiscale network modeling, MNMA

1 Introduction

Alzheimer's disease (AD) is the most common form of degenerative dementia and is expected to affect over 100 million people worldwide in 2050 [1]. Rare mutations in *APP*, *PSEN1* and *PSEN2* have been identified in early-onset familial AD. For the

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_28,

more common late-onset alzheimer's disease (LOAD), a recent whole-genome scan of common variants in 17,000 LOAD patients identified 11 new genetic loci (*CASS4*, *CELF1*, *DSG2*, *FERMT2*, *HLA-DRB5–HLA-DRB1*, *INPP5D*, *MEF2C*, *NME8*, *SORL1*, *SLC24A4-RIN3* and *ZCWPW1*) in addition to the ten well-established GWAS-defined risk factors (*ABCA7*, *APOE*, *BIN1*, *CD33*, *CLU*, *CR1*, *CD2AP*, *EPHA1*, *MS4A6A-MS4A4E* and *PICALM*) [2]. Increasingly available large scale genomic, genetic and pathophysiological data in LOAD have made it possible to more comprehensively address the complex mechanisms and effectors of LOAD through application of advanced systems biology approaches. This approach provides insights into the interacting pathways underlying LOAD but also the detailed regulatory circuits, which in turn enable identification of novel causative pathways and potential key causal regulators. In the past decade, several complementary systems approaches have been developed to understand the causal mechanisms underlying LOAD based on large molecular profiling data.

Weighted gene co-expression network analysis (WGCNA) [3] was employed to identify genes and pathways involved in regional and cell type changes in LOAD [4–6]. Co-expressed gene modules associated with immune response, synaptic transmission, metabolic process were identified in different brain regions from different LOAD stages that were defined by the MiniMental State Examination (MMSE) score and neurofibrillary tangle (NFT) burden. Hub genes in each module were also identified and they are predicted to play key roles in these pathways. However, none of these genome-wide expression studies (GWES) can pinpoint causal pathways and key causal regulators underlying LOAD as WGCNA can only highlight correlation/association relationship among genes and the number of samples is too small to capture sufficiently perturbed complex biological processes underlying AD.

In another systems biology approach [7], differential co-expression analysis (DCA) was applied to identify *APOE* ε4 effectors in LOAD. DCA basically integrates differential expression and differential co-expression (correlation) into a single score to capture potential master regulators of a disease state such as LOAD. Twenty *APOE*-dependent candidate modulators of LOAD such as *RNF219*, *SV2A* and *HDLBP* have been identified. These potential APOE ε4 effectors include an *APP* processing gene *ITM2B* and three *APP* trafficking genes, *TMEM59L*, *FYN* and *APBA2*. However, it is still unclear how these genes are modified in LOAD and how they function in the context of interacting pathways to promote LOAD as again no effort was made to construct causal networks to dissect these co-regulated genes.

To comprehensively and objectively identify and characterize molecular interactions associated with LOAD, we developed a unique network-based integrative approach to integrate large scale

genetic, genomic and pathophysiological data from multiple post-mortem brain regions of LOAD patients as well non-demented normal controls [8]. The key component of the approach is to combine multiscale networks including gene co-expression and causal networks for identification of novel subnetworks and key regulators that are causally linked to LOAD.

2 Materials

The gene expression and genotype data were generated in 1,647 brain specimens from three brain regions including dorsolateral prefrontal cortex (PFC), Visual cortex (VC) and Cerebellum (CB) in 376 LOAD patients and 273 non-demented subjects. The microarray design, RNA sample preparation, amplification, hybridization, and DNA isolation were previously described in detail [8]. Briefly, tissue samples were profiled on a custom-made Agilent 44 K array of 40,638 DNA probes and genotyped for 838,958 single nucleotide polymorphisms (SNPs) on the Illumina HumanHap 650Y array and a custom Perlegen 300 K array. Neuropathology traits such as Braak stage, general and regional atrophy, gray and white matter atrophy and ventricular enlargement were assessed for all the LOAD subjects. The gene expression data as well as the sample information are available at the GEO database under accession number GSE44772.

3 Methods

3.1 Architecture of a Multiscale Network Modeling Approach to Characterize Genetic Networks

A core component of MNMA is integration of correlation/interaction and causal networks, as shown in Fig. 1. At first, correlation/association analyses are performed on each pair of data types and this gives rise to genes that are significantly correlated with each clinical trait and single nucleotide polymorphisms (SNPs) that are associated with gene expression (eSNPs) and/or clinical traits (cSNPs). Gene expression data are then employed to construct gene co-expression networks using an extended WGCENA termed as weighted interaction network analysis (WINA) and gene modules that are comprised of highly interacting genes are then correlated with clinical traits to determine their relevance to clinical endpoints. The average connectivity between the genes in each module of co-expression networks in LOAD is then compared with that of the same set of genes in the normal control using modular differential connectivity (MDC). MDC captures the change of coordination/co-expression among a group of genes and thus complements the conventional differential expression analysis for individual genes. The causal relationships among the genes in each module are

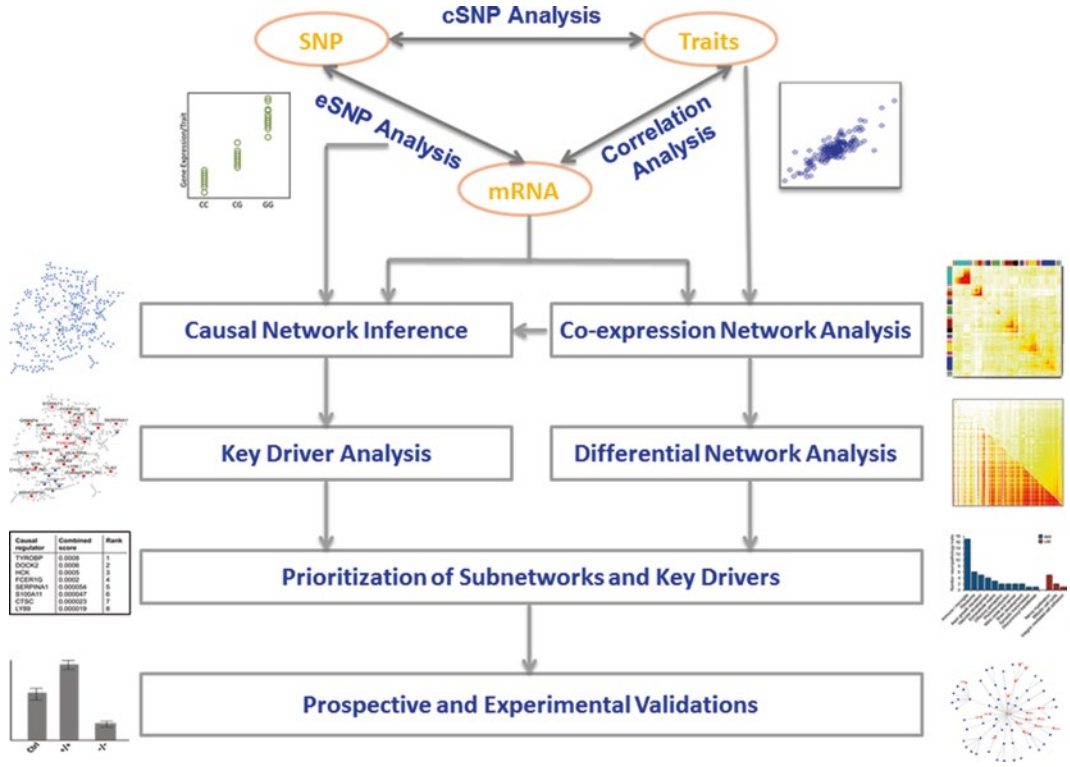


Fig. 1 Architecture of the multiscale network modeling approach to integrating DNA, mRNA expression and clinical data. Correlation/association analyses are first performed on each pair of data types. Gene expression data is then employed to construct gene co-expression networks using gene co-expression network analysis to identify gene modules that are comprised of highly interacting genes are then correlated with clinical traits to determine their relevance to clinical endpoints. Differential network analysis is then employed to identify modules with significant change in connectivity. The causal relationships among the genes in each module are then determined based on gene expression data and eSNP information by Bayesian network (BN) inference. BNs are then used to identify key regulators based on the key driver analysis (KDA). Co-expressed gene modules and key causal regulators are then prioritized and go through in silico validation and the top ranked drivers in the most relevant subnetworks/modules are then selected for experimental validation

then determined by Bayesian network (BN) inference. BNs are constructed based on gene expression data and prior information derived from the eSNP analysis. BNs are then used to identify key causal regulators based on the key driver analysis (KDA). To identify subnetworks most relevant or causally linked to AD, we combine module relevance, module differential connectivity, enrichment of differential expression and enrichment LOAD GWAS hits to compute a total relevance score for rank ordering all the modules. To further prioritize key causal regulators, we employ a machine learning approach to formally combine these causal drivers from multiple brain regions. Causal regulators then go through in silico validation and the top ranked drivers in the most relevant subnetworks/modules are then selected for experimental validation.

3.2 Association Analysis of DNA, mRNA and Clinical Data

Association of mRNA expression and clinical traits is computed by either Pearson or Spearman correlation. False discovery rate is estimated through the permutation of samples. Association of SNPs with mRNA expression and clinical traits is calculated using Kruskal Wallis test. Expression quantitative trait loci (e)QTLs for gene expression traits were determined by identifying the SNP most strongly associated with each expression trait profiled on the array over all the N genotyped SNPs. *Cis* analysis was limited to SNPs located within 1 Mb of either side of the transcription start or end within the gene body, while *trans* effects were defined as the associated SNPs located on a different chromosome to the physical location of the corresponding probe [9]. The association P-value was adjusted to control for testing of multiple SNPs and traits (expression or clinical) through the Bonferroni correction and the empirical FDR [10].

6993, 8836, 4634 eSNP genes were identified in CB, PFC and VC in LOAD, respectively while 5116, 6375, 4385 eSNP genes were found in CB, PFC and VC in the normal control. For all the three brain regions, eSNPs in LOAD overlap very significantly (Fisher's exact test $P<10^{-241}$) with those in the normal control though less one third of them are conserved, suggesting genetics plays a big role in regulating gene expression in LOAD. eSNPs, especially *cis* eSNPs, will be used to test the genetic signal in the co-expressed gene modules and to serve as prior information for causal network inference.

3.3 Constructing Tissue-Specific and Multi-Tissue Co-expression Networks

The association analysis discussed previously establishes links between individual features (gene expressions, SNPs and clinical traits), but it doesn't present any global patterns of those interactions/associations. A pressing task in analyzing large-scale genomic and genetic data is to identify and visualize a global landscape of interactomes that contribute to clinical endpoints such as AD severity. WGCNA has emerged to solve this problem through identification of gene modules comprised of highly interconnected genes over a gene–gene interaction heatmap [3] (*see* **Note 1**). Given the gene expression data from multiple tissues in this study, we can use WGCENA to construct both tissue-specific and multi-tissue gene co-expression networks in LOAD and the normal control. Multi-tissue gene networks can simultaneously uncover both intra-tissue and inter-tissue gene–gene interactions.

The weighted network analysis begins with a matrix of the similarities such as Pearson correlations between all gene pairs (*see* **Note 2**), then converts the correlation matrix into an adjacency matrix using a power function $f(x)=x^{\beta}$ (*see* **Note 3**). The parameter β of the power function is determined in such a way that the resulting adjacency matrix, i.e., the weighted co-expression network, is approximately scale-free. To measure how well a network satisfies a scale-free topology, we use the fitting index, i.e., the model fitting index R^2 of the linear model that regresses $\log(p(k))$

on $\log(k)$, where k is connectivity and $p(k)$ is the frequency distribution of connectivity [3]. The fitting index of a perfect scale-free network is 1. The connectivity between genes or *kij* is a transformed correlation between the expression profiles of two genes, $|r(i,j)|\beta$, with r as the similarity. The parameter $\beta(>0)$ of the power function is determined in such a way that the global probability distribution of the resulted connectivity values for all the gene pairs is scale free.

To explore the modular structures of a co-expression network, we further transform the corresponding adjacency matrix into a topological overlap matrix (TOM) and then employ average linkage hierarchical clustering to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules [11]. To distinguish between modules, each module is assigned a unique color identifier, with the remaining, less well connected genes colored grey.

To compare and contrast two multi-tissue networks, we combined their TOM heat-maps into a single large network. In the combined heat-map, the upper panel shows the hierarchical clustering on the TOM of the LOAD network while the color bar below represents the gene modules. Similarly, the lower panel represents the TOM from normal multi-tissue network. The color intensity in the map represents the interaction strength between genes. This connectivity map highlights how genes in the multi-tissue transcriptional networks fall into distinct network modules, where genes within a given module are more highly interconnected with each other (blocks along the diagonal of the matrix) than with genes in other modules.

In construction of a multi-tissue gene network, we assign each probe in a tissue with a new unique probe identifier by combining its original probe ID and the tissue name and then select the N most varying probes in each tissue to build up a multi-tissue gene expression data that includes $3*N$ probes and the subjects with all the three brain regions profiled.

Figure 2a–c show the tissue-specific gene co-expression networks in CB, PFC and VC in LOAD, respectively. Figure 2d shows a multi-tissue network based on the one-third most varying genes in each of the three brain regions in the LOAD subjects. 111 modules were identified in the LOAD multi-tissue network (Fig. 2d), each containing between 30 and 1,446 gene members, while the network generated from non-demented samples has 89 modules ranging in size from 30 to 2,278 genes. Figure 2e is the unweighted counterpart of the weighted multi-tissue gene co-expression network. Many gene modules are enriched for the genes in known pathways. The most enriched pathways are immune response (P=8.1e–91, 3.3-fold), extracellular matrix (P=1.3e–32, 2.9-fold), transmission of nerve impulse (1.2e–25, 4.8-fold) and

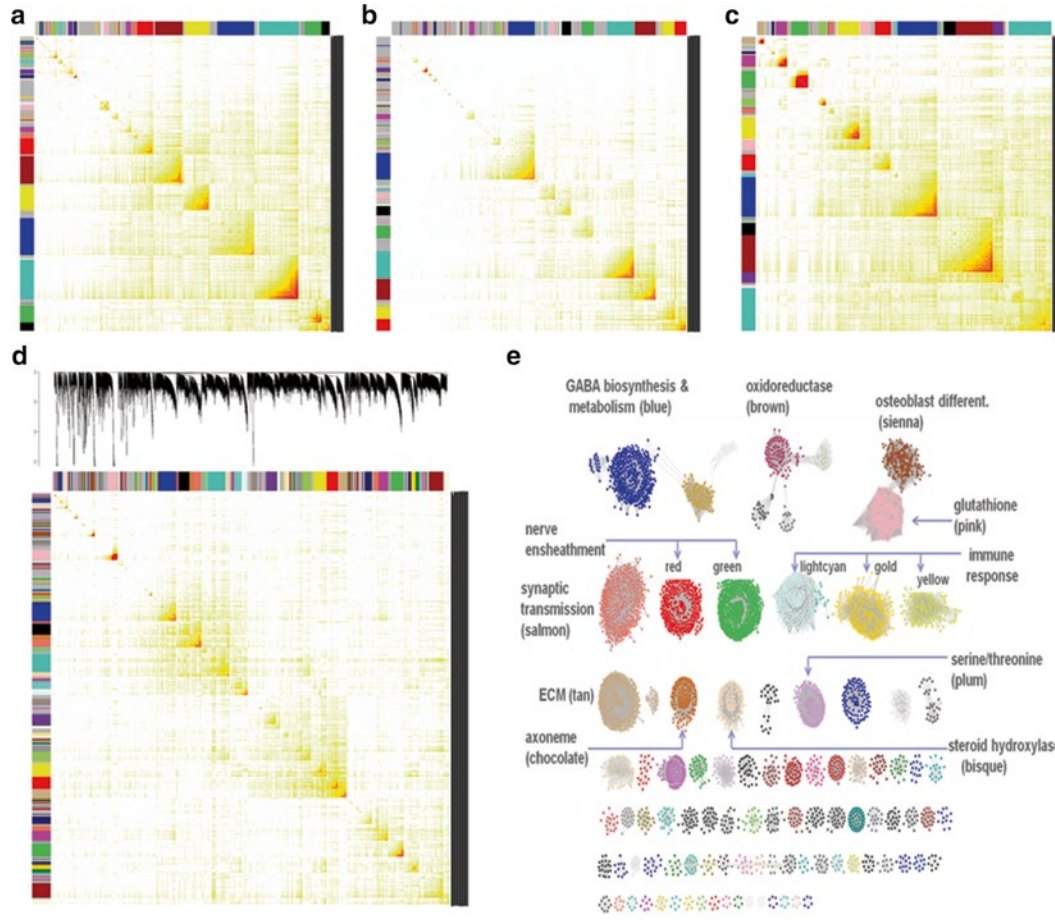


Fig. 2 Tissue-specific and multi-tissue weighted gene co-expression networks. The symmetric heat map with rows and columns as genes represents the network connection strength (indicated by the intensity of *red* color) between any pair of nodes (genes) in the network. The network connection strength is measured as the topological overlap between genes. The network modules highlighted as colored bars along the rows and columns were identified via an average linkage hierarchical clustering algorithm using topological overlap as the dissimilarity metric. (**a**) CB gene co-expression network. (**b**) PFC gene co-expression network. (**c**) VC gene co-expression network. (**d**) Multi-tissue gene co-expression network using the gene expression data profiled from CB, PFC and VC. (**e**) Visualization of the un-weighted counterpart of the weighted co-expression network shown in (**d**)

chaperone (P=1.8e−23, 23.9-fold). Among the 111 modules, only 33 modules are comprised of genes from a single brain regions and a vast majority include genes from two or three regions, suggesting very strong cross-region interactions.

3.4 Differential Network Analysis

To characterize dysregulation of gene networks in LOAD, we have defined modular differential connectivity (MDC) to quantify the difference between the connectivity among a set of genes (or module, denoted Ω) in LOAD versus normal nondemented networks.

Given a set of N genes and two networks, x and y, MDC is the ratio of the average connectivity among the N genes in the network x to that among the same gene set in network y (*see* **Note 4**), specified by the formula below:

$$\delta_{\Omega}(x,y) = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} k_{ij}^{x}}{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} k_{ij}^{y}}, \quad (1)$$

where, k_{ij} is the connectivity between two genes i and j in a given network.

The significance or false discovery rate (FDR) of the statistic MDC is accessed by permuting the data underlying the two networks. We differentiate two scenarios, gain of connectivity $\left(\delta_{\Omega}(x,y) > 1\right)$ and loss of connectivity $\left(\delta_{\Omega}(x,y) < 1\right)$. Given M permutations, FDR of MDC is computed as follows:

$$\mathrm{FDR}\left(\delta_{\Omega}(x,y) > 1\right) = \frac{1}{M}\sum_{p=1}^{M} \delta_{\Omega}(x,y) < \delta_{\Omega}\left(x^{p}, y^{p}\right),$$

or,

$$\mathrm{FDR}\left(\delta_{\Omega}(x,y) < 1\right) = \frac{1}{M}\sum_{p=1}^{M} \delta_{\Omega}(x,y) > \delta_{\Omega}\left(x^{p}, y^{p}\right),$$

where x^{p} and y^{p} are the networks derived from the permuted data. To rigorously access significance of MDC, we estimate two types of FDR estimates, one based on shuffled samples, i.e. networks with non-random nodes but random connections, and the other based on shuffled gene labels, i.e. networks with random nodes but non-random connections, and then we select the larger value as the final FDR estimate.

Among the 111 modules in the multi-tissue network (Fig. 2d, e), 70 modules have significant change in connectivity at FDR < 10 %. Five of the 70 modules with significant MDC have a loss of connectivity (MDC < 1) and they are enriched for genes in nerve ensheathment, gamma-aminobutyrate (GABA) biosynthesis and metabolism, neurotrophin TRK signaling and integrin-mediated cell adhesion, respectively. Figure 3 contrasts the paired TOM plots of the global and modular co-expression networks. In Fig. 3a, the topological overlap matrix (TOM) plots correspond to the LOAD (the upper right panel) and non-demented (the lower left panel) multi-tissue co-expression networks. The rows and columns represent the same set of the top one-third most variably expressed genes in each of the three brain tissues and states, expressed in a symmetric fashion and sorted by the hierarchical clustering tree of the LOAD network. Figure 3b shows individual TOM plots of three differentially connected modules in LOAD

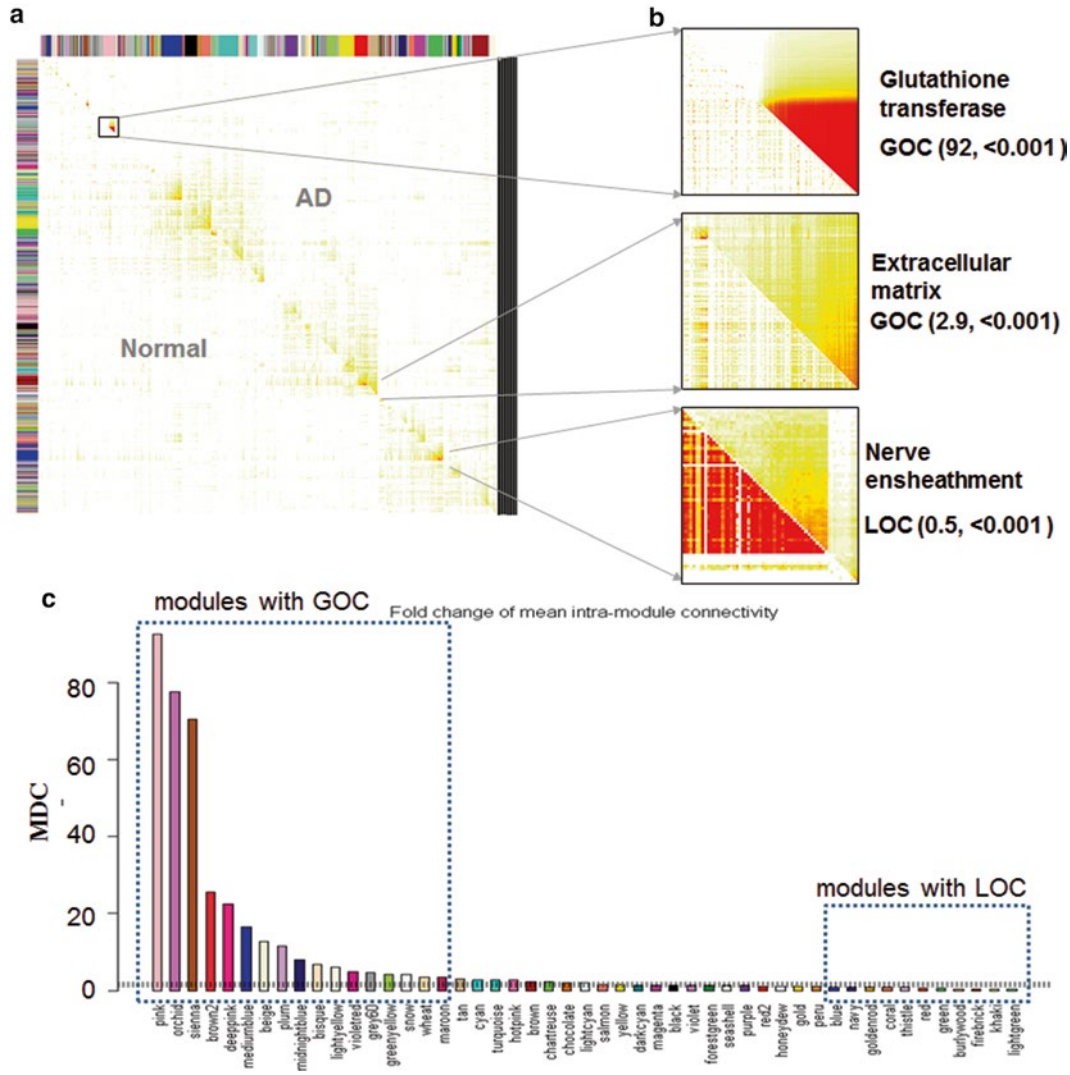


Fig. 3 Differential network analysis. (**a**) The topological overlap matrix (TOM) plots correspond to the LOAD (the *upper right panel*) and non-demented (the *lower left panel*) multi-tissue co-expression networks. The rows and columns represent the same set of the top one-third (13,193 in total) of the most variably expressed genes in each of the three brain tissues and states, expressed in a symmetric fashion and sorted by the hierarchical clustering tree of the LOAD network. (**b**) Individual TOM plots of three differentially connected modules in LOAD (the *upper right panel* of each module) versus that in the non-demented state (the *lower left panel* of each module). Differential connectivity (MDC) and FDR estimate is specified in each panel in parenthesis (MDC, FDR). Two modules involved in Glutathione transferase and extracellular matrix have a gain of connectivity (GOC) and the one associated with nerve ensheathment has a loss of connectivity (LOC). (**c**) MDC plot of 49 modules with at least 100 genes

(the upper right panel of each module) versus that in the non-demented state (the lower left panel of each module), demonstrating that LOAD reconfigures specific portions of the molecular interaction structure. Differential connectivity (MDC) and FDR estimate is specified in each panel in parenthesis (MDC, FDR).

Two modules involved in Glutathione transferase and extracellular matrix have a gain of connectivity (GOC) and the one associated with nerve ensheathment has a loss of connectivity (LOC). Figure 3c shows the MDC of 49 modules with at least 100 genes (*see* also Table 1).

3.5 Determination of Module Relevance to LOAD Pathology

To examine how each gene module was related to LOAD neuropathology traits, we first performed principal components analysis (PCA) for each module and then computed module-trait relevance using two complementary approaches: (1) the correlation between the first principal component (Module Eigengene) and each trait and (2) the correlation (the square root of R-square) between the top principal components and each trait through multivariate regression model. The significance (*P*-value) and FDR of each correlation was also calculated. FDR was estimated through random permutation of sample names of the trait data. A module is associated with a trait if both correlation *P*-value and FDR are below 0.05. The total number of traits associated with a module is used to quantify the association of a module with LOAD.

Of all modules, the immune/microglia showed correlation to the greatest number of LOAD-related neuropathology traits. Expression of the PFC immune/microglia module correlated to atrophy levels in multiple brain regions, including frontal cortex ($r=0.27$, FDR = 0.018), parietal ($r=0.20$, FDR = 0.016), temporal ($r=0.19$, FDR = 0.022) and neostriatum regions ($r=0.28$, FDR = 3.3e−09) as well as ventricular enlargement ($r=0.17$, FDR = 0.031).

3.6 Reconstruction of the Bayesian Causal Networks

Integration of genetic, genomic information and gene expression data into causal networks has been successfully used in dissecting causal relationships in complex human diseases such as diabetes and obesity [12, 13] and in a yeast model [14–16]. Bayesian network (where a joint probability can be decomposed to several conditional probabilities as $p(X_1,,\ldots,,X_n)=\prod_{i=1}^{n} p\left(X_i \mid Pa\left(X_i\right)\right)$) is one type of probabilistic causal networks, providing a natural framework for integrating highly dissimilar types of data (*see* **Note 5**). Integration of genetic information and gene expression data has led to the identification of several novel genes that are causal for obesity [13, 17]. For example, we constructed structure priors for integrating genetics information as [18]

$$p(A->B)=\frac{2\times\sum_i p(A->B\mid A,B,l_i)}{\sum_i p(A->B\mid A,B,l_i)+p(B->A\mid A,B,l_i)},$$

$$p(A->B)=1-\frac{\sum_i p(A\perp B\mid A,B,l_i)}{\sum_i 1} \text{ and}$$

Table 1
Top 35 modules most relevant to LOAD

Module	Ranking order	module size	CB number	PFC number	VC number	GO gene category	GO FET *P*	GO corrected *P*	GO fold enrichment	MDC	MDC FDR
Yellow	1	1318	4	1312	2	Response to biotic stimulus	1.90E–96	8.10E–91	3.25	1.49	0
Pink	2	983	90	149	744	Glutathione transferase activity	2.10E–03	1	7.09	92.67	0
Gray12	3	52	1	0	51	Calcium-dependent cell–cell adhesion	6.80E–03	1	16.08	4.26	0.04
Seashell	4	322	0	322	0	Cytoskeletal protein binding	4.50E–07	0.1887	3.2	1.29	0.16
Gray1	5	60	0	60	0	Cell junction	2.00E–05	1	10.42	0.82	1
Gray17	6	47	47	0	0	Phagocytosis	2.60E–04	1	81.57	1.09	0.34
Cyan	7	735	5	3	727	Response to external biotic stimulus	1.90E–22	7.99E–17	2.68	2.84	0
Green	8	1217	1211	3	3	Nerve ensheathment	6.50E–11	2.78E–05	11.56	0.5	0
Gray35	9	34	5	26	3	Fertilization (sensu Metazoa)	9.60E–06	1	64.87	4.13	0.62
Gray31	10	36	36	0	0	Exocytosis	1.30E–05	1	26.57	0.67	0.88
Cyan4	11	62	19	24	19	Ribosomal protein	6.80E–05	1	34.41	0.85	1
Gold2	12	99	3	96	0	Receptor-mediated axon growth repulsion	4.50E–09	0.001887	28.34	3.27	0
Red3	13	77	3	63	11	Ribosome	6.80E–08	0.02886	27.56	24.93	0
Green yellow	14	861	0	859	2	Response to unfolded protein	8.70E–15	3.77E–09	10.43	4.15	0
Red	15	1189	1	1188	0	Nerve ensheathment	1.50E–08	0.006549	10.11	0.68	1
Turquoise	16	1741	5	32	1704	Oxidoreductase activity	5.20E–05	1	8	2.79	0
Gray6	17	53	0	28	25	Antibacterial response protein	4.40E–04	1	63.56	1.16	0.46
Gray29	18	36	1	23	12	B-cell- and antibody-mediated immunity	1.40E–03	1	34.96	34.85	0
Goldenrod	19	114	39	36	39	Heart development	2.10E–03	1	29.13	0.88	1

(continued)

Table 1
(continued)

Module	Ranking order	module size	CB number	PFC number	VC number	GO gene category	GO FET *P*	GO corrected *P*	GO fold enrichment	MDC	MDC FDR
Gray14	20	49	20	13	16	mRNA transcription	3.40E−03	1	293.68	0.99	1
Light cyan	21	624	3	9	612	Defense response	2.10E−87	9.10E−82	4.46	2.01	0
Tan	22	830	0	830	0	Extracellular region	3.00E−38	1.33E−32	2.89	2.88	0
Blue	23	1673	14	20	1639	Cadherin	2.70E−12	1.22E−06	4.24	0.92	1
Gray21	24	44	44	0	0	Odorant binding	9.90E−05	1	31.24	1.19	0.32
Gray18	25	46	46	0	0	Vitamin biosynthesis	1.80E−04	1	97.89	3.82	0.02
Light yellow	26	506	362	127	17	mRNA cleavage factor complex	3.40E−04	1	20.74	6.09	0
Gray7	27	53	0	53	0	Protein processing	8.70E−04	1	15.58	1.8	0.14
Midnight blue	28	687	665	20	2	Synaptic transmission	1.50E−13	6.55E−08	4.07	8.03	0
Light green	29	519	1	0	518	Nerve ensheathment	2.50E−10	0.00010989	20.02	0.25	0
Navy	30	287	283	4	0	Immune response	1.10E−08	0.004773	2.42	0.92	1
Gold3	31	75	1	73	1	Microtubule family cytoskeletal protein	1.30E−06	0.5661	16.68	12.12	0
Gray13	32	52	52	0	0	Neuronal activities	1.00E−05	1	7.13	0.86	1
Gray19	33	45	13	0	32	Homeobox transcription factor	2.80E−05	1	21.29	122.65	0
Gray4	34	54	0	54	0	Adenosine receptor activity	2.00E−04	1	91.77	1.03	0.44
Brown	35	1435	1423	6	6	Oxidoreductase activity	3.10E−04	1	7.69	2.34	0

The modules are rank-ordered by a combined score based on correlations to LOAD pathophysiological traits, enrichment for differentially expression genes in LOAD, enrichment for LOAD specific eSNPs, enrichment for known GWAS hits and enrichment for AD related genes curated in AlzhDB. Column names: "module," module assignment; "module size," number of probes in a module; "CB number," number of probes from CB; "PFC number," number of probes from PFC; "VC number," number of probes from VC; "GO Gene Category," most enriched Gene Ontology (GO) Category; "GO FET *P*," Fisher's exact test (FET) *P* value for enrichment of the GO Category; "GO Corrected *P*," multiple-test corrected FET *P* value for enrichment of the GO Category; "GO Fold Enrichment," fold enrichment of the GO enrichment test; "MDC," Module Differential Connectivity (MDC); "MDC FDR," False Discovery Rate of MDC

$p(A->B) = 2 \times \frac{1+n(B)}{2+n(A)+n(B)}$ for causal/reactive, independent and undetermined relationships, respectively.

In general, Bayesian networks can only be solved to Markov equivalent structures, so that it is often not possible to determine the causal direction of a link between two nodes even though Bayesian networks are directed graphs. However, the Bayesian network reconstruction algorithm can take advantage of the experimental design by incorporating genetic data to break the symmetry among nodes in the network that lead to Markov equivalent structures, thereby providing a way to infer causal directions in the network in an unambiguous fashion [19]. Because *cis* eSNPs are in LD with causal variants that affect the expression levels of a neighboring gene or they are the causal variant themselves, they serve as an excellent source of natural perturbation to infer causal relationships among genes and between genes and higher order phenotypes like disease [9, 20]. We modified the reconstruction algorithm to incorporate eSNP data as priors, in the following way: genes with *cis* eSNP [21] are allowed to be parent nodes of genes without *cis* eSNPs, but genes without *cis* eSNPs are not allowed to be parents of genes with *cis* eSNPs, $p(trans \rightarrow cis) = 0$. A software package *Reconstructing Integrative Molecular Bayesian Networks* (RIMBANet) [15, 16, 18, 19, 22] has been developed to construct probabilistic causal networks for integrating diverse data and is free available at http://icahn.mssm.edu/departments-and-institutes/genomics/about/software/rimbanet.

The computational complexity of our MCMC method described above is expressed as $O(N^4)$, where the number of nodes included in the network reconstruction process is N. It is practically impossible to construct a global Bayesian network including all 39,000 genes from three different brain regions. Thus, we constructed a Bayesian network for each individual co-expression module.

3.7 Identification of Key Causal Regulators

One of the primary goals of network analysis is to distinguish causal regulators from passengers. Our data-driven multiscale networks present a great opportunity to robustly and accurately identify causal drivers of the biological processes underlying AD. Through co-expression network analyses, we identified a set of gene modules, for which Bayesian networks are reconstructed to derive module-based Bayesian subnetworks. For each Bayesian subnetwork, we then identify the key drivers (regulators) by leveraging several centrality measurements in the directed network [15, 23, 24]. By definition, each driver modulates a set of genes, i.e., its downstream nodes. Previous work has shown that a gene's function can be predicted by its neighboring genes in the network [25]. Moreover, a series of validation experiments have shown that the

downstream nodes of a driver predicted by Bayesian networks significantly overlap with its knockout signature [15]. For each Bayesian network of individual modules, we further identified the regulators by examining the number of N-hob downstream nodes (NHDN) for each gene in the directed network [18, 24, 26]. For a given network, let μ be the numbers of N-hop downstream nodes and d be the out degrees for all the genes. Genes with a number of NHDN greater than $\bar{\mu}+\sigma(\mu)$, were nominated as causal regulators. The regulators with degree above $\bar{d}+2\sigma(d)$, where d denotes the number of downstream genes, become key causal regulators of a corresponding network module associated with LOAD differential connectivity. These criteria identified genes with number of downstream nodes and number of out links significantly above the corresponding average value.

Figure 4 shows the Bayesian network of an immune/microglia enriched gene module and its key drivers (colored in red). The BN was constructed based on the aforementioned procedure. The top key drivers of this network are *TYROBP*, *FCER1G*, *CYBA*, *FYB*, *SLC7A7*, *DOCK2*, *TBXAS1*, *SERPINA1*, *FPR1*, *TLR2*, and *SYK*. *TYROBP*, TYRO protein tyrosine kinase binding protein, encodes a transmembrane signaling polypeptide which contains an immunoreceptor tyrosine-based activation motif (ITAM) in its cytoplasmic domain. *TYROBP* along with its putative receptor, triggering receptor expressed on myeloid cells 2 (*TREM2*) plays a key role in immune response. Mutations in *TYROBP* are linked to Nasu-Hakola disease [27]. *TYROBP* signals via spleen tyrosine kinase (*SYK*) to activate *SYK* mediating neutrophils and macrophages integrin-mediated activation. Interestingly, mutations in *TYROBP* have been found to lead to presenile dementia in human and impaired immune responses in mice [28, 29].

3.8 Prioritization of Subnetworks/ Modules and Key Drivers

To identify subnetworks most relevant or causally linked to AD, we combine module relevance, module differential connectivity, enrichment of LOAD specific eSNP genes, enrichment of differential expression, enrichment of differential expression and enrichment AD GWAS hits to compute a total relevance score so as to rank order all the 111 modules in the multi-tissue network.

To further prioritize key drivers in all three different brain regions, we employed a machine learning approach [30] to formally combine the ranking orders of key drivers from multiple brain regions:

$$G_j = \prod_i g_{ji}, \tag{2}$$

where, g_{ji} is the discriminant value of a candidate driver j in one brain region i, and is defined based upon the ranking order r_{ij}:

$$\left(\max_i\left(r_{ji}\right)+1-r_{ji}\right)/\sum_j r_{ji} \quad (3)$$

As we reported previously [8], a PFC-specific microglia/immune enriched module is the most relevant to LOAD, followed by the modules involved in glutathione transferase activity (pink module), calcium-dependent cell–cell adhesion (gray12 module), cytoskeletal protein binding (seashell), cell junction (gray1), phagocytosis (gray17), response to external biotic stimulus (cyan), and nerve ensheathment (green).

The three region-specific microglia modules share eight causal regulators, *CTSC*, *HCK*, *TYROBP*, *SERPINA1*, *S100A11*, *LY86*, *DOCK2* and *FCER1G*. Through the combined ranking score by Eq. 2, *TYROBP* ranked the highest. *TYROBP* is also the top driver of the CB-specific immune/microglia module, as shown in Fig. 4. Causal regulators then go through in silico validation and the top ranked drivers in the most relevant subnetworks/modules are then selected for experimental validation.

3.9 Summary and Future Work

Complex human diseases like AD involve many different pathways across many different tissues. The integration of large scale molecular profiling data, genotypic data, clinical data and other biologically relevant information is critical for us to understand more comprehensively how various molecular pathways interplay with each other to give rise to disease. As we showed here and in previous systems level studies of AD, the disease state can be depicted by complex networks of co-regulated genes and proteins. Accurate reconstruction of such networks will not only establish a global picture of interacting pathways underlying AD but also yield detailed regulatory circuits, which can further be employed to identify potential key drivers in an unbiased manner. Indeed, the drivers we identified here based on the multiscale networks are more likely to be the known AD susceptibility genes [31] than the non-causal regulators. In particular, our microglia/immune enriched network not only provides a global model to unite the previous GWAS risk loci including *MS4A4A*, *MS4A6A*, *CD33* and *TREM2* but also presents detailed direct or indirect connections between them. Concentration of the known genetic risk factors of LOAD in the networks strongly suggests the causal role of this microglia/immune network, thus demonstrating the power of the multiscale network approach. More importantly, these data-driven networks enable us to connect many missing components so as to develop more biological meaningful hypotheses for experimental validation.

Given the high complexity of molecular pathology underlying LOAD, insufficient data, limited modeling techniques and ineffective low-throughput validation platforms, systems/network approaches to LOAD is still at an early stage. The rapid development

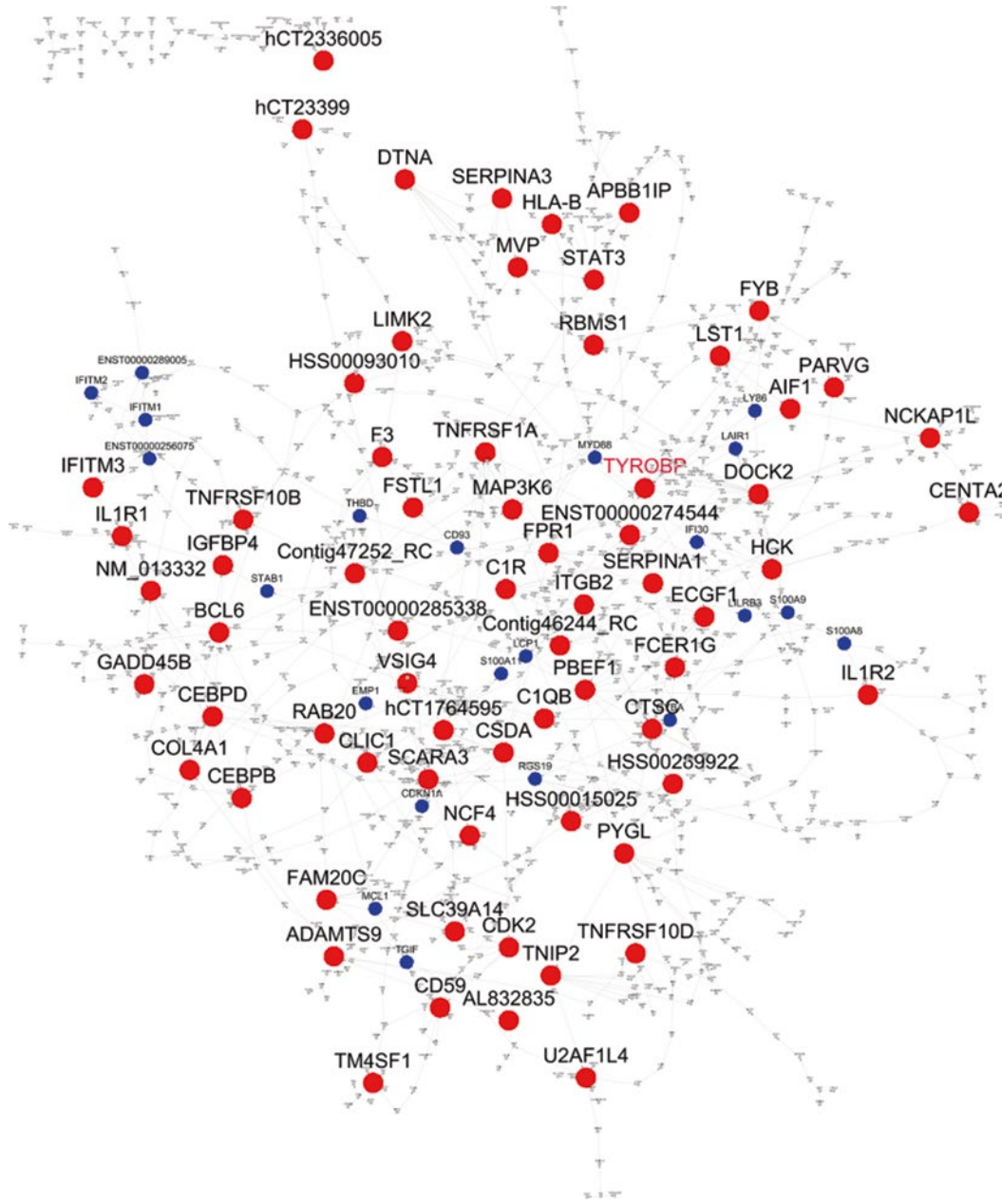


Fig. 4 An immune/microglia enriched Bayesian network and its key drivers. The BN was constructed based on expression data and eSNP information. The key drivers (*red circles*) are identified based on connectivity, as described in the text. The top key drivers of the network are *TYROBP*, *FCER1G*, *CYBA*, *FYB*, *SLC7A7*, *DOCK2*, *TBXAS1*, *SERPINA1*, *FPR1*, *TLR2*, and *SYK* (in the descendant order of connectivity)

of DNA/RNA sequencing [32–34], systems/network biology [35, 36] and in vitro modeling [37–42] technologies will enable building up of a more comprehensive picture of the mechanisms underlying LOAD in the next decade. One would expect to see that LOAD results from a large set of interacting complex networks

at various biological scales, from DNA, molecular, biochemical/cellular, tissue to organism levels. Systems/network biology holds the key to tackling one of the most challenging tasks in biological sciences.

4 Notes

1. Traditional clustering analysis is also effective in identification of co-expressed gene clusters [43].
2. A similarity matrix can be calculated by other similarity measurements such as Euclidean distance or mutual information [44].
3. An adjacency matrix can be derived based on other soft- or hard-thresholding functions [3] like the sigmoid and signum functions:

$$\text{sigmoid}\left(x,\alpha,\tau_0\right)=\frac{1}{1+e^{-\alpha\left(x-\tau_0\right)}} \tag{4}$$

$$\text{signum}\left(x,\tau_0\right)=\begin{cases}1 & \text{if } x\geq\tau_0\\ 0 & \text{if } x<\tau_0\end{cases} \tag{5}$$

However, soft-thresholding functions can identify more biologically meaningful modules.

4. Alternatively, we can define MDC as the difference between the average connectivity among the N genes in the network x to that among the same gene set in network y [45]:

$$\delta_\Omega\left(x,y\right)=\frac{2}{N\left(N-1\right)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(k_{ij}^{x}-k_{ij}^{y}\right) \tag{6}$$

where, *kij* is the connectivity between two genes i and j in a given network.

Obviously, there is a gain of connectivity if $\delta\Omega(x,y)$ is greater than 0 and a loss of connectivity if $\delta\Omega(x,y)$ is smaller than 0. Given M permutations, FDR of MDC is computed as follows:

$$\text{FDR}\left(\delta_\Omega\left(x,y\right)>0\right)=\frac{1}{M}\sum_{p=1}^{M}\delta_\Omega\left(x,y\right)<\delta_\Omega\left(x^p,y^p\right),$$

or

$$\text{FDR}\left(\delta_\Omega\left(x,y\right)<0\right)=\frac{1}{M}\sum_{p=1}^{M}\delta_\Omega\left(x,y\right)>\delta_\Omega\left(x^p,y^p\right),$$

where x^p and y^p are the networks derived from the permuted data.

The significance of this new MDC measure can be accessed by shuffled samples and shuffled gene labels.

5. Random Forest based Causal Networks (RFCN) inference emerges as a powerful approach (the best performer in the DREAM4 In Silico Multifactorial Challenge to predict genetic regulatory networks) though it has not been validated in human [35]. RFCN inference basically decomposes the prediction of a regulatory network between *n* genes into *n* different regression problems. For a given gene *G*, those genes that have significant prediction power in the corresponding regression trees are considered as causal regulators of *G*.

Acknowledgements

This work was supported in part by the National Institutes of Health (NIH)/National Institute on Aging (NIA) Award R01AG046170 (to B.Z. and J.Z.); NIH/National Institute of Mental Health (NIMH) Award R21MH097156-01A1 (to B.Z.); NIH/National Cancer Institute (NCI) Award R01CA163772 (to B.Z. and J.Z..); and NIH/National Institute of Allergy and Infectious Diseases (NIAID) Award U01AI111598-01 (to B.Z. and J.Z.).

References

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3:186–191
2. Lambert JC, Ibrahim-Verbaas CA, Harold D et al (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45: 1452–1458
3. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:17
4. Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 107: 12698–12703
5. Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J Neurosci 28:1410–1420
6. Miller JA, Woltjer RL, Goodenbour JM et al (2013) Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med 5:48
7. Rhinn H, Fujita R, Qiang L et al (2013) Integrative genomics identifies APOE epsilon4 effectors in Alzheimer's disease. Nature 500: 45–50
8. Zhang B, Gaiteri C, Bodea LG et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease. Cell 153:707–720
9. Emilsson V, Thorleifsson G, Zhang B et al (2008) Genetics of gene expression and its effect on disease. Nature 452:423–428
10. Storey JD, Tibshirani R (2003) Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol 224:149–157
11. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics 24:719–720
12. Schadt EE, Sachs A, Friend S (2005) Embracing complexity, inching closer to reality. Sci STKE 2005:pe40
13. Yang X, Deignan JL, Qi H et al (2009) Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. Nat Genet 41:415–423
14. Zhu J, Zhang B, Schadt EE (2008) A systems biology approach to drug discovery. Adv Genet 60:603–635

15. Zhu J, Zhang B, Smith EN et al (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet 40:854–861
16. Zhu J, Sova P, Xu Q et al (2012) Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. PLoS Biol 10:e1001301
17. Schadt EE, Lamb J, Yang X et al (2005) An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 37:710–717
18. Zhu J, Wiener MC, Zhang C et al (2007) Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. PLoS Comput Biol 3:e69
19. Zhu J, Lum PY, Lamb J et al (2004) An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet Genome Res 105: 363–374
20. Chen Y, Zhu J, Lum PY et al (2008) Variations in DNA elucidate molecular networks that cause disease. Nature 452:429–435
21. Schadt EE, Molony C, Chudin E et al (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6:e107
22. Zhu J, Chen Y, Leonardson AS et al (2010) Characterizing dynamic changes in the human blood transcriptional network. PLoS Comput Biol 6:e1000671
23. Tran LM, Zhang B, Zhang Z et al (2011) Inferring causal genomic alterations in breast cancer using gene expression data. BMC Syst Biol 5:121
24. Wang I-M, Zhang B, Yang X et al (2012) Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. Mol Syst Biol 8:594
25. Sharan R, Ulitsky I, Shamir R (2007) Network-based prediction of protein function. Mol Syst Biol 3:88
26. Yang X, Zhang B, Molony C et al (2010) Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. Genome Res 20:1020–1036
27. Kondo T, Takahashi K, Kohara N et al (2002) Heterogeneity of presenile dementia with bone cysts (Nasu-Hakola disease): three genetic forms. Neurology 59:1105–1107
28. Lanier LL, Bakker ABH (2000) The ITAM-bearing transmembrane adaptor DAP12 in lymphoid and myeloid cell function. Immunol Today 21:611–614
29. Paloneva J, Kestilä M, Wu J et al (2000) Loss-of-function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. Nat Genet 25:357–361
30. Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley, New York
31. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23
32. ADNI. http://ida.loni.ucla.edu/
33. BrainSpan. http://www.brainspan.org/
34. GTEx. http://commonfund.nih.gov/GTEx/
35. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS One 5:e12776
36. Marbach D, Costello JC, Küffner R et al (2012) Wisdom of crowds for robust gene network inference. Nat Methods 9:796–804
37. Brennand KJ, Gage FH (2011) Concise review: the promise of human induced pluripotent stem cell-based studies of schizophrenia. Stem Cells 29:1915–1922
38. Brennand KJ, Landek-Salgado MA, Sawa A (2014) Modeling heterogeneous patients with a clinical diagnosis of schizophrenia with induced pluripotent stem cells. Biol Psychiatry 75:936–944
39. Brennand KJ, Simone A, Jou J et al (2011) Modelling schizophrenia using human induced pluripotent stem cells. Nature 473:221–225
40. Qiang L, Fujita R, Abeliovich A (2013) Remodeling neurodegeneration: somatic cell reprogramming-based models of adult neurological disorders. Neuron 78:957–969
41. Qiang L, Fujita R, Yamashita T et al (2011) Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. Cell 146:359–371
42. Tran NN, Ladran IG, Brennand KJ (2013) Modeling schizophrenia using induced pluripotent stem cell-derived and fibroblast-induced neurons. Schizophr Bull 39:4–10
43. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–14868
44. Butte AJ, Kohane IS (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput 5:418–429
45. Gill R, Datta S, Datta S (2010) A statistical framework for differential network analysis from microarray data. BMC Bioinformatics 11:95

Chapter 29

Network-Based Analysis for Uncovering Mechanisms Underlying Alzheimer's Disease

Masataka Kikuchi, Soichi Ogishima, Satoshi Mizuno, Akinori Miyashita, Ryozo Kuwano, Jun Nakaya, and Hiroshi Tanaka

Abstract

Alzheimer's disease (AD) is known to be a multifactorial neurodegenerative disorder, and is one of the main causes of dementia in the elderly. Many studies have demonstrated molecules involved in the pathogenesis of AD, however its underlying mechanisms remain obscure. It may be simplistic to try to explain the disease based on the role of a few genes only. Accumulating new, huge amount of information from e.g. genome, proteome and interactome datasets and new knowledge, we are now able to clarify and characterize diseases essentially as a result of dysfunction of molecular networks. Recent studies have indicated that relevant genes affected in human diseases concentrate in a part of the network, often called as "disease module." In the case of AD, some disease-associated pathways seem different, but some of them are clearly disease-related and coherent. This suggests the existence of a common pathway that negatively drives from healthy state to disease state (i.e., the disease module(s)). Additionally, such disease modules should dynamically change through AD progression. Thus, network-level approaches are indispensable to address unknown mechanisms of AD. In this chapter, we introduce network strategies using gene co-expression and protein interaction networks.

Key words Alzheimer's disease, Systems biology, Network dysfunction, Network perturbation, Disease module, Gene expression profile, Gene co-expression network, Protein interaction network

1 Introduction

Alzheimer's disease (AD) is neuropathologically characterized by extracellular plaques of amyloid-beta (Aβ) peptide and intraneuronal accumulation of neurofibrillary tangles (NFTs). However, the molecular mechanisms of AD pathology remain obscure. What molecules accelerate production of Aβ or NFTs? How do those molecules lead to neuronal cell death? It is difficult to take into account pathological mechanisms of AD by only the known molecules. The new approaches need to identify the remaining essential molecules and pathways.

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_29,

Recently, developments of high-throughput technologies have emerged as a new paradigm for elucidation of biological complexity including complex diseases. Biological molecular interactions (e.g. gene–gene, protein–protein, protein–DNA interactions and others) obtained by comprehensive resources are analyzed using a network representation. In terms of such network biology, there are some attempts to characterize diseases as network perturbations [1–4]. These studies indicate that genes affected in diseases concentrate in a part of the network, often called as "disease module" [5]. AD-associated pathways seem superficially different, but some of them are clearly not incoherent [6, 7]. A common pathway that negatively drives from healthy state to disease state (i.e., the disease module) may exist in AD. Moreover, pursuing the AD-specific modules may help to understand the other neurodegenerative diseases. Actually, a rare mutation in *t*riggering *r*eceptor *e*xpressed on myeloid cells 2 (TREM2) relates not only to AD [8, 9] but also to Nasu-Hakola disease [10] and frontotemporal dementia [11], which suggests that those diseases share common modules or pathways centered on TREM2. The network-level approaches would shed light on the uncharacterized cellular phenomena within AD brains. In this chapter, we present methods for understanding AD pathology through network-based, but not single molecules analysis. To this end, we introduce available gene expression profiles from AD postmortem brains and the human protein–protein interaction datasets in Subheading 2, and in Subheading 3 we provide the actual approaches from recent studies.

2 Materials

A gene co-expression network is generally reconstructed using available gene expression profiles. On the other hand, the protein interaction network is assembled from protein–protein interaction data in open access repositories. We here provide available gene expression profiles of AD and protein–protein interaction databases.

2.1 Gene Expression Profiles of Alzheimer's Disease

Many of systems biology studies have yielded important insights into mechanisms underlying AD using gene expression profiles from postmortem brains and autopsied tissues of AD subjects. Public gene expression datasets are basically registered in the Gene Expression Omnibus (GEO) database [12] (*see* **Note 1**), with GEO accession numbers assigned. Below, we introduce some available gene expression profiles from AD subjects.

The Braak stage is used as the neuropathological staging in AD, which is diagnosed based on expansion of neurofibrillary tangles (NFTs) across brain regions. NFTs deposit in the following order; the transentorhinal region (Braak stage I–II), the limbic

system (Braak stage III–IV) and the isocortical region (Braak stage V–VI) [13]. Liang et al. provided gene expression profiles from postmortem brains of 14 healthy subjects (Braak stage 0–II) and 34AD-affected subjects (Braak stage III–VI) (GEO accession number: GSE5281) [14, 15]. Postmortem brains were laser-captured in six brain regions (entorhinal cortex, hippocampus, medial temporal gyrus, posterior cingulate, superior frontal gyrus and primary visual cortex). The gene expression profiles were obtained with Affymetrix Human Genome U133 Plus 2.0 microarrays (Affymetrix Inc., Santa Clara, CA, USA).

The MiniMental State Examination (MMSE) test is a clinical assessment for cognitive function [16]. Blalock et al. stratified 35 subjects by MMSE score into four groups, "Control" (MMSE score > 25), "Incipient AD" (MMSE score 20–25), "Moderate AD" (MMSE score 14–19), and "Severe AD" (MMSE score < 14) (GSE1297) [17]. The CA1 and CA3 regions were dissected from the frozen hippocampal tissues and they were profiled on Affymetrix Human Genome U133A Array.

In order to construct gene regulatory network in late-onset alzheimer's disease (LOAD) and non-demented healthy controls, Zhang et al. collected 690 autopsied tissues from dorsolateral prefrontal cortex BA9, visual cortex BA17 and cerebellum in brains of LOAD patients, and utilized custom microarrays manufactured by Agilent Technologies (GSE44772) [18].

2.2 The Human Protein–Protein Interaction Data

Each interaction between proteins has been identified by established methods (i.e. two-hybrid system, immunoprecipitation method, and others). Over the past decade, high-throughput technologies including large-scale yeast two-hybrid screenings and mass spectrometry have enabled to obtain comprehensive protein–protein interaction (PPI) datasets in human [19–21]. At present, PPI datasets curated from published studies regardless of small- or large-scale experiments are integrated in databases as those in Table 1.

3 Methods

3.1 Construction of the Gene Co-expression Network

In order to build the gene co-expression network, associations between genes are determined by Pearson correlation coefficient (PCC) using gene expressions across samples. PCC ranges from −1 (negative correlation) to 1 (positive correlation). When PCC between gene i and gene j ($PCC_{i,j}$) exceeds a threshold, two genes are linked (co-expression). The PCC value (e.g., $|PCC| > 0.5$) and p-value can be used directly as test for no correlation. However, these thresholds depend on sample size and are often arbitrary. To overcome these difficulties, the weighted gene co-expression network analysis (WGCNA) (*see* **Note 2**) [22, 23], which is widely applied in some studies including AD [18, 24, 25], determines a

Table 1
Protein interaction databases

Database name	URL
The Biological General Repository for Interaction Datasets (BioGRID)	http://thebiogrid.org/
The Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
The Human Protein Reference Database (HPRD)	http://www.hprd.org/
The IntAct	http://www.ebi.ac.uk/intact/
The Interologous Interaction Database (I2D)	http://ophid.utoronto.ca/ophidv2.204/
IRefIndex	http://irefindex.org/wiki/index.php?title=iRefIndex
The Molecular INTeraction database (MINT)	http://mint.bio.uniroma2.it/mint/Welcome.do
STRING	http://string-db.org/

threshold based on the fact that biological networks are essentially scale-free (*see* **Note 3**). First, $PCC_{i,j}$ is transformed into "similarity," $s_{i,j}$, taking from 0 to 1 (*see* **Note 4**):

$$s_{i,j} = \left| PCC_{i,j} \right|$$

If you preserved the sign of $PCC_{i,j}$,

$$s_{i,j} = \frac{1 + PCC_{i,j}}{2}$$

is used (*see* **Note 5**). Next, $s_{i,j}$ is assigned into the power function:

$$a_{i,j} = \left| s_{i,j} \right|^{\beta}$$

where β is the parameter. The parameter β should be set to be higher than the scale-free topology model fit (R^2) that is the slope between $\log_{10}(p(k))$ and $\log_{10}(k)$ (*see* **Note 3** about $p(k)$). A stringent parameter brings the higher R^2, but it may lead to networks with very few interactions because of trade-off relationships between R^2 and the number of interactions.

3.2 Construction of the Protein Interaction Network

Some studies analyze PPI datasets that combine data from several databases and repositories, however the curation policies of each database are different. In addition, registered proteins are often maintained with different identifiers (e.g. Entrez gene ID and UniProt ID). The International Molecular Exchange (IMEx) consortium recently developed common strategies and attempts to

provide a nonredundant dataset through the participating databases [26]. To avoid problems in some efforts at unifying IDs for example, iRefIndex provides an index across 13 primary databases [27].

3.3 Module Detection

The enormous amount of information in biological networks makes it difficult to be analyzed. Therefore, networks are usually divided into modules, which are defined as subsets of nodes (genes or proteins) that densely interact with each other (represented as links or edges).

There are mainly two methods for module detection. Basically, either a node belongs to only one module or to multiple modules. We here present the Infomap algorithm and the topological overlap as the first method, and the link clustering algorithm as the second method.

The Infomap algorithm that proposed by Rosvall and Bergstrom, which detects modules based on the random walk [28]. The algorithm divides a network into m modules with an optimal number of modules, M. Here, the module is defined as the region in which the random walker tends to stay for a long time. The efficiency on M is assessed by the map equation [29]:

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p_{\circlearrowright}^{i} H\left(p^{i}\right)$$

where $q_{\curvearrowright}$ and $H(Q)$ are the probability and the entropy of the movement of the random walker between modules, $p_{\circlearrowright} i$ and $H\left(P^{i}\right)$ are the fraction and the entropy of the movement within module i. This equation takes/results in a low value when a random walker has less module transitions and less within-module movements. It seeks the best number of modules to minimize the map equation over all possible partitions. The Infomap algorithm is reported to have the best-performance compared to several algorithms [30].

Next, the topological overlap is the method focused on a link similarity between node i and j ($\omega_{i,j}$), which is given by the formula below:

$$\omega_{i,j} = \frac{l_{i,j} + a_{i,j}}{\min\left(k_i, k_j\right) + 1 - a_{i,j}}$$

where $l_{i,j}$ is the number of common nodes connected between node i and j, $a_{i,j}$ is adjacency function ($a_{i,j} = 1$, if i and j are linked. and $a_{i,j} = 0$, otherwise), and k_i is the connection degree (the number of interacting partners) of i [31]. The topological overlap calculated across all nodes is displayed as a matrix. Hierarchical clustering is implemented to its matrix, and the classified clusters are regarded as modules. This method can be applied to unweighted

and weighted networks. In practice, the weighted gene co-expression network analysis (WGCNA) detects modules by the topological overlap matrix (TOM) from a constructed gene co-expression network.

The two methods above essentially assign a node to a module, whereas in real networks a node could participate/belong to multiple modules. For instance, proteins that have a lot of functions may associate with several protein complexes in the biological network. In particular, such proteins are called as "date hubs" in the context of systems biology [32]. The link clustering is a method to classify links into distinct modules [33]. The originality of this method is to calculate similarity between links e_{ik} and e_{jk} that share a node k as:

$$S\left(e_{ik},e_{jk}\right)=\frac{\left|n_{+}\left(i\right)\cap n_{+}\left(j\right)\right|}{\left|n_{+}\left(i\right)\cup n_{+}\left(j\right)\right|},$$

where $n_{+}\left(i\right)$ is the node set of node i and the neighbors. Calculated similarities are reordered by the application of hierarchical clustering and the results are represented as a dendrogram. In order to determine the best threshold to cut branches in a dendrogram, the partition density, D, is used:

$$D=\frac{2}{M}\sum_{c}m_{c}\frac{m_{c}-\left(n_{c}-1\right)}{\left(n_{c}-2\right)\left(n_{c}-1\right)}$$

where M is the number of links in the network, c is the number of the modules, m_c is the number of links in a module and n_c is the number of the nodes in a module. The partition density, D, indicates the average density across each module and takes the value from 0 (sparse) to 1 (dense). D is computed at each height of the dendrogram. The height at which D takes the maximum value is adopted as the cutting threshold.

3.4 Application Using a Genetic Interaction Network

Zhang and coworkers analyzed the gene expression profiles of three brain regions (dorsolateral prefrontal cortex (PFC), visual cortex (VC) and cerebellum (CB)) from LOAD and non-demented individuals described in Subheading 2.1 (GSE44772) [22]. They first obtained 13,193 (one-third) of the most variable probesets in each brain. The probesets were assigned a unique identifier, combined probeset ID and brain region name, and those expression data were merged. Based on these multi-tissue expression data sets containing each 39,579 probesets in LOAD and non-demented brains, multi-tissue co-expression networks were constructed by WGCNA. From the topological overlap matrices (*see* Subheading 3.3), 111 and 89 modules were identified in LOAD and non-demented brains, respectively. Next, they measured the modular differential connectivity

(MDC) to compare the connectivity among modules in LOAD and normal healthy brains. MDC is defined by the following:

$$\delta_{\Omega}\left(\text{LOAD},\ \text{Normal}\right)=\frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}k_{ij}^{\text{LOAD}}}{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}k_{ij}^{\text{Normal}}}$$

where N is the number of genes in a module, k_{ij} is the connectivity between genes i and j. Here, k_{ij} equals to $a_{i,j}$ in Subheading 3.1. The modules with MDC > 1 indicate gain of connectivity (GOC), in contrast, those with MDC <1 indicate loss of connectivity (LOC). The GOC modules were found more than ten times greater than the LOC modules. In GOC modules with at least 100 genes, the immune/microglia module was identified, and 99.5 % of genes in this module were differentially expressed in PFC, which is commonly affected in AD. Interestingly, expressions of genes in the PFC immune/microglia module correlated with atrophy levels in several brain regions. Furthermore, expression quantitative trait loci (eQTL) analyses were performed to identify SNPs associated with gene expressions (eSNPs). Many genes in the PFC immune/microglia module were significantly enriched *cis*-eSNPs located within around 1 Mb of the gene body. Finally, the directed Bayesian networks for the immune/microglia module were constructed. As a result of calculation of the combined score, based on the number of downstream genes and differential expression, TYRO protein tyrosine kinase-binding protein (TYROBP) was ranked the highest score, indicating TYROBP is a key causal regulator. TYROBP is also known as DNAX-activating protein of 12 kD (DAP12), and works as a signaling adaptor protein of TREM2. A rare variant of TREM2 was recently reported increases the risk to develop LOAD in cohorts from North America and Europe [6, 7].

3.5 Application Using a Protein Interaction Network

The biggest risk for AD is aging. AD progresses slowly over years or decades, rather than a rapid transition from healthy to disease state. We therefore have to consider dynamic, temporal changes of the AD-associated networks and modules.

We recently identified modules disrupted with the progression of AD by combining a protein interaction network with gene expression profiles of brains from AD and normal aging individuals [34]. The AD gene expression profiles used were from postmortem brains of AD subjects (GSE5281), introduced in Subheading 2.1. We also used the gene expression profiles from postmortem brains (entorhinal cortex, hippocampus, superior frontal gyrus and postcentral gyrus) of cognitively intact subjects aged 60–99 years as normal aging [35]. Normal aging subjects were classified into the following four age groups: 60–69, 70–79,

80–89, and 90–99 years old. We analysed gene expression profiles from common three brain regions (entorhinal cortex (EC), hippocampus (HIP) and superior frontal gyrus (SFG)) between two datasets. First, gene expression datasets were normalized using the MAS 5.0 algorithm (Affymetrix, Santa Clara, CA). Then, we used probe sets marked as "present" by the detection call algorithm (Affymetrix) and averaged their expression values through samples in same brain region in same stage (Braak stage or age group). Here, we considered that a gene is expressed if the average expression values exceeded 200, and assumed direct protein expression from gene expression (RNA expression) datasets (*see* **Note 6**). We next retrieved the human interaction dataset from BioGRID [36]. Adding physical interactions between expressed proteins, the expressed protein interaction networks (PINs) were constructed in each stage, and they were divided into modules using the Infomap algorithm (*see* Subheading 3.3). To observe trajectories of modules through AD progression (Braak stages), we performed the brute-force approach to compute similarities of interactions (C_L) and cellular functions of proteins (C_{GO}) between modules in a stage and the next stage. The similarity was defined as follows:

$$C(t) = \frac{\left|A(t) \cap A(t+1)\right|}{\left|A(t) \cup A(t+1)\right|}$$

where $A(t)$ is a set of interactions to obtain similarities of interactions (C_L), or cellular functions, to obtain similarities of cellular functions (C_{GO}), in a module at time t (i.e., Braak stage or age group) (Fig. 1). A similarity takes 1 when the modules in time t and $t+1$ have same interactions or same cellular functions (*see* **Note 7**). To estimate whether two modules in time t and $t+1$ are conserved, we considered that the both modules were conserved if a module pair has the highest C_L and their C_L and C_{GO} exceed 0.5 (*see* **Note 8**). Otherwise, they are not conserved. Repeating this procedure, the conserved relationships between modules in consecutive stages were linked as a module lineage. Next, we sought AD-specific disrupted module lineages, which are defined as module lineages that are fully conserved across all age groups in normal aging but are not conserved across Braak stages in AD. AD-specific, disrupted module lineages are classified into the early-disrupted type and the late-disrupted type. In entorhinal cortex (EC), affected in the incipient Braak stage, 4.0 % of all module lineages indicated early-disrupted type, and 40.0 % of all module lineages indicated late-disrupted type (*see* **Note 9**). Of the late-disrupted type in EC, we found a module that lost the most interactions across Braak stages. The members in the module are

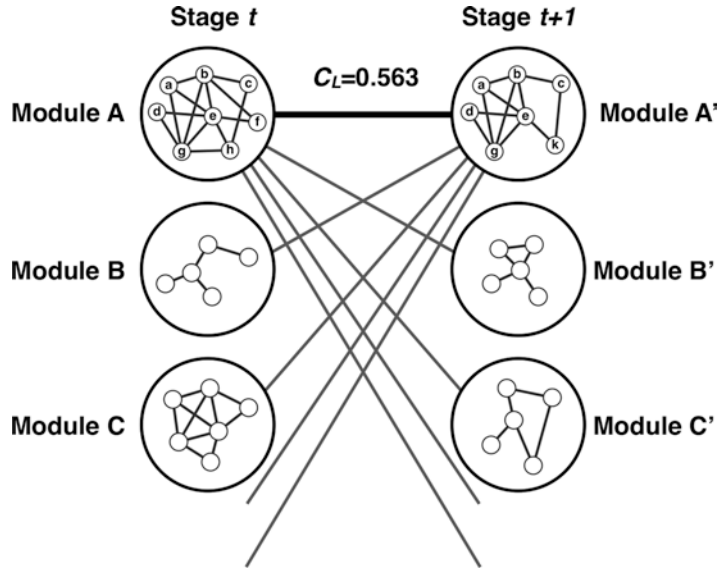


Fig. 1 Calculation for similarities between modules stage t and $t+1$. Similarities of interactions (C_L) and cellular functions (C_{GO}) are calculated over all possible module pairs between stage t and $t+1$. We considered that the both modules were conserved if a module pair has the highest C_L and their C_L and C_{GO} exceed 0.5

significantly associated with the histone acetyltransferase (HAT) complex. We also found that the HAT module tightly interacted with the proteasome module via the deubiquitinating enzyme UCHL5 in Braak stage I (Fig. 2). However, interactions between UCHL5 and some members in the HAT module (INO80B/C, NFRKB and others) were beginning to disappear in Braak stage II, and fully collapsed in Braak stage IV. UCHL5 has been reported to interact with the INO80 complex via NFRKB [37]. This complex could alter chromatin conformation and regulate gene transcription or DNA repair [38]. Furthermore, the deubiquitinating enzyme UCHL5 is also associated with the 26S proteasome. In healthy cells, abnormal toxic proteins (e.g., Aβ in AD) are decomposed by protein quality control systems such as the ubiquitin-proteasome system (UPS). However, the degradation of toxic proteins does not seem to work efficiently in AD compared to healthy subject. Indeed, an impairment in ubiquitin-proteasome system function has recently been observed in AD [39, 40]. Our findings suggest that down-regulated UCHL5 and affected network interactions may disturb proteolysis, with also presence of aberrant gene expression in AD.

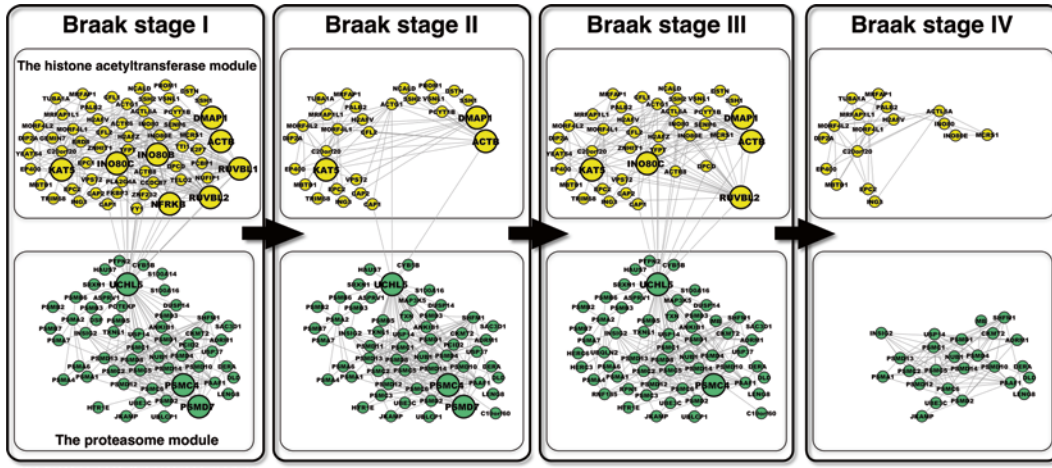


Fig. 2 Dynamics of module interactions in the entorhinal cortex during AD progression. The *upper yellow* and *lower green* nodes are components of the histone acetyltransferase and proteasome modules respectively. Hub proteins disappearing with Braak stage are depicted as large nodes. Figure obtained, adapted from studies/data in [34]

4 Notes

1. The Gene Expression Omnibus (GEO) is provided at the National Center for Biotechnology Information (NCBI), and is freely accessible at (http://www.ncbi.nlm.nih.gov/geo/).
2. WGCNA is implemented based on the R project for statistical computing package (http://www.r-project.org).
3. A network is composed of nodes (e.g., genes or proteins) and edges/links (e.g., co-expression relationships or physical interactions). In a scale-free network, the frequency of connection degree (number of partners a node interacts with) is $p(k) \sim k^{-\gamma}$, where k is the connection degree and γ is the degree exponent. This indicates the presence of many nodes with a few interactions and a few nodes with many interactions. Many biological networks are scale-free [41]. In WGCNA, the users can determine the parameter β to conserve scale-free topology.
4. Besides Pearson correlation coefficient, the other measurements (e.g. biweight midcorrelation, mutual information) are calculable.
5. The users can select "unsigned" or "signed" from the variables in corresponding functions ("type" and "networkType").
6. To determine whether the gene is expressed or not, we adopted a 200 threshold based on the method proposed by Bossi et al. [42].

An expression value of 200 represents approximately 3–5 copies per cell [43].

7. Based on the "biological process" functions of the Gene Ontology Annotation (GOA), we assigned proteins with cellular functions. Note that one protein can have several functions. Next, we assigned an interaction with the GOA common to both proteins constituting the interaction. Using interaction sets with the GOA functions, we next sought significantly enriched functions within each module by hypergeometric test. If the probability by hypergeometric distribution was less than 0.05 and the ratio to expected value was greater than 2, we assigned the GOA function to the module. As an example of calculation of C_{GO}, we consider a module with functions A, B and C at time t (M_t^1), and a module with functions B and D at time $t+1$ $\left(M_{t+1}^1\right)$. The common function is B, and the union of functions is A, B, C and D, therefore the C_{GO} is $1/4$.
8. This criterion has two steps: (1) filtering module pairs with the highest C_L and, (2) extracting module pairs with C_L and $C_{GO} > 0.5$ from module pairs filtered in step (1). In the first step, if the modules at time t and $t+1$ are conserved, each interaction that constitute the two modules has to be highly shared. For instance, when a module at time t M_t^1 shows the highest C_L with a module at time $t+1$ M_{t+1}^1 and M_{t+1}^1 also shows the highest C_L with M_t^1, $M_t^1 - M_{t+1}^1$ pair moves to the next step. On the other hand, if M_{t+1}^1 shows the highest C_L with a different module at time t M_t^2, $M_t^1 - M_{t+1}^1$ pair is omitted from this criterion. Note that the highest C_L can be same value (e.g., when M_t^1 equally splits into M_{t+1}^1 and M_{t+1}^2 at time $t+1$). The second step is a process to filter out pairs with same highest C_L and lowest conserved pairs. Summation of C_L of a module is ≤ 1. From this, it follows that with a threshold >0.5, the pair satisfying this threshold is determined uniquely. Conversely, summation of C_{GO} of a module can be >1 because cellular functions can be redundant. The threshold of C_{GO} is therefore arbitrary.
9. We did not verify the statistical significance of the disrupted modules in [34]. To do this here, we propose bootstrap analysis as a useful approach. More specifically, we randomly resample protein sets (e.g., 1,000) with the same number of proteins as the observed module from expressed proteins (i.e. "resampling set" and "observed set"). We compare statistics (e.g., number of interactions lost across Braak stages) between the observed set and the resampling sets. If the statistics of the observed set are significantly different with those of the resampling sets, we evaluate that the observed module is a disrupted module.

Acknowledgement

We thank Drs. Takeshi Ikeuchi and Kensaku Kasuga of Niigata University for helpful discussions.

References

1. Lewis NE, Schramm G, Bordbar A et al (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nat Biotechnol 28:1279–1285
2. Mine KL, Shulzhenko N, Yambartsev A et al (2013) Gene network reconstruction reveals cell cycle and antiviral genes as major drivers of cervical cancer. Nat Commun 4:1806
3. Pichlmair A, Kandasamy K, Alvisi G et al (2012) Viral immune modulators perturb the human molecular network by common and unique strategies. Nature 487:486–490
4. Rozenblatt-Rosen O, Deo RC, Padi M et al (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature 487:491–495
5. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12:56–68
6. Mizuno S, Iijima R, Ogishima S et al (2012) AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst Biol 6:52
7. Huang Y, Mucke L (2012) Alzheimer mechanisms and therapeutic strategies. Cell 148:1204–1222
8. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116
9. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127
10. Paloneva J, Manninen T, Christman G et al (2002) Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. Am J Hum Genet 71:656–662
11. Guerreiro RJ, Lohmann E, Bras JM et al (2013) Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. JAMA Neurol 70:78–84
12. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210
13. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82:239–259
14. Liang WS, Dunckley T, Beach TG et al (2007) Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. Physiol Genomics 28:311–322
15. Liang WS, Dunckley T, Beach TG et al (2008) Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol Genomics 33:240–256
16. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198
17. Blalock EM, Geddes JW, Chen KC et al (2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci U S A 101:2173–2178
18. Zhang B, Gaiteri C, Bodea LG et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease. Cell 153:707–720
19. Rual JF, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437:1173–1178
20. Stelzl U, Worm U, Lalowski M et al (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122:957–968
21. Ewing RM, Chu P, Elisma F et al (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol 3:89
22. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:17
23. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559
24. Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J Neurosci 28:1410–1420

25. Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A 107: 12698–12703
26. Orchard S, Kerrien S, Abbani S et al (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods 9:345–350
27. Razick S, Magklaras G, Donaldson IM (2008) iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics 9:405
28. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 105:1118–1123
29. Rosvall M, Axelsson D, Bergstrom CT (2008) The map equation. Eur Phys J Spec Top 178: 13–23
30. Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. Phys Rev E 80:056117
31. Ravasz E, Somera AL, Mongru DA et al (2002) Hierarchical organization of modularity in metabolic networks. Science 297:1551–1555
32. Han JD, Bertin N, Hao T et al (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430:88–93
33. Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466:761–764
34. Kikuchi M, Ogishima S, Miyamoto T et al (2013) Identification of unstable network modules reveals disease modules associated with the progression of Alzheimer's disease. PLoS One 8:e76162
35. Berchtold NC, Cribbs DH, Coleman PD et al (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. Proc Natl Acad Sci U S A 105:15605–15610
36. Stark C, Breitkreutz BJ, Chatr-Aryamontri A et al (2011) The BioGRID interaction database: 2011 update. Nucleic Acids Res 39: D698–D704
37. Yao T, Song L, Jin J et al (2008) Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. Mol Cell 31:909–917
38. Zediak VP, Berger SL (2008) Hit and run: transient deubiquitylase activity in a chromatin-remodeling complex. Mol Cell 31:773–774
39. Keller JN, Hanni KB, Markesbery WR (2000) Impaired proteasome function in Alzheimer's disease. J Neurochem 75:436–439
40. Lam YA, Pickart CM, Alban A et al (2000) Inhibition of the ubiquitin-proteasome system in Alzheimer's disease. Proc Natl Acad Sci U S A 97:9902–9906
41. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113
42. Bossi A, Lehner B (2009) Tissue specificity and the human protein interaction network. Mol Syst Biol 5:260
43. Su AI, Cooke MP, Ching KA et al (2002) Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 99:4465–4470

Chapter 30

The SDREM Method for Reconstructing Signaling and Regulatory Response Networks: Applications for Studying Disease Progression

Anthony Gitter and Ziv Bar-Joseph

Abstract

The Signaling and Dynamic Regulatory Events Miner (SDREM) is a powerful computational approach for identifying which signaling pathways and transcription factors control the temporal cellular response to a stimulus. SDREM builds end-to-end response models by combining condition-independent protein–protein interactions and transcription factor binding data with two types of condition-specific data: source proteins that detect the stimulus and changes in gene expression over time. Here we describe how to apply SDREM to study human diseases, using epidermal growth factor (EGF) response impacting neurogenesis and Alzheimer's disease as an example.

Key words Pathway discovery, Transcriptional regulation, Time series gene expression, Protein–protein interaction network, Knockdown prediction, Network dynamics, EGF response, Alzheimer's disease, Human disease progression

1 Introduction

Studies of the cellular response to external stimuli or stresses provide valuable information regarding the coordinated activation of genes and their protein products. Similarly, studying how biological processes are disrupted in human diseases provides insights into their underlying organization and can lead to the identification of the underlying causes of those diseases. Because most biological processes are inherently dynamic, it is important to monitor the temporal component of a response by collecting time series data [1]. Although some types of high-throughput data are often collected as a time series (e.g. gene expression profiles), others are either static (for example DNA) or only measured at a single time point due to technological and other challenges (e.g. protein interactions). Gene expression alone, while very informative for identifying the effects of a stimulus, does not directly reveal the upstream mechanisms that

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_30,

drive transcriptional changes. To address this challenge, and to construct a mechanistic model for the activation of genes during stress response and disease progression, we developed the SDREM algorithm [2]. SDREM is able to recover upstream proteins and pathways—both transcription factors (TFs) and the signaling cascades that active the TFs—by combining condition-specific temporal gene expression with biological networks.

SDREM is divided into two phases: the first detects TFs that are likely to be controlling changes in gene expression, and the second searches for signaling pathways that can activate these TFs (Fig. 1). In the first phase, putative TFs are identified by combining the TF-gene binding interactions with temporal expression data [3, 4]. When genes are initially similarly expressed and then split into sub-groups, the divergence is likely caused by the influence of active TFs that bind the genes in one sub-group but not the other(s). Our method identifies such splits and assigns TFs to them based on the subset of genes they are known or predicted to regulate. This analysis provides a set of TFs that are potentially active in the response, but in order to be activated the TFs themselves must receive signals from the upstream proteins that initiate the response. Therefore, SDREM's second phase discovers pathways in a protein–protein interaction (PPI) network that connect the upstream source proteins (either sensory proteins or other proteins that interact with the environment, stimulus, or pathogen) to the downstream putative target TFs using a network orientation algorithm [5]. Prior information, for example from RNA interference (RNAi) screens, can be used to suggest which proteins should be included in the signaling pathways [6]. The output of the second phase feeds back into the first, and the entire process is repeated until SDREM converges upon TFs that are well-connected in the signaling pathways and modulate the differentially expressed genes. Given an SDREM model, it is possible to prioritize experimental validation by predicting the effects of single or double gene knockdowns [6].

We first applied SDREM to study the high osmolarity glycerol and target of rapamycin pathways in yeast as well as pathogen response in *Arabidopsis thaliana* [2]. Subsequent extensions enabled SDREM to be applied to the study of human diseases. In our SDREM analysis of H1N1 influenza infection, we recovered signaling pathways and TFs that correspond well to canonical representations of immune response and previous literature [6]. Furthermore, we showed that SDREM can accurately predict the effects of RNAi screens and used it to estimate the effects of double gene knockdowns on viral load. We identified several proteins that are likely to be involved in the response to pathogenic H5N1 influenza but not seasonal H1N1 influenza.

Here we demonstrate SDREM's potential for Alzheimer's disease research with a case study of epidermal growth factor (EGF) stimulus.

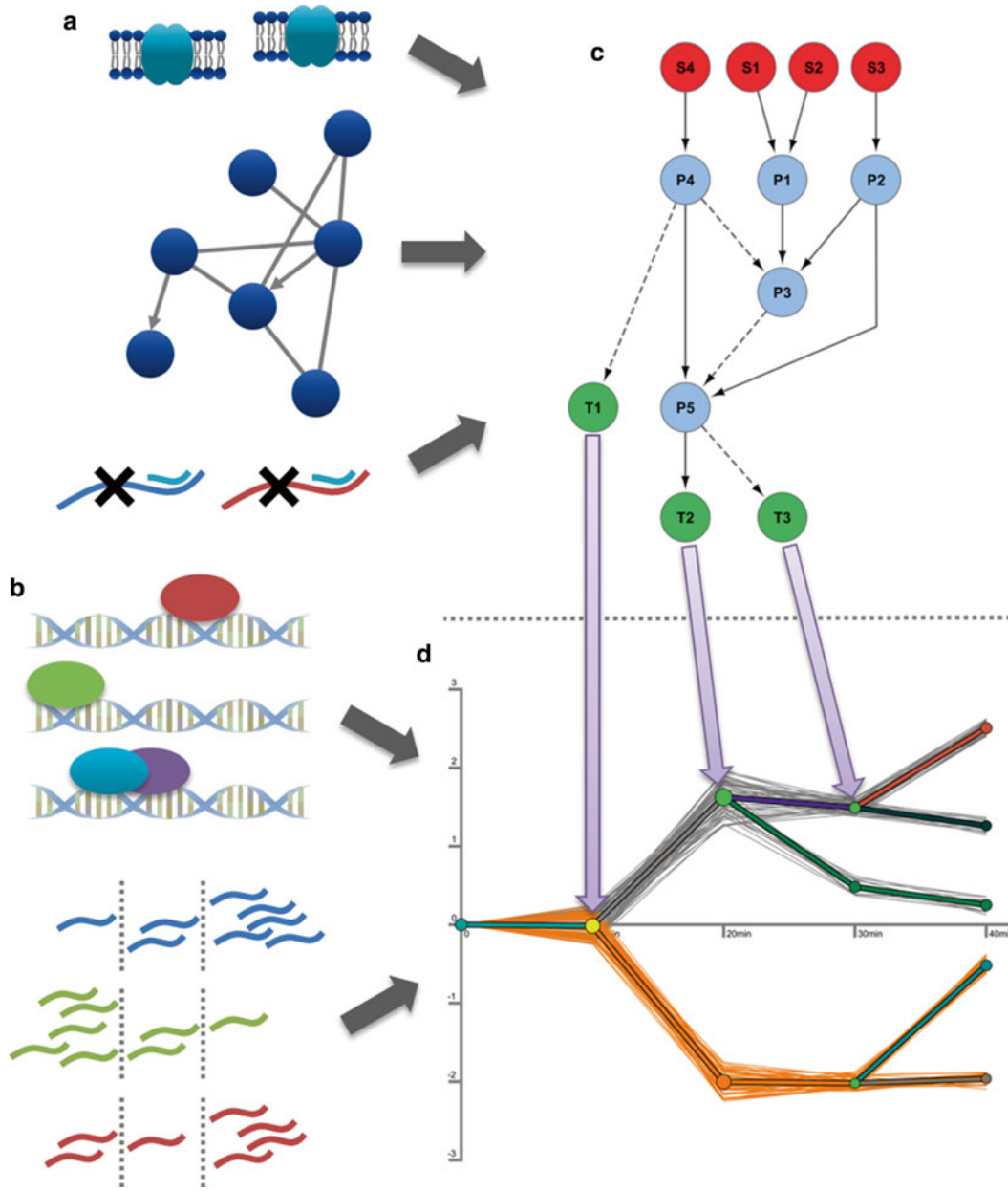


Fig. 1 SDREM input and output. (**a**) Three types of data are provided as input to the signaling pathway inference component of SDREM. The list of source proteins includes receptors or other proteins that directly interact with the activation agent used in the experiment to initiate the signaling response (for example, a pathogen or an overexpressed protein). The PPI network is a list of physical interactions between proteins. Node priors are optional and specify the prior probability that a protein is a member of the signaling pathway. They may be derived from RNAi screens (as pictured), reference pathways, or other data. (**b**) SDREM integrates static protein–DNA interactions and temporal gene expression (for example, RNA-Seq data measured at multiple time points as pictured) to determine when TFs are actively controlling the transcriptional component of the response. (**c**) The input in panel (**a**) and the predicted active TFs from SDREM are used to learn signaling pathways from the sources to the TFs. *See* Fig. 3 for additional details. (**d**) The input from panel (**b**) and the connectivity of the TFs in SDREM's predicted signaling pathways are used to refine the estimate of which TFs are active in the stress response and their times of activity. Full color figure appears on the SDREM website: http://www.sb.cs.cmu.edu/sdrem. *See* Fig. 2 for additional details

EGF has been shown to impact neurogenesis [7], and it has been speculated that therapies to increase neurogenesis could benefit patients with Alzheimer's disease [8]. In addition, presenilin 1 (*PS1*), which can cause early onset familial Alzheimer's disease when mutated, is an important regulator of the EGF receptor (*EGFR*) [9]. As can be seen in Figs. 2 and 3, when applied to EGF stimulus data, SDREM is able to correctly recover a large fraction of known EGFR pathway members including core signaling cascade members *SHC1*, *GRB2*, *SOS1*, *HRAS*, *MAP2K1*, and *MAPK1* as well as the TFs *ELK1*, *FOS*, and *JUN*. The EGF study described in this chapter serves as an example of how SDREM can be applied to study other stimuli applied to brain cells and tissues, and disease progression. All data referenced in Methods are provided with the SDREM software.

2 Materials

2.1 Software

1. If Java is not already installed, download and install it from http://www.java.com/. SDREM requires Java 5 or higher.
2. Download SDREM from http://sb.cs.cmu.edu/sdrem/ (*see* **Note 1**). No installation is necessary. SDREM version 1.2 is described here.
3. Download Cytoscape [10, 11] from http://www.cytoscape.org/. Cytoscape version 2.8 is described here (*see* **Note 2**).

2.2 Required Input Data

1. Time series gene expression data (*see* **Note 3**).
2. A list of the source proteins in the protein–protein interaction network with one protein per line.

2.3 Optional Input Data (Defaults Available)

1. TF-gene binding interactions and a list of all potential TFs (*see* **Notes 4** and **5**). SDREM requires these inputs, but if user-provided data are not available, default interactions are provided with the SDREM software (*see* **Note 6**).
2. Protein–protein interaction network (*see* **Note 7**). A default network is provided (*see* **Note 8**).
3. Node priors for the proteins in the PPI network (*see* **Note 9**).

3 Methods

1. Prepare the properties file (e.g. store.props) for the preprocessing step that enumerates paths in the PPI network and stores them to disk.
2. Store paths using the command 'java -Xmx16g -jar StorePaths.jar store.props'. -Xmx is a Java option that sets the maximum heap size (to 16 gigabytes in this example), which can be increased

or decreased depending on the size of the input data, the amount of RAM available on the machine, and the operating system (*see* **Notes 10** and **11**).

3. Prepare the two SDREM properties files. One specifies DREM parameters (e.g. drem_settings.txt) and the other contains all other SDREM parameters (e.g. sdrem.props) as well as a reference to the DREM properties file (*see* **Note 12**).
4. Run SDREM using the command 'java -Xmx16g -jar sdrem.jar sdrem.props'. As above, the maximum heap size can be changed as needed (*see* **Notes 13** and **14**).
5. After SDREM terminates, use the DREM graphical user interface to view the temporal TF activity and groups of co-regulated genes (Fig. 2). Execute the command 'java -Xmx1g -jar drem.jar -d drem_settings.txt'. Most of the DREM options will be automatically populated from the properties file. Set the TF-gene Interactions File field to tfActivityPriors_round<N-1>.txt and the Saved Model File field to <N>.model, which SDREM writes in the model directory specified by the 'model.dir' parameter in sdrem.props. In this and subsequent steps, <N> is the value of the 'iterations' parameter in sdrem.props. For example, if iterations = 10, then load the files tfActivityPriors_round9.txt and 10.model. Click the Execute button to load and view the saved model (*see* **Notes 15** and **16**).
6. Use Cytoscape to view the SDREM signaling pathways (Fig. 3). Open Cytoscape and load the pathways by selecting 'File → Import → Network (Multiple File Types)…' and choosing topPathEdges_itr<N>.sif (*see* **Note 17**). Load the role the predicted proteins play in the signaling pathways (source, internal signaling protein, or target TF) by selecting 'File → Import → Node Attributes…' and selecting topPathNodes_itr<N>.noa. Arrange the nodes by selecting a layout in the Layout menu. Visually distinguish the different types of nodes on the pathways by switching to the VizMapper™ tab in the Control Panel. Scroll to Node Color in the Visual Mapping Browser and double click it. Select Role from the Node Color attribute menu that appears and choose the Discrete Mapper. Then select a different color for each of the three types of nodes.
7. If the SDREM 'predict.knockdown' parameter was set and gene knockdown effects were predicted, then analyze the top-ranked genes. Single knockdown predictions are written to the file singleKnockdown_itr<N>.txt. Open this tab-delimited text file using spreadsheet software such as Microsoft Excel and sort the data by the topTargAvgRank column in ascending order. The corresponding values in the topTargAvg column are in the range [0, 1] and specify the fraction of TF connectivity that remains after the gene is removed from the top-ranked paths, which means that smaller values signify stronger predicted effects.

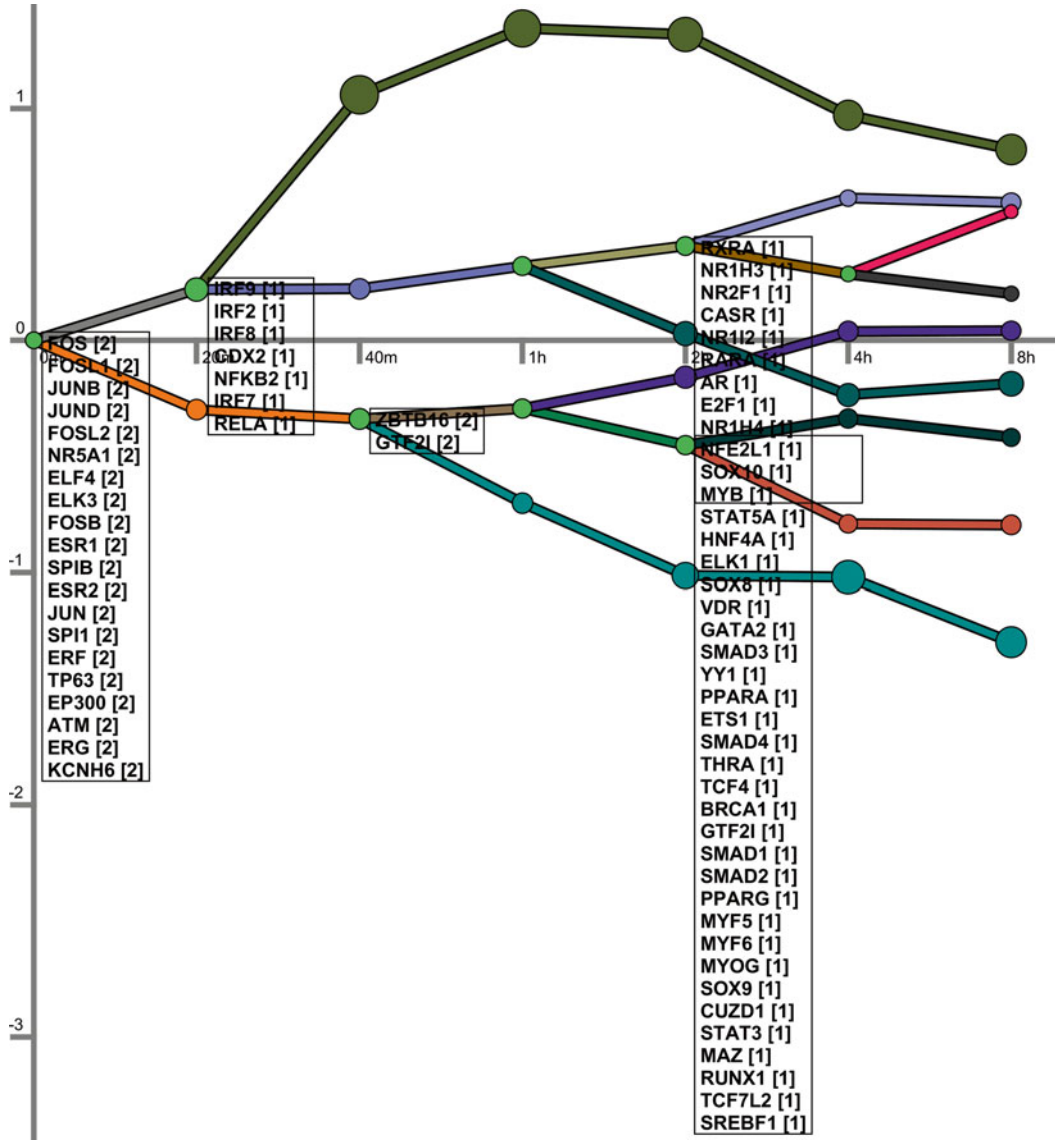


Fig. 2 SDREM active TFs and co-regulated genes in the EGF response. Each colored path summarizes the transcriptional changes of a group of genes that is co-expressed and co-regulated over time. The *x*-axis is time and the *y*-axis is $\log_2$ fold change relative to the 0 min time point. TF names are shown next to the time point on a regulatory path at which SDREM predicts that the TF actively up- or down-regulates its target genes on the path. TF names are only shown the first time the TF is active on a path. The numbers in brackets indicate which branch out of the path the TF controls after it diverges. One group of active TF labels at the 2 h time point has been hidden because it overlaps with another group of labels. Full color figure appears on the SDREM website: http://www.sb.cs.cmu.edu/sdrem

If double knockdown effects were predicted as well, open the tab-delimited file doubleKnockdown_itr<N>.txt and sort by the topTargAvgRank in ascending order. The values in the topTargAvg_int column are the predicted genetic interactions.

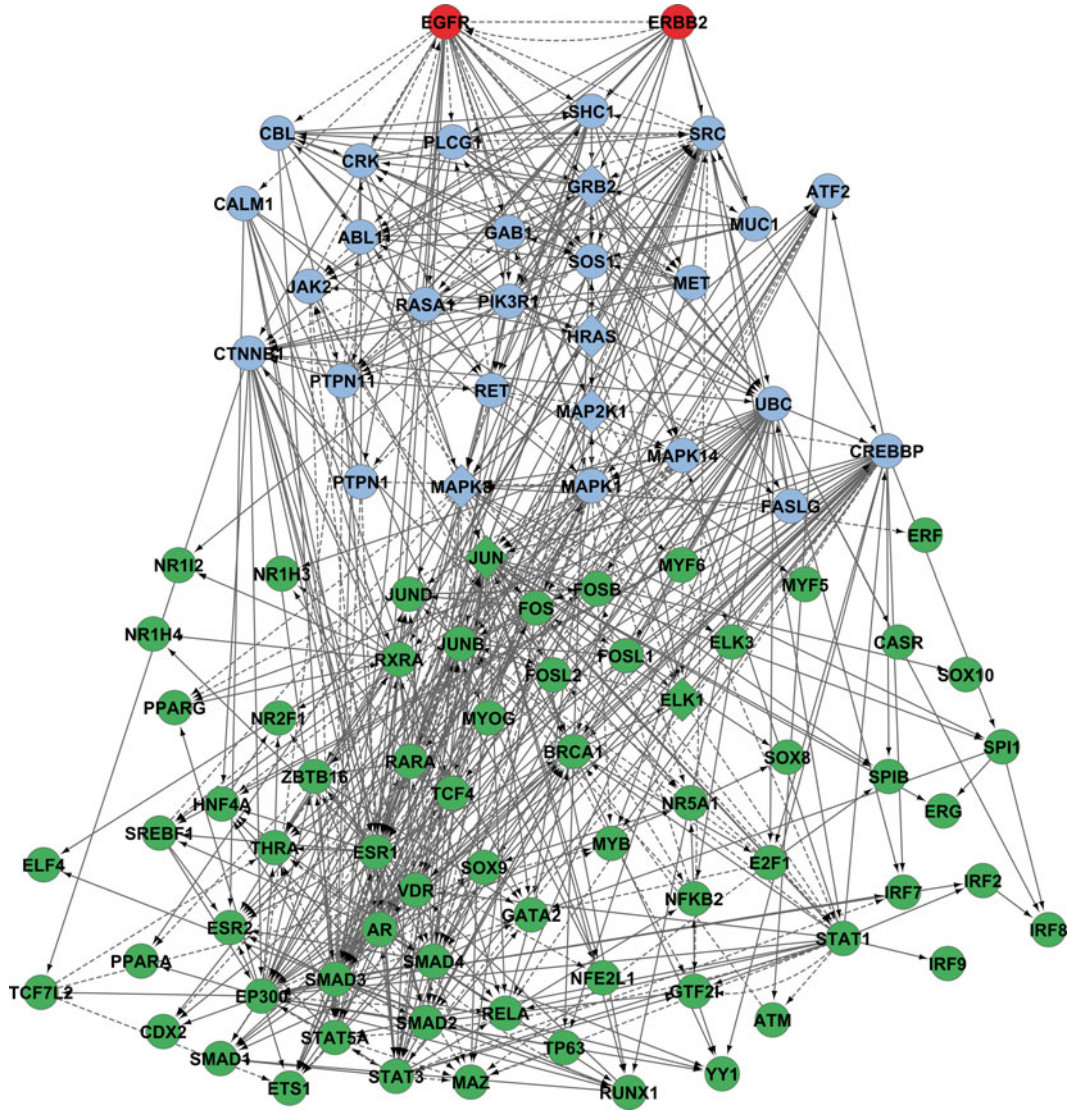


Fig. 3 SDREM EGF response signaling pathways. There are three types of nodes along the signaling pathways. *Red* nodes (*top*) are the sources given as input. *Green* nodes (*bottom*) are the active TFs on the regulatory paths (which are also assigned by SDREM as shown in Fig. 2). *Blue* nodes (*middle*) are intermediate proteins that are used to connect the sources and target TFs. Diamond shaped nodes (*ELK1*, *GRB2*, *HRAS*, *JUN*, *MAP2K1*, and *MAPK8*) were assigned a large node prior making them more likely to be included on the predicted pathway. *Circles* received the default prior. All edges among these nodes are displayed. *Solid* edges are PPI whose orientation was inferred by SDREM. *Dashed* edges are interactions with a previously known orientation

The strongest negative genetic interactions appear at the top of the sorted table, but strong positive genetic interactions at the bottom of the table may also be of interest (*see* **Notes 18** and **19**).

8. Assess the functional enrichment of the SDREM pathways. Obtain a list of all proteins on the predicted signaling pathways by opening topPathNodes_itr<N>.noa with spreadsheet software

and parsing it as a space-delimited file (for example, with the Text to Columns feature of Microsoft Excel). Copy all protein ids and paste them into a functional annotation tool such as the DAVID [12, 13] web site http://david.abcc.ncifcrf.gov/home.jsp to determine if Gene Ontology terms [14], canonical pathways, or other functional categories are enriched (*see* **Note 20**).

4 Notes

1. The SDREM download includes executable .jar files, the Java source code, default interaction networks, example data that can be used to build a model of human EGF response, and example properties files that can be modified for applications in other settings. The sample input data demonstrate the expected file formats. The EGF expression data is from [15], the sources that directly interact with EGF are from KEGG's human "ErbB signaling pathway" [16], and the node priors were assigned arbitrarily to give preference to a small number of nodes on this same KEGG pathway. The default interaction networks can be used to model other human responses and are described below.
2. Downloading and installing Cytoscape is an optional step but is recommended in order to visualize how signal propagates along the predicted pathways. Cytoscape version 3 is not fully backwards compatible with version 2.8 so the exact steps needed to load the SDREM pathways in Cytoscape 3 may vary.
3. The DREM manual, which is available from http://sb.cs.cmu.edu/drem/, describes the three supported expression file formats and which value of the DREM parameter 'Normalize_Data' they correspond to. The example expression data is formatted to use the 'Log normalize data' value of this parameter. In this format gene ids are given in the first column and gene expression values (e.g. microarray probe intensity or RNA-Seq fragments per kilobase of transcript per million mapped reads) are provided in subsequent columns. The first row provides the time points at which expression was measured, and the first time point is the baseline 0 min measurement.
4. The TF-gene interactions are provided as a tab-delimited text file, and a default file is provided (*see* **Note 6**). The first line gives the TF ids (column labels) using the same identifiers that appear in the PPI network. The first entry of all subsequent lines (row labels) is a gene id using the same identifiers that appear in the gene expression data. DAVID (http://david.abcc.ncifcrf.gov/conversion.jsp) [17], UniProt (http://www.uniprot.org/mapping/) [18], or other websites can be used to convert protein and gene identifiers if needed. Each row-column

entry corresponds to the interaction between that gene and TF. The value should be 0 if the TF does not bind the gene and 0.5 if it does (assuming a uniform TF activity prior). TFs and genes may not appear in multiple columns and rows, respectively. The filename must end with '0.txt', and an updated version of the file will be written at each iteration as the TF activity priors are refined based on the predicted signaling pathways.

5. The TF list is a list of TF protein ids with one TF per line that uses the same identifiers that appear in the PPI network. During randomization testing, random targets are drawn from this list of TFs. They may be the same TFs present in the TF-gene binding data or a superset that includes other TFs for which binding data are not available. The 'random.target.ratio' parameter and the number of active TFs determine how many random targets will be drawn at each iteration. SDREM will throw an error if not enough random targets are available, which can be avoided by adding more TFs to the TF list (preferred solution) or decreasing the 'random.target.ratio' parameter.
6. The default TF-gene interactions packaged with SDREM are predicted interactions from [19] that were processed as described in ref. [4]. All node ids are Entrez gene ids. When possible it is preferable to prepare condition-specific TF-gene interactions that are specific to the cell or tissue type and/or biological condition being studied. Data for many human cell lines can be obtained from the ENCODE project [20].
7. The PPI network file is a text file with four tab-separated columns. A default network is provided with SDREM (*see* **Note 8**). Each line specifies an interaction between proteins A and B. The first column is protein A, the second is the type of interaction, the third is protein B, and the fourth is the interaction score, which is typically a probability in the range [0, 1]. Three types of interactions can be specified in the second column: 'pp' for an undirected PPI, 'ptm' for a directed post-translational modification from A to B, or 'pd' for a directed protein–DNA interaction from A to B (protein A regulates the gene that encodes B). Only one undirected and one directed interaction between A and B are allowed, but directed interactions from A to B and from B to A are acceptable.
8. The default protein interaction network packaged with SDREM is derived from the predicted TF-gene interactions, PPI from BioGRID [21], and PPI and post-translational modifications from HPRD [22] as described in ref. [6]. All protein ids are Entrez gene ids. Alternative PPI networks can be obtained from one of the many databases cataloged by Pathguide [23]. In addition, PSISCORE scoring servers or

other resources that integrate interactions from multiple databases with PSICQUIC [24] can be used to obtain weighted PPI networks.

9. The node priors file is a two column tab-separated file. The first column is the protein id using the same type of identifiers as the PPI network. The second column is the prior probability that the protein is involved in the response in the range [0, 1]. For all proteins not listed in the node priors file, the default prior specified by the 'default.node.prior' parameter will be used. Node priors can be used to give preference to genes implicated by functional screens (such as RNAi screens, as in ref. [6]) or on canonical representations of relevant pathways [16, 25], as in our EGF example.
10. The paths enumerated by StorePaths.jar can be reused by SDREM in many cases. As long as the sources, TF list, PPI network, node priors, 'max.path.length' parameter, and 'default.node.prior' parameter do not change, SDREM can be started from **step 3** of Subheading 3. This can save time when testing different TF-gene binding interaction or gene expression datasets or tuning SDREM's parameters.
11. Stored paths can consume tens or hundreds of gigabytes (GB) for human input data when using the default parameters. However, if the 'path.enum.bound' will not be changed after enumerating paths (i.e. the full set of paths written to the 'stored.paths.dir' directory will not be filtered again using a different value) then the 'stored.paths.dir' directory can be deleted after StorePaths.jar terminates. SDREM's 'stored.paths.dir' parameter can be set to the StorePaths.jar 'filtered.paths.dir' parameter so that the StorePaths.jar 'stored.paths.dir' directory is not read.
12. Several SDREM parameters can be tuned to control the runtime and the size of the predicted pathways. The number of retained paths from sources to targets has a substantial impact on runtime and is controlled by the 'path.enum.bound' parameter. Decreasing the number of paths makes SDREM faster but less accurate, as the score for an oriented network will be a worse approximation of the true score that considers all possible source-target paths. The value of 'path.enum.bound' in the properties file passed to StorePaths.jar must be greater than or equal to the value in the SDREM properties file if the filtered paths directory is the input directory for SDREM ('stored.paths.dir' parameter). The number of possible paths grows exponentially with the path length bound so increasing 'max.path.length' has a considerable impact on the time needed to enumerate paths. However, once paths have been stored in the pre-processing step, SDREM's runtime is much more dependent on 'path.enum.bound' than 'max.path.length'.

Lowering 'dist.thresh' increases the number of active TFs in the model. Decreasing 'target.thresh' has a similar effect because it causes more TFs to be considered well-connected in the signaling network, which raises their binding prior at the next SDREM iteration and increases their activity score. Reducing 'node.thresh' causes more nodes to be predicted as internal signaling nodes in the SDREM model and may raise the binding prior of TFs that were not selected as active TFs in the current iteration.

13. We recommend running SDREM on a multi-core machine or cluster for human analyses. SDREM automatically uses all processors available to the Java Virtual Machine to execute critical parts of the algorithm in parallel. An example of how to run SDREM on a cluster using the Portable Batch System job queueing software is included with SDREM, but SDREM can be run with other cluster configurations as well. Even when running in parallel, SDREM may take multiple days to run on large datasets. If SDREM is run on a cluster, the contents of the model directory can be downloaded to a local machine for visualization in DREM and Cytoscape.
14. SDREM writes the file targetsByIteration.txt in the model directory, which can be used to quickly assess whether it has converged. This file contains the active TFs at each iteration, which should not change substantially during the final few iterations. If SDREM has not converged, it can be run again with a larger "iterations" parameter.
15. Once the SDREM model is loaded into DREM for visualization, the Key TFs Labels button can be used to visualize active TFs based on SDREM's activity scores. Choose to display key TFs based on Activity Score, which shows a TF each time its score exceeds the threshold, or use the option to only annotate TFs the first time they are active on a path. The SDREM output file <N>.targetsStd provides the activity score for all active TFs, which can be used to set a key TF activity score threshold. Find the minimum activity score in the fourth column of this file, take the $\log_{10}$ of the minimum score, and use the slider in the Key TFs Labels window to set X to that value. The number that appears after a TF name indicates the primary path out of the split that it controls. At a split with k outgoing paths, [1] is the lowest path and [k] is the highest path. If TF labels overlap (as in Fig. 2), left clicking a TF text box will hide its text.
16. The Save Image button captures an image of the DREM model. Other buttons on the DREM interface are explained in the DREM manual (http://sb.cs.cmu.edu/drem/).
17. The .sif file that can be loaded into Cytoscape contains all edges between proteins on the high-confidence pathways (sources, internal signaling nodes, and target TFs), not the

top-ranked paths. The file satisfiedPaths_itr<N>.txt.gz in the SDREM model directory, where N is the final iteration number, can be used to obtain only the top-ranked paths. The compressed file contains two tab-delimited columns. The first column contains a text representation of a path in which edges are separated by '|'. Each edge has the format <protein id>:dir:<protein id> or <protein id>:undir:<protein id> depending on whether the edge was directed or undirected in the input network. The path source is the leftmost protein and the target TF is the rightmost. Paths can be sorted by the second column, the path weight, to obtain only the top-ranked paths and visualize only the edges on these paths.

18. There are four possible values for the SDREM 'predict.knockdown' parameter: SingleTop, SingleAll, DoubleTop, and DoubleAll. Single means that only single gene knockdown effects are predicted and Double denotes that both single and pairwise effects are predicted. Top means that predictions are made for the sources and for nodes with at least 'node.thresh' (from the SDREM parameters) fraction of top-ranked paths passing through them. All makes predictions for all nodes in the network. DoubleAll is supported but not recommended because it is very slow, and it is typically sufficient to focus on only those pairs in which at least one of the proteins is expected to have an impact upon knockdown. If the 'predict.knockdown' field is left empty, then knockdown predictions will not be made.
19. The topTargAvg metric is the recommended criterion for sorting genes by their predicted knockdown effect because it was shown to be effective when studying H1N1 influenza [6]. The additional ranking metrics described in ref. [6] are available as well and are written to the same output files. When selecting genes for experimental validation, the Source and Degree columns can be considered in addition to the rankings based on the predicted knockdown effects. Both sources and high-degree nodes are more likely to have source-target paths pass through them, leading to large predicted effects. Therefore it is more surprising when low-degree non-source nodes have relatively substantial impacts and these may be more promising candidates for validation.
20. Many other tools, as reviewed in [13], can be used as alternatives to DAVID for functional analysis of SDREM's predicted pathway members. In addition to analyzing all of the nodes on the predicted signaling pathways, we also recommend assessing the enrichment of only the internal nodes and targets (filtering the sources). The sources are already known to be relevant to the response and therefore can bias the enrichment analysis. Enrichment analysis of the top-ranked knockdown effect

predictions is also valuable. In addition, to obtain accurate significance calculations in a tool such as DAVID the background should be set manually to include only the proteins in the interaction network because only these proteins can appear in the SDREM predictions. A list of all proteins in the network can be obtained from the node scores output file.

Acknowledgements

This work was supported by National Institutes of Health (1RO1 GM085022) and National Science Foundation (DBI-0965316) awards to Z.B.J. A.G. is supported by Microsoft Research.

References

1. Bar-Joseph Z, Gitter A, Simon I (2012) Studying and modelling dynamic biological processes using time-series gene expression data. Nat Rev Genet 13:552–564
2. Gitter A, Carmi M, Barkai N, Bar-Joseph Z (2013) Linking the signaling cascades and dynamic regulatory networks controlling stress responses. Genome Res 23:365–376
3. Ernst J, Vainas O, Harbison CT et al (2007) Reconstructing dynamic regulatory maps. Mol Syst Biol 3:74
4. Schulz MH, Devanny WE, Gitter A et al (2012) DREM 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data. BMC Syst Biol 6:104
5. Gitter A, Klein-Seetharaman J, Gupta A, Bar-Joseph Z (2011) Discovering pathways by orienting edges in protein interaction networks. Nucleic Acids Res 39:e22
6. Gitter A, Bar-Joseph Z (2013) Identifying proteins controlling key disease signaling pathways. Bioinformatics 29:i227–i236
7. Wiltrout C, Lang B, Yan Y et al (2007) Repairing brain after stroke: a review on post-ischemic neurogenesis. Neurochem Int 5:1028–1041
8. Jin K, Peel AL, Mao XO et al (2004) Increased hippocampal neurogenesis in Alzheimer's disease. Proc Natl Acad Sci U S A 101:343–347
9. Repetto E, Yoon I-S, Zheng H, Kang DE (2007) Presenilin 1 regulates epidermal growth factor receptor turnover and signaling in the endosomal-lysosomal pathway. J Biol Chem 282:31504–31516
10. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504
11. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
12. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
13. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13
14. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29
15. Amit I, Citri A, Shay T et al (2007) A module of negative feedback regulators defines growth factor signaling. Nat Genet 39:503–512
16. Kanehisa M, Goto S, Sato Y et al (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114
17. Huang DW, Sherman BT, Stephens R et al (2008) DAVID gene ID conversion tool. Bioinformation 2:428–430
18. Huang H, McGarvey PB, Suzek BE et al (2011) A comprehensive protein-centric ID mapping service for molecular data integration. Bioinformatics 27:1190–1191
19. Ernst J, Plasterer HL, Simon I, Bar-Joseph Z (2010) Integrating multiple evidence sources to predict transcription factor binding in the human genome. Genome Res 20:526–536

20. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
21. Chatr-aryamontri A, Breitkreutz B-J, Heinicke S et al (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41:D816–D823
22. Prasad TSK, Goel R, Kandasamy K et al (2009) Human protein reference database—2009 update. Nucleic Acids Res 37:D767–D772
23. Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. Nucleic Acids Res 34:D504–D506
24. Aranda B, Blankenburg H, Kerrien S et al (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Meth 8:528–529
25. Croft D, O'Kelly G, Wu G et al (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39:D691–D697

Part VI

Systems Biology of Alzheimer's Disease in Practice: From Systems Biology to Early Diagnostics and Systems Medicine

Chapter 31

Advanced Neuroimaging Methods Towards Characterization of Early Stages of Alzheimer's Disease

Jorge Sepulcre and Joseph C. Masdeu

Abstract

In the past 5 years, imaging network properties in the brain of patients with Alzheimer's disease (AD) has revolutionized our understanding of this disorder. Postmortem data had already suggested that the damage spreads along functional neural networks, but postmortem studies do not provide information on the temporal evolution of the damage in the same patient, essential to determine spreading. These data can be provided by functional and structural neuroimaging, which allow for the visualization over time of the progressive damage inflicted by AD. Functional networks can be mapped by determining the synchrony across brain regions of the blood oxygenation level dependence (BOLD) signal on functional magnetic resonance imaging (MRI) during quiet wakefulness. Other less extensively used techniques are also useful. For instance, amyloid deposition can be imaged and its progression mapped to determine whether it follows brain networks, and, if so, which are affected earliest. Network patterns of neurobiological changes, including tau deposition, may prove critical to our understanding of the neurobiology of AD and therefore open the way for therapeutic interventions.

Key words Alzheimer's disease, Neuroimaging, Network analysis, Amyloid, Graph theory, Early stages, Functional connectivity magnetic resonance imaging (fcMRI), Positron emission tomography (PET)

1 Alzheimer's Disease Is a Network Disease: Need for Network-Wise Approaches

Alzheimer's disease (AD) is a devastating neurological illness that poses a biomedical challenge for current and future generations. AD is characterized clinically by cognitive impairment and histologically by neurodegenerative changes that include the progressive accumulation of misfolded phosphorylated tau protein into intracellular neurofibrillary tangles (NFT) and the formation of extracellular neuritic plaques (NP) from insoluble self-aggregating amyloid-β (Aβ) [1–10]. However, the correlation between the degree of cognitive impairment and the amount of NFT and NP deposition in the brain is far from linear: older individuals with normal cognition who die from other diseases may have pronounced brain AD-related pathology [11, 12].

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_31,

Although brain changes in AD are complex, several histological patterns caught the attention of researchers already several decades ago. For instance, NFT and NP are not randomly distributed across the neuronal systems of the brain but have characteristic spatial patterns, such as NFT in transentorhinal cortex, and NP in ventral regions of temporal and frontal lobes in early stages of the disease. Both neuropathology landmarks largely affect heteromodal and association areas when AD is well established [1, 2, 13–20]. The evidence of deposits of NFT and NP in interrelated brain systems suggests that AD pathology may be associated with anatomical interconnectivity, particularly between medial temporal lobe and association cortices in the precuneus, posterior cingulate, and inferior parietal and lateral temporal cortices—regions that are now known to belong to the default mode network (DMN) or cortical hubs network [21, 22].

A remarkable consequence of interpreting AD as a brain network disease is that it establishes a plausible framework to understand how it spreads via neuronal connectivity [1, 13, 15]. Although the molecular basis by which toxicity transfers neuron-to-neuron remains largely elusive, it is likely that a transneuronal spread by a "prion-like" mechanism may take place in the progression of AD. Neurobiology studies with animal models support the idea that *trans*-axonal and *trans*-synaptic propagation mechanisms through interconnected neural networks might contribute to Aβ deposition in AD. The over-expression of the amyloid precursor protein in the entorhinal cortex of transgenic mice can lead to amyloid plaques in its synaptic terminal zone located in the dentate gyrus [23]. Moreover, lesioning the perforant pathway, which connects the entorhinal cortex to the hippocampal formation, decreases Aβ deposition in the hippocampus of APP transgenic mice, which are prone to overproducing Aβ [24]. Another possibility that may be the basis of the spreading nature of the disease is the transneuronal transport of abnormal tau [25].

The paradigm of AD as a network disease has opened new avenues of research not only at the neuronal level in basic experimental settings but also at higher scale levels, especially in systems neuroscience and neuroimaging. For instance, in the past several years we have witnessed a major transformation in the studies of neurodegeneration with neuroimaging moving from analysis of atrophy and intensity changes to investigations of functional and structural brain networks [26, 27]. This transition is not surprising. If AD encompasses complex brain disruptions that follow neuronal circuits, then network tools are more appropriate to capture its fingerprints of neurodegeneration. In a similar manner, as biology has evolved into systems biology, neuroimaging has complemented its conventional methods with the addition of network neuroimaging approaches. The old Hebbian principle has been reformulated to state not only "neurons that fire together, wire

together" [28], but also, "neurons that wire together, die together" [29]. This has formalized a novel framework to study the network nature of AD and engendered a renewed enthusiasm in the neuroimaging community. In the two following sections, we present recent advances of two network neuroimaging techniques: one using functional connectivity magnetic resonance imaging (fcMRI) and the other based on positron emission tomography (PET) imaging. Of note, although other imaging connectivity techniques such as MRI diffusion tensor imaging (MSI DTI) or diffusion spectrum magnetic resonance imaging (DSI) and electroencephalography/magnetoencephalography (EEG/MEG) are also important in the field of AD, we have focused this chapter only on functional and molecular imaging approaches.

2 Disruption of Functional Brain Networks in AD

Despite the fact that AD has been recognized as a network disorder for several decades now, only recently neuroimaging approaches have focused on the study of neurodegenerative brain networks. The neglect of a network-oriented focus was, in part, due to the limitations of appropriate neuroimaging technologies for surveying distributed brain systems. However, the development of functional and structural connectivity approaches in the MRI field opened new possibilities to reveal brain network breakdown in AD. Of special relevance is the use of fcMRI in the study of AD. Network analysis based on fcMRI technique has created great expectations for its potential to reveal functional network disruptions. fcMRI is a neuroimaging technique that uses the signal couplings of spontaneous low-frequency blood oxygenation level dependence (BOLD) fluctuations at rest, also referred as intrinsic activity, to investigate functional connections—or disconnections—between brain regions [30, 31]. fcMRI has several clinical advantages compared to conventional functional MRI (fMRI). It is a task-free and data-driven neuroimaging approach that can be easily used in cognitively impaired patients that otherwise would have difficulty performing tasks.

Brain functionality changes in AD are system-specific rather than just compensatory or deleterious effects in isolated regions [32]. fcMRI studies using independent component analysis (ICA) and seed-based approaches have shown that functional disruptions in AD occur within the boundaries of well-known connectivity networks [33, 34], confirming in vivo what previous histological findings pointed to before. Interestingly, early disconnectivity between the medial temporal lobe and the default mode network (DMN) has been described in preclinical stages, even in the absence of amyloid deposits detectable by PET [35]. However, functional connectivity changes in individuals with AD compared to control

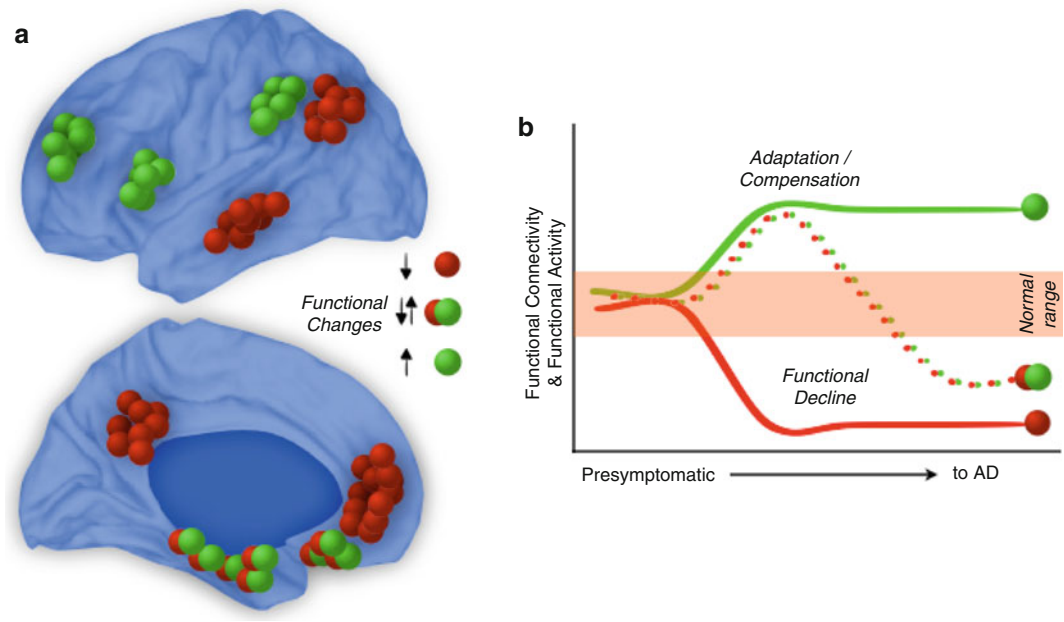


Fig. 1 Brain networks functional changes during neurodegeneration. Brain networks present divergent functional changes on MRI BOLD activation and default functional connectivity during neurodegeneration (**a**, **b**). Default mode network (*red* nodes) shows consistent decreases whereas fronto-parietal network (*green* nodes) shows increases of functional connectivity. Medial temporal lobe tends to exhibit an initial increase of functional activity, with a subsequent decrease (*red-green* nodes/line in **b**)

subjects are not limited to decreased connectivity in some networks (particularly in the DMN; red nodes in Fig. 1a), as would be expected, but also increased connectivity in others (particularly in fronto-parietal networks; green nodes in Fig. 1a; *see* Sheline and Raichle [36], for a recent review of this topic). In one sense, the idea that accumulating pathology disrupts brain networks accommodates well the findings of decreased connectivity in the preclinical or clinical AD brain. Conversely, the fact that some AD-affected areas increase their functional activity and connectivity makes the overall scenario more complex to interpret.

A tentative explanation of this phenomenon is the compensatory adaptation theory [37, 38]. Enhancement of BOLD activity during memory-task performances—particularly in frontal areas during aging or AD stages—has been frequently described before [37]. More recently, Zamboni et al. [39] have shown that this BOLD phenomenon during memory tasks overlaps with an increase of functional connectivity at rest, without any cognitive demand. Therefore, what once was thought to be an adaptive or compensatory task-related phenomenon it is now seen as an intrinsic feature that emerges in the AD brain network reorganization. But is this a real intrinsic network adaptation of the neurodegenerating brain? If so, why do only specific areas in the medial temporal network or

the frontal-parietal network undergo this adaptive phenomenon? Why do other key AD-related areas, such as those within the DMN, not adapt their functional connectivity? These questions still require clarification.

A factor that may play a role in the functional discrepancies between networks is their differential affinity for Aβ deposits. Networks that display functional increase, such as the medial temporal and fronto-parietal networks, have less absolute affinity for Aβ deposition than networks that show functional decrease, such as the DMN. Therefore, it is possible that the degree of functional activity and connectivity adaptation may be related to the anatomical distribution of amyloid accumulation, and perhaps to the distribution of other abnormal proteins, such as hyper-phosphorylated tau. In this sense, the divergent functional changes found in AD are probably rooted in histological features of neurons and glia associated with each network. Anatomical studies intended to disentangle the connectivity underpinnings of brain systems [40] as well as the differential susceptibility to histological neurodegeneration of networks will help in the future to understand the large-scale functional connectivity changes in AD.

One further step in the study of neurodegeneration in AD has been the use of neuroimaging in the framework of graph theory. The study of brain functional and structural networks as graphs, which transform neuroimaging connectivity to nodes (vertices) and links (edges), has facilitated the comprehension and visualization of their complex organization. A classic and frequently used metric in graph theory is degree centrality [41]. It refers to the number of nodes that a given node is directly linked to in the graph, and therefore, it quantifies a type "network hubs" or central nodes in the network. This metric can be applied to resting state fcMRI data to show which regions (nodes) are the most central areas as a result of their connections to the rest of the brain (*see* [21] and [22] for methodological details of voxel-based degree centrality implemented in fcMRI of high resolution). This strategy has uncovered remarkable findings with regard to AD [27]. For instance, we now know that there is a striking overlap between the centers of higher functional connectivity in the cortex and the spatial distribution of Aβ deposits [22]. Moreover, cortical hub regions have been related to potential main routes by which the disease may progress in the human brain [42, 43].

The overlap between functional hubs in normal individuals and the Aβ deposits in AD subjects has a significant implication. It has been postulated that continuous sustained functional activity in cortical hubs may result in amyloid accumulation, leading to progressive neurodegeneration [22]. The "synaptic excitatory toxicity hypothesis" states that neural damage can result from the over-activation of *N*-methyl-D-aspartic acid receptors by glutamate at the synaptic level [44–46]. Regions such as the cortical hubs

with high basal metabolism and connectivity, and presumably high synaptic activity, could be targets for an Aβ-related neurodegenerative cascade [9, 47–50]. Although this assumption needs further testing, the spatial match between cortical hubs and Aβ deposits suggests that high degree centrality may lead to activity-dependent neurodegeneration.

3 Characterization of Pathological Brain Networks with PET Imaging in AD

The introduction of PET imaging using Pittsburgh Compound-B (PIB), or ^{11}C-labeled agent *N*-methyl^{11}C-2-(4-methylaminophenyl)-6-hydroxybenzothiazole tracer, has advanced our capabilities to detect in vivo Aβ deposition in clinical and preclinical AD [12, 51–54]. Although PET-PIB is the most frequently used tracer to detect AD-related pathology, PET imaging is constantly evolving and new tracers based on an F-18-radiolabel have also been reported with promising results to visualize NP and NFT deposits [55–63]. Using molecular imaging, we have learned that Aβ accumulates in vivo from healthy aging to AD in a continuous manner [64–67], reaching a plateau or saturation phase soon after the disease can be diagnosed on the basis of incipient cognitive impairment [64]. Therefore, due to this saturation effect, PET amyloid imaging may be especially relevant for studying early phases of AD in preclinical and normal individuals [66].

PET imaging is not only useful to detect pathological accumulation of abnormal proteins in the human brain but also to characterize its pathological networks at a large-scale level. Similar to structural MRI network approaches, spatial co-variation patterns of PET imaging data can be used to extract network properties of molecular tracers for AD. The theoretical assumption is that NP and/or NFT are not independently accumulating in brain regions; on the contrary, they relate to each other and reflect fingerprints of neurodegeneration between interconnected neurons. By using this approach, we can better understand how distributed systems are pathologically affected during disease progression. A graph theory method that characterizes spatial associations between amyloid deposits using PET-PIB imaging uses stepwise connectivity (SC) analysis (*see* Sepulcre et al. [68] for methodological details). SC analysis is able to detect, (a) direct spatial correlations of a seed region and, (b) indirect associations of that seed region to the rest of the brain through its linked neighbors. For instance, if we are interested in investigating how Aβ load in the hippocampal formation is associated with Aβ load in other parts of the brain, we can use SC analysis to determine their local spatial associations and also their indirect connections to distributed regions such as the DMN regions [68].

In a recent study, we investigated this specific matter. The medial temporal lobe in individuals with AD and elderly controls displayed striking similarities in their SC patterns. Preclinical and clinical AD

both show Aβ deposits in the hippocampus that directly relate to accumulations in areas of the orbitofrontal cortex, lateral temporal cortex, and precuneus cortex [68]. In further steps of connectivity, amyloid load in the medial temporal lobe is also spatially associated to load in wide areas in the DMN and cortical hub regions, such as midline and lateral prefrontal cortex. Interestingly, elderly controls with low Aβ burden show a prominent spatial association between the hippocampus and the orbitofrontal cortex (*see* SC maps of medial temporal lobe of elderly controls with low Aβ burden in Fig. 2a). A display of the amyloid network in elderly controls with low Aβ burden illustrates the centrality of the hippocampus,

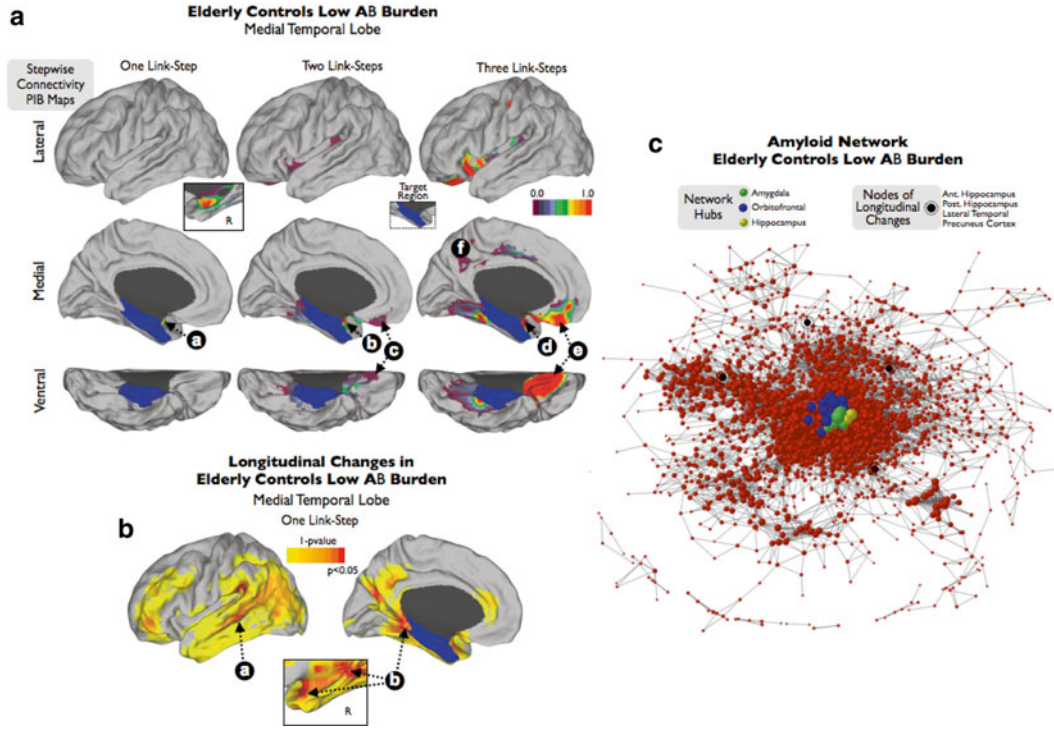


Fig. 2 Stepwise connectivity analysis of amyloid imaging on medial temporal lobe (**A**, **B**) and visualization of the amyloid network (**C**) in elderly controls with very low Aβ burden. Stepwise connectivity maps in **A** reveal that amyloid load in the medial temporal lobe of elderly controls with low Aβ burden is associated with amyloid load in ipsilateral (**a** in **A**) and contralateral amygdala and anterior hippocampus (detailed inset in **A**). In further link-step distances (two and three link-step maps to medial temporal lobe, **A**), amyloid load in the medial temporal lobe is associated with Aβ load in hippocampus formation and parahippocampal cortex, amygdala (**b** and **d** in **A**), and orbitofrontal cortex (**c** and **e** in **A**), and weakly associated with load in cingulum areas (**f** in **A**). In the 1-year follow-up longitudinal analysis of stepwise connectivity maps, medial temporal lobe shows significant changes in direct spatial associations with the lateral temporal lobes (**a** in **B**), precuneus, contralateral HFPG, and amygdala (**b** in **B**). A network graph in **C** shows the pathological amyloid network of elderly controls with very low Aβ burden (size of nodes represents degree centrality values, and larger nodes denote a high number of connections to the rest of the network). Color bars represent normalized z-score scale in **A** and 1 minus *p*-value scale in **B**. Reproduced from Sepulcre et al. [68]. In vivo characterization of the early states of the beta-amyloid network, 2013, with permission from Oxford University Press © 2013

orbitofrontal cortex, and amygdala in amyloid accumulation at the very early stages of pathology (yellow, blue and green nodes in graph of Fig. 2c). In other words, these regions represent core hubs of the spatial association between Aβ deposits (Fig. 2) and, in this sense, it is plausible to assume that they are excellent candidates for regions where the Aβ pathology may originate. Interestingly, nodes that show longitudinal changes (dark nodes in Fig. 2c), such as the posterior hippocampus, lateral temporal cortex, ventrolateral prefrontal cortex, and precuneus (Fig. 2b), are located in the periphery of the graph, suggesting that they hold potential for contributing to the expansion of the Aβ pathology in the human brain. In previous studies, it has been postulated that interactions between distant brain systems, such as the medial temporal lobe and the cingulate cortex, may be a key factor for conversion to AD stages [69]. A network approach such as the SC analysis of Aβ accumulations may help in the future to describe putative pathways for the spread of pathology through interconnected neurons, revealing the neurobiology of progression to cognitive decline in individual cases.

4 Conclusions

The emergence of network analytical approaches in neuroimaging [70, 71] has revolutionized the study of AD neurodegeneration and the functional and molecular breakdowns attending its progression. A renewed enthusiasm to (re)interpret AD as a network disorder brings the attractive possibility of approaching its complex nature with more suitable tools. Much work is needed to characterize the whole spectrum of brain histological and functional changes in AD, but network neuroimaging has already provided key findings to understand spatial associations and connectivity patterns that may be useful for disease staging in individuals.

Acknowledgements

Author thanks Elizabeth Beam for their helpful suggestions and comments about the manuscript. This work was supported by the Alzheimer's Association Grants NIRG-11-205690 (J.S.).

References

1. Arnold SE, Hyman BT, Flory J et al (1991) The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. Cereb Cortex 1: 103–116
2. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82:239–259
3. Braak H, Braak E (1991) Demonstration of amyloid deposits and neurofibrillary changes in whole brain sections. Brain Pathol 1:213–216

4. Mann DM (1985) The neuropathology of Alzheimer's disease: a review with pathogenetic, aetiological and therapeutic considerations. Mech Ageing Dev 31:213–255
5. Hyman BT, Phelps CH, Beach TG et al (2012) National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 8:1–13
6. Ikonomovic MD, Klunk WE, Abrahamson EE et al (2008) Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain 131:1630–1645
7. Mattson MP (2004) Pathways towards and away from Alzheimer's disease. Nature 430: 631–639
8. Ogomori K, Kitamoto T, Tateishi J et al (1989) Beta-protein amyloid is widely distributed in the central nervous system of patients with Alzheimer's disease. Am J Pathol 134: 243–251
9. Selkoe DJ (2006) The ups and downs of Abeta. Nat Med 12:758–759, discussion 759
10. Walsh DM, Selkoe DJ (2004) Oligomers on the brain: the emerging role of soluble protein aggregates in neurodegeneration. Protein Pept Lett 11:213–228
11. Brewer JB, Sepulcre J, Johnson KA (2013) Structural, functional and molecular neuroimaging biomarkers for Alzheimer's disease. In: Charney DS, Nestler EJ, Sklar P, Buxbaum JD (eds) Neurobiology of mental illness, 4th edn. Chapter 62. Oxford University Press, NY, USA, pp. 821–833.
12. Masdeu JC, Kreisl WC, Berman KF (2012) The neurobiology of Alzheimer disease defined by neuroimaging. Curr Opin Neurol 25: 410–420
13. Pearson RC, Powell TP (1989) The neuroanatomy of Alzheimer's disease. Rev Neurosci 2:101–122
14. Saper CB, Wainer BH, German DC (1987) Axonal and transneuronal transport in the transmission of neurological disease: potential role in system degenerations, including Alzheimer's disease. Neuroscience 23:389–398
15. Braak H, Braak E, Ohm T, Bohl J (1989) Alzheimer's disease: mismatch between amyloid plaques and neuritic plaques. Neurosci Lett 103:24–28
16. Hyman BT, Van Hoesen GW, Damasio AR, Barnes CL (1984) Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. Science 225:1168–1170
17. Kalus P, Braak H, Braak E, Bohl J (1989) The presubicular region in Alzheimer's disease: topography of amyloid deposits and neurofibrillary changes. Brain Res 494:198–203
18. Lewis DA, Campbell MJ, Terry RD, Morrison JH (1987) Laminar and regional distributions of neurofibrillary tangles and neuritic plaques in Alzheimer's disease: a quantitative study of visual and auditory cortices. J Neurosci 7:1799–1808
19. Thal DR, Rub U, Orantes M, Braak H (2002) Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology 58:1791–1800
20. Braak H, Alafuzoff I, Arzberger T et al (2006) Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol 112:389–404
21. Sepulcre J, Liu H, Talukdar T et al (2010) The organization of local and distant functional connectivity in the human brain. PLoS Comput Biol 6:e1000808
22. Buckner RL, Sepulcre J, Talukdar T et al (2009) Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J Neurosci 29:1860–1873
23. Lazarov O, Lee M, Peterson DA, Sisodia SS (2002) Evidence that synaptically released beta-amyloid accumulates as extracellular deposits in the hippocampus of transgenic mice. J Neurosci 22:9785–9793
24. Buxbaum JD, Thinakaran G, Koliatsos V et al (1998) Alzheimer amyloid protein precursor in the rat hippocampus: transport and processing through the perforant path. J Neurosci 18:9629–9637
25. De Calignon A, Polydoro M, Suarez-Calvet M et al (2012) Propagation of tau pathology in a model of early Alzheimer's disease. Neuron 73:685–697
26. Seeley WW, Crawford RK, Zhou J et al (2009) Neurodegenerative diseases target large-scale human brain networks. Neuron 62:42–52
27. Tijms B, Wink A, De Haan W et al (2013) Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. Neurobiol Aging 34:2023–2036
28. Hebb DO (1961) Distinctive features of learning in the higher animal. In: Delafresnaye JF (ed) Brain mechanisms and learning. Oxford University Press, London, UK
29. Sepulcre J, Sabuncu MR, Johnson KA (2012) Network assemblies in the functional brain. Curr Opin Neurol 25:384–391
30. Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor

cortex of resting human brain using echo-planar MRI. Magn Reson Med 34:537–541

31. Van Dijk KR, Hedden T, Venkataraman A et al (2010) Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol 103: 297–321
32. Jagust W (2013) Vulnerable neural systems and the borderland of brain aging and neurodegeneration. Neuron 77:219–234
33. Greicius MD, Srivastava G, Reiss AL, Menon V (2004) Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci U S A 101:4637–4642
34. Greicius M (2008) Resting-state functional connectivity in neuropsychiatric disorders. Curr Opin Neurol 21:424–430
35. Sheline Y, Morris J, Snyder A et al (2010) APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF Aβ42. J Neurosci 30:17035–17040
36. Sheline YI, Raichle ME (2013) Resting state functional connectivity in preclinical Alzheimer's disease. Biol Psychiatry 74:340–347
37. Dennis N, Daselaar S, Cabeza R (2007) Effects of aging on transient and sustained successful memory encoding activity. Neurobiol Aging 28:1749–1758
38. Prvulovic D, Bokde A, Faltraco F, Hampel H (2011) Functional magnetic resonance imaging as a dynamic candidate biomarker for Alzheimer's disease. Prog Neurobiol 95:557–569
39. Zamboni G, Wilcock GK, Douaud G et al (2013) Resting functional connectivity reveals residual functional activity in Alzheimer's disease. Biol Psychiatry 74:375–383
40. Mesulam M (2012) The evolving landscape of human cortical connectivity: facts and inferences. Neuroimage 62:2182–2189
41. Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge, UK
42. Raj A, Kuceyeski A, Weiner M (2012) A network diffusion model of disease progression in dementia. Neuron 73:1204–1215
43. Zhou J, Gennatas ED, Kramer JH et al (2012) Predicting regional neurodegeneration from the healthy brain functional connectome. Neuron 73:1216–1227
44. Gouras GK, Relkin NR, Sweeney D et al (1997) Increased apolipoprotein E epsilon 4 in epilepsy with senile plaques. Ann Neurol 41: 402–404
45. Mackenzie IR, Miller LA (1994) Senile plaques in temporal lobe epilepsy. Acta Neuropathol 87:504–510
46. Palop JJ, Chin J, Mucke L (2006) A network dysfunction perspective on neurodegenerative diseases. Nature 443:768–773
47. Bero AW, Bauer AQ, Stewart FR et al (2012) Bidirectional relationship between functional connectivity and amyloid-beta deposition in mouse brain. J Neurosci 32:4334–4440
48. Hardy JA, Higgins GA (1992) Alzheimer's disease: the amyloid cascade hypothesis. Science 256:184–185
49. Hardy J (1992) An "anatomical cascade hypothesis" for Alzheimer's disease. Trends Neurosci 15:200–201
50. Mintun MA, Larossa GN, Sheline YI et al (2006) [11C]PIB in a nondemented population: potential antecedent marker of Alzheimer disease. Neurology 67:446–452
51. Bacskai BJ, Hickey GA, Skoch J et al (2003) Four-dimensional multiphoton imaging of brain entry, amyloid binding, and clearance of an amyloid-beta ligand in transgenic mice. Proc Natl Acad Sci U S A 100:12462–12467
52. Klunk WE, Engler H, Nordberg A et al (2004) Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. Ann Neurol 55:306–319
53. Klunk WE, Wang Y, Huang GF et al (2001) Uncharged thioflavin-T derivatives bind to amyloid-beta protein with high affinity and readily enter the brain. Life Sci 69:1471–1484
54. Nordberg A (2004) PET imaging of amyloid in Alzheimer's disease. Lancet Neurol 3:519–527
55. Jureus A, Swahn BM, Sandell J et al (2010) Characterization of AZD4694, a novel fluorinated Abeta plaque neuroimaging PET radioligand. J Neurochem 114:784–794
56. Nelissen N, Van Laere K, Thurfjell L et al (2009) Phase 1 study of the Pittsburgh compound B derivative 18F-flutemetamol in healthy volunteers and patients with probable Alzheimer disease. J Nucl Med 50:1251–1259
57. Rowe CC, Ackerman U, Browne W et al (2008) Imaging of amyloid beta in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. Lancet Neurol 7:129–135
58. Small GW, Kepe V, Ercoli LM et al (2006) PET of brain amyloid and tau in mild cognitive impairment. N Engl J Med 355:2652–2663
59. Thompson PW, Ye L, Morgenstern JL et al (2009) Interaction of the amyloid imaging tracer FDDNP with hallmark Alzheimer's disease pathologies. J Neurochem 109:623–630
60. Tolboom N, Van der Flier WM, Yaqub M et al (2009) Relationship of cerebrospinal fluid markers to 11C-PiB and 18F-FDDNP binding. J Nucl Med 50:1464–1470

61. Tolboom N, Van der Flier WM, Boverhoff J et al (2010) Molecular imaging in the diagnosis of Alzheimer's disease: visual assessment of [11C]PIB and [18F]FDDNP PET images. J Neurol Neurosurg Psychiatry 81:882–884
62. Wong DF, Rosenberg PB, Zhou Y et al (2010) In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (florbetapir [corrected] F 18). J Nucl Med 51:913–920
63. Xia CF, Arteaga J, Chen G et al (2013) (18) FT807, a novel tau positron emission tomography imaging agent for Alzheimer's disease. Alzheimers Dement 9:666–676
64. Jack CR Jr, Lowe VJ, Weigand SD et al (2009) Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. Brain 132: 1355–1365
65. Sojkova J, Zhou Y, An Y et al (2011) Longitudinal patterns of beta-amyloid deposition in nondemented older adults. Arch Neurol 68:644–649
66. Villain N, Chetelat G, Grassiot B et al (2012) Regional dynamics of amyloid-β deposition in healthy elderly, mild cognitive impairment and Alzheimer's disease: a voxelwise PiB-PET longitudinal study. Brain 135:2126–2139
67. Villemagne VL, Pike KE, Chetelat G et al (2011) Longitudinal assessment of Abeta and cognition in aging and Alzheimer disease. Ann Neurol 69:181–192
68. Sepulcre J, Sabuncu M, Becker A et al (2013) In vivo characterization of the early states of the amyloid-beta network. Brain 136:2239–2252
69. Tosun D, Schuff N, Mathis CA et al (2011) Spatial patterns of brain amyloid-beta burden and atrophy rate associations in mild cognitive impairment. Brain 134:1077–1088
70. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198
71. Sporns O, Chialvo DR, Kaiser M, Hilgetag CC (2004) Organization, development and function of complex brain networks. Trends Cogn Sci 8:418–425

Chapter 32

Plasma Proteomics Biomarkers in Alzheimer's Disease: Latest Advances and Challenges

Robert Perneczky and Liang-Hao Guo

Abstract

The recent paradigm shift towards a more biologically oriented definition of Alzheimer's disease (AD) in clinical settings increases the need for sensitive biomarkers that can be applied in population-based settings. Blood plasma is easily accessible and contains a large number of proteins related to cerebral processes. It is therefore an ideal candidate for AD biomarker discovery. The present chapter provides an overview of the current research landscape in relation to blood-based AD biomarkers. Both clinical and methodological issues are covered. A brief summary is given on two relevant laboratory techniques to ascertain blood biomarker changes due to AD; methodological and clinical challenges in the field are also discussed.

Key words Alzheimer's disease, Dementia, Biomarker, Early diagnosis, Prognosis, Proteomics

1 Introduction

Alzheimer's disease (AD) is typically characterized by a slowly progressive neurodegenerative process with a duration of several decades [1]. The disease can be pragmatically divided into three stages: an initial preclinical or asymptomatic stage; a subsequent pre-dementia stage (termed "mild cognitive impairment," MCI), in which cognitive performance starts to deteriorate but normal activities of daily living are still largely preserved [2]; and finally a dementia stage, in which the impairment of cognitive abilities becomes severe enough to have a significant negative impact on patient autonomy [3]. Until recently, an AD diagnosis could not be established in a clinical setting without a dementia syndrome. However, a paradigm shift is currently taking place towards a more biologically oriented diagnosis of AD. Newly conceptualized guidelines such as the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria encompass all AD stages [1, 4] and therefore also refer to individuals with neurodegenerative tissue changes but without cognitive impairment [5]. This way of

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_32,

thinking relies on the belief that an early diagnosis of AD will pave the way for early disease-modifying therapies in individuals with a functionally and structurally intact brain. To achieve this ambitious goal, there is an urgent need for improved biomarkers, which are not only more sensitive for early changes but can also be applied for population-based screening purposes, to accurately identify individuals with early clinical AD or at risk for future cognitive deterioration.

Substantial evidence exists that currently established fluid and imaging biomarkers such as the cerebrospinal fluid (CSF) proteins total-Tau (tTau), phosphorylated-Tau $(pTau)_{181}$ and Amyloid-β $(Aβ)_{1\text{-}42}$ [6] as well as structural and functional imaging techniques including magnetic resonance imaging (MRI) of the mediotemporal lobe and fluorodeoxyglucose positron emission tomography (PET) [7] show good diagnostic accuracy. More recently published CSF and imaging biomarker candidates such as CSF soluble amyloid precursor proteins (sAPPs) [8] and amyloid imaging [9] may also have superior diagnostic properties compared to the established markers. However, costly neuroimaging studies and invasive CSF sampling are not suitable for large-scale screening programs, which limits their clinical usefulness, even though lumbar puncture is a relatively safe and well tolerated procedure [10]. Hence, biomarkers must be developed that can be obtained with relative ease from peripheral body fluids to replace or assist the existing fluid and imaging markers.

In contrast to CSF, blood sampling does not include a laborious and painful lumbar puncture, which furthers the assessment of larger populations and the repeated examination of individual patients. Since blood is in constant contact and exchange with all organs and tissue including the brain, it reflects a wide range of physiological and pathophysiological processes, and therefore is an ideal medium for the discovery of new biomarkers for many disorders [11]. Importantly, signaling proteins, which the brain uses to exert control over many body functions, can be detected in blood [12] and changes in these signaling proteins associated with AD are likely to be related to a specific blood patterns [13]. Research into blood-based AD markers addresses several key shortcoming of the available biomarkers, namely the need for less invasive ascertainment procedures, and the restriction of the established markers to two, albeit central, aspects of the multifactorial nature of AD pathophysiology.

The present chapter provides an overview of the current research landscape in relation to blood-based AD biomarkers. Both clinical and methodological issues are covered. A brief summary is given on two relevant laboratory techniques to ascertain blood biomarker changes due to AD; methodological and clinical challenges in the field are also discussed.

2 Materials

2.1 Blood Handling

Whole blood samples are collected by venipuncture into commercially available anticoagulant-treated tubes (e.g. EDTA-treated or citrate-treated), and put into the freezer at 4 °C within 120 min in most cases. Plasma is isolated from blood cells (*see* **Note 1**) by centrifugation at 2,500×*g* for 15 min. If the sample is not analysed immediately, it should be apportioned into 500 μL aliquots; the aliquots should be immediately transferred to −20 °C for not longer than 2 weeks and stored long-term at −80 °C.

2.2 Multicentre Studies

Joint projects between different laboratories are common in biomarker development. Multicentre studies require each individual laboratory to follow exactly the same protocol for plasma collection. Aliquots of the samples are kept at −20 °C for not longer than 2 weeks, or at −80 °C until use. Plasma samples must be shipped on dry ice to the central laboratory, and are required to be immediately stored at −20 °C for not longer than 2 weeks or at −80 °C until further analysis (*see* **Note 2**).

3 Methods

3.1 Enzyme-Linked Immunosorbent Assay (ELISA)

ELISA is usually performed in duplicate and according to the commercial manufacturer's instructions. An ELISA assay includes coating, blocking, incubation, and detection according to the following general guidelines.

1. Determine wells for reagent blank, test sample blank, test sample and diluted standard, then coat the microwells with 100 μL appropriate diluted antigens (*see* **Note 3**). Incubate the plate at room temperature (RT) for 2 h or at 4 °C after covering it with plate lid overnight.
2. Vigorously wash all unbound antigen off the plate to prevent false positive for at least three times with 300 μL wash buffer (*see* **Note 4**). Then block non-specific binding by adding 200 μL of blocking buffer to each well. Incubate at RT for 1 h or at 4 °C overnight, then wash plate as above.
3. Pipette 100 μL of diluted samples (*see* **Notes 2** and **5**) and standards to appropriate wells, incubate for 1 h at room temperature (RT) or at 4 °C overnight, then repeat washing step.
4. Add 100 μL of second step antibody to wells and incubate for 1 h at RT, then repeat washing step.
5. For color development, 100 μL of chromogenic substrate is added to each well. Cover the plate and incubate for 15 min, or until a suitable color has developed. The plate should preferably be protected against light during this incubation.

6. Determine the optical density of each well within 30 min through the bottom of the microwell plate using an automated or semi-automated photometer (ELISA-reader) with appropriate wavelength. Determine the concentration of the samples from the standard curve using curve fitting software (*see* **Note 6**).

3.2 Luminex xMAP Platform

All reagents should be stored at 2–8 °C. All reagents shall be brought to RT approximately 30 min before use.

1. Prepare the antibody-immobilized polystyrene microsphere beads by sonicating each antibody-bead vial for 30 s, and vortexing for 1 min, with subsequent dilution in the appropriate volume (*see* **Note 7**).
2. Determine the number of wells of a filter-bottom microplate required for the assay. Standard curves and samples may be run in duplicates. Pre-wet the wells with working wash solution for 10 min at RT, then remove wash buffer by vacuum.
3. Sonicate prepared bead bottle for 30 s and vortex for 1 min, then pipette 25 μL of the beads into the designated wells. Once dispensed the beads should be protected from light (*see* **Note 8**).
4. Gently wash plate twice, then add 50 μL incubation buffer into each well, followed by pipetting of 50 μL of each standard, control and samples into the appropriate wells. Seal the plate and incubate for 2 h at RT on an orbital shaker to keep beads suspended during the incubation.
5. Gently remove fluid by vacuum and wash plate twice, then add 100 μL of the prepared detection antibodies to each well and incubate the plate for 1 h at RT on an orbital shaker.
6. Gently wash plate twice, and add 100 μL of the prepared streptavidin-conjugated fluorescent protein, R-Phycoerythrin (SAV-RPE) to bind to the biotinylated detector antibodies, forming a four-member, solid-phase sandwich, and incubate the plate for 30 min at RT on an orbital shaker.
7. Remove unbound SAV-RPE by washing the wells with the vacuum manifold twice, then resuspend the beads on a plate shaker for 5 min.
8. Run plate on the Luminex instrument (Luminex Corp., Austin, TX, USA) and analyse the samples (*see* **Note 9**). Determine the concentration of the samples from the standard curve using curve fitting software.

3.3 Challenges

3.3.1 Clinical Challenges

One important aspect that should be considered when planning biomarker validation studies is that the clinical diagnosis (or any measure of clinical progression) may not always be the most appropriate measure to reflect one single key player of AD

pathophysiology. An appropriate clinical marker should ideally be related to an individual's condition in cross-sectional and prospective observations. No perfect association between clinical and biological changes should be expected because of a plethora of interactions with other relevant factors such as cerebrovascular changes [14], brain reserve [15] or lifestyle (e.g. nutrition, sport). Excluding brain biopsy, which is unethical for research purposes [16], alternative benchmarks could be established using validated indicators of AD neuropathology including MRI-derived hippocampal volume [17] or PET amyloid imaging [18]. These indicators of neuropathology may however not be appropriate references if upstream markers of AD pathogenesis such as sAPP or BACE1 are studied, which reflect processes that occur before structural brain changes or Aβ deposition manifest. Lastly, the limited prognostic accuracy of both the established and the novel biomarkers has important ethical implications. Biomarker-positive, but otherwise healthy, older people with pathological biomarker results, who have only minor or no cognitive problems may experience significant psychological distress if labelled as harboring a beginning neurodegenerative disease with an uncertain prognosis and no treatment available. As long as no effective disease-modifying treatment options are available, diagnosis in individuals with presymptomatic or pre-dementia AD will have to be accompanied by appropriate psychosocial counselling in order to prevent unnecessary psychological distress.

3.3.2 Analytical Challenges

Plasma proteomics is a powerful tool to evaluate novel candidate biomarkers for diagnosis and treatment of AD by measuring changes in global protein levels and monitoring specific protein interactions. Several analytic platforms have been developed, and each of these technologies is evolving rapidly. Antibody-based technology has major advantages if quantitative results are required and large numbers of samples need to be analysed. Currently, ELISA is the gold standard in the quantification of protein biomarkers. ELISA methodology combines the specificity of antibodies with the sensitivity of simple enzyme assays by using antibodies or antigens coupled to an easily-assayed enzyme, which is capable of measuring the concentration of one protein at a time, but in a high-throughput and highly specific fashion. ELISA can detect proteins in complex matrices like blood, urine or saliva, often down to the pg per mL range, and numerous standardized commercial kits are available. Furthermore, ELISA are low cost and convenient, and many researchers have expertise in developing or applying ELISA. Therefore, the standard 96-well ELISA is available to most laboratories without the need for investment in special equipment or skills. However, ELISA also suffers from certain disadvantages. Antibodies have limited sensitivity and large amounts of samples may be required. The numerous wash steps make it

difficult to automate ELISA, and manual sample handling may be failure prone if not executed by skilled hands. A recent study across 14 clinical neurochemistry laboratories in Europe (Germany, Austria, and Switzerland) reported that the coefficient of variation CV (CV) of ELISAs were in the 20–30 % range when using commercially available ELISA kits for AD diagnosis [19].

In order to overcome some of the shortcomings of the ELISA technology, multiplex testing was developed. The Luminex xMAP platform is an alternative to microtiter-plate ELISA-type assays. This microsphere-based technology is a flow-cytometric method involving covalent coupling of specific monoclonal capturing antibodies to the surface of microspheres. When these color-coded beads are added to the sample, the specific coated beads bind to their individual targets, which are labelled by specific conjugated antibodies giving unique fluorescent signals. This allows flow-cytometric discrimination of mixed microsphere sets of extensive numbers of analytes and the coverage of a large number of biological pathways [20]. Currently beads are available in 500 different color-codes; thus up to 500 unique bioassays can in theory be performed within a single sample. Luminex xMAP provides an efficient and state-of-the-art platform to perform multiplex detection. It is well-suited to a wide range of applications for drug discovery, diagnostic testing, and basic research because it as easy, reliable and robust, but requires less amounts of consumables and reagents compared to conventional ELISA. However, most multiplex technologies, including Luminex xMAP, require the use of dedicated instruments with limited flexibility for additional analytical applications. Multiplex assays also require complex mixtures of numerous antibodies against a wide variety of biomarkers in the same sample. Although validated panels are generally available for broad applications, these panels might not contain specific targets of interest, especially if an unbiased selection strategy was employed for biomarker identification.

A recently developed bead-based Multi-Analyte Profiling (MAP) panel (Human DiscoveryMAP, Rules-Based Medicine Inc., Austin, TX), based on Luminex technology, allows for the simultaneous measurement of multiple disease related analytes. This platform is the product of many years of biomarker research in cancer, cardiovascular disease, metabolic disorders, inflammation, pre-natal screening for Down syndrome and AD [13, 21–24].

The MAP panel currently includes 307 analytes (http://rbm.myriad.com/). So far, only half of them have been analysed in AD plasma [25, 26]. The MAP panel has the major advantage that it covers a wide spectrum of biological pathways, but it is biased towards peptides which are thought to be involved in AD pathogenesis, often based on limited human in vivo evidence [27]. The MAP panel was empirically shown to be associated with less intra- and inter-assay variability [28], but it requires sophisticated

equipment and the necessity to develop multiple assays, which may be sample and time consuming as well as costly. Nevertheless, the MAP panel seems to be a promising alternative to more established platforms for biomarker discovery and validation in body fluids such as blood and CSF [29].

Variability across laboratories and assays limits the usefulness of fluid biomarkers to expert centres [30]. This variability of measurements can be introduced at multiple points, such as in the production of assay materials, during sample collection and storage, at the testing laboratory as a result of operator and instrumentation difference, and in the process of data collection, entry, and calculation. Several quality control initiates have been launched worldwide with this in mind to establish standardized protocols for biomarker assessment, including the Alzheimer's Association Global Biomarkers Consortium [31]. The large number of discovery phase proteomic studies is in contrast to the limited number of validation studies, which highlights the challenges of replication [32, 33]. Problems may arise from different experimental designs and analytical methods as well as heterogeneous cohorts [34]. Peripheral markers of disease identified in proteomics or genomics studies are influenced by a number of internal and external factors such as gender [35], age, concomitant diseases [36] and medications [37]. Future research will have to carefully adjust for these nuisance factors.

4 Notes

1. Compared with serum, plasma sampling needs no clotting time, and the isolation of cells and liquid phase is easily accomplished, thus is less time-consuming. Furthermore, plasma volume yield is about 10–20 % higher compared to serum; the concentration of proteins in plasma is also greater than in serum, which contains clotting factors and related constituents [38].
2. Test samples should be measured soon after collection. For the stored frozen samples, it is important to avoid additional freeze-thaw cycles. When using frozen samples, it is recommended to thaw the samples at a low temperature and mix them completely by vortexing prior to use. Haemolysed, icteric or lipaemic samples might invalidate certain tests.
3. Wells reserved for chromogen blanks should be left empty.
4. Always remove the wash buffer completely by tapping the precoated plate on paper towel. Do not wipe wells with paper towel.
5. Plasma samples are recommended to be diluted 1:50 to 1:100.
6. The dose-response will be non-linear beyond the standard point and inaccurate, thus do not extrapolate the standard curve

beyond the highest point; diluted samples that are greater than the highest standard should be reanalysed and the results should be multiplied by the appropriate dilution factor.

7. It is recommended to count the number of microspheres actually recovered after each coupling reaction. This can be performed by a cell counter or haemocytometer. The typical yield recovery is over 90 %.
8. Return unused beads to 2–8 °C until the expiration date noted on the assay.
9. If the plate cannot be read on the day of the assay, they may be covered and stored in a dark location at 2–8 °C overnight to read on the following day without significant loss of fluorescent intensity.

References

1. Jack CR Jr, Albert MS, Knopman DS et al (2011) Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7:257–262
2. Lopez OL (2013) Mild cognitive impairment. Continuum (Minneap Minn) 19:411–424
3. Mckhann GM, Knopman DS, Chertkow H et al (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7:263–269
4. Mckhann GM (2011) Changing concepts of Alzheimer disease. JAMA 305:2458–2459
5. Giaccone G, Arzberger T, Alafuzoff I et al (2011) New lexicon and criteria for the diagnosis of Alzheimer's disease. Lancet Neurol 10:298–299, author reply 300–301
6. Blennow K, Hampel H, Weiner M et al (2010) Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol 6:131–144
7. Frisoni GB, Fox NC, Jack CR Jr et al (2010) The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6:67–77
8. Perneczky R, Tsolakidou A, Arnold A et al (2011) CSF soluble amyloid precursor proteins in the diagnosis of incipient Alzheimer's disease. Neurology 77:35–38
9. Drzezga A, Grimmer T, Henriksen G et al (2008) Imaging of amyloid plaques and cerebral glucose metabolism in semantic dementia and Alzheimer's disease. Neuroimage 39:619–633
10. Zetterberg H, Tullhog K, Hansson O et al (2010) Low incidence of post-lumbar puncture headache in 1,089 consecutive memory clinic patients. Eur Neurol 63:326–330
11. Chan K, Lucas D, Hise D et al (2004) Analysis of the human serum proteome. Clin Proteomics 1:101–226
12. Wyss-Coray T (2006) Inflammation in Alzheimer disease: driving force, bystander or beneficial response? Nat Med 12:1005–1015
13. Ray S, Britschgi M, Herbert C et al (2007) Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. Nat Med 13:1359–1362
14. Mortimer JA (2012) The Nun Study: risk factors for pathology and clinical-pathologic correlations. Curr Alzheimer Res 9:621–627
15. Valenzuela M, Brayne C, Sachdev P et al (2011) Cognitive lifestyle and long-term risk of dementia and survival after diagnosis in a multicenter population-based cohort. Am J Epidemiol 173: 1004–1012
16. Noel-Storr AH, Flicker L, Ritchie CW et al (2013) Systematic review of the body of evidence for the use of biomarkers in the diagnosis of dementia. Alzheimers Dement 9:e96–e105
17. Jack CR Jr, Dickson DW, Parisi JE et al (2002) Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. Neurology 58:750–757
18. Driscoll I, Troncoso JC, Rudow G et al (2012) Correspondence between in vivo (11)C-PiB-PET amyloid imaging and postmortem, region-matched assessment of plaques. Acta Neuropathol 124:823–831
19. Lewczuk P, Beck G, Ganslandt O et al (2006) International quality control survey of neurochemical dementia diagnostics. Neurosci Lett 409:1–4

20. Siderowf A, Xie SX, Hurtig H et al (2010) CSF amyloid {beta} 1-42 predicts cognitive decline in Parkinson disease. Neurology 75:1055–1061
21. Amonkar SD, Bertenshaw GP, Chen TH et al (2009) Development and preliminary evaluation of a multivariate index assay for ovarian cancer. PLoS One 4:e4599
22. Gurbel PA, Kreutz RP, Bliden KP et al (2008) Biomarker analysis by fluorokine multianalyte profiling distinguishes patients requiring intervention from patients with long-term quiescent coronary artery disease: a potential approach to identify atherosclerotic disease progression. Am Heart J 155:56–61
23. Koster MP, Pennings JL, Imholz S et al (2009) Bead-based multiplexed immunoassays to identify new biomarkers in maternal serum to improve first trimester Down syndrome screening. Prenat Diagn 29:857–862
24. Steinacker P, Mollenhauer B, Bibl M et al (2004) Heart fatty acid binding protein as a potential diagnostic marker for neurodegenerative diseases. Neurosci Lett 370:36–39
25. Burnham SC, Faux NG, Wilson W et al (2014) A blood-based predictor for neocortical Abeta burden in Alzheimer's disease: results from the AIBL study. Mol Psychiatry 19:519–526
26. Guo LH, Alexopoulos P, Wagenpfeil S et al (2013) Plasma proteomics for the identification of Alzheimer disease. Alzheimer Dis Assoc Disord 227:337–342
27. O'Bryant SE, Xiao G, Edwards M et al (2013) Biomarkers of Alzheimer's disease among Mexican Americans. J Alzheimers Dis 34: 841–849
28. Hu WT, Chen-Plotkin A, Arnold SE et al (2010) Biomarker discovery for Alzheimer's disease, frontotemporal lobar degeneration, and Parkinson's disease. Acta Neuropathol 120:385–399
29. Guo LH, Alexopoulos P, Perneczky R (2013) Heart-type fatty acid binding protein and vascular endothelial growth factor: cerebrospinal fluid biomarker candidates for Alzheimer's disease. Eur Arch Psychiatry Clin Neurosci 263:553–560
30. Verwey NA, Van Der Flier WM, Blennow K et al (2009) A worldwide multicentre comparison of assays for cerebrospinal fluid biomarkers in Alzheimer's disease. Ann Clin Biochem 46:235–240
31. Carrillo MC, Blennow K, Soares H et al (2013) Global standardization measurement of cerebral spinal fluid for Alzheimer's disease: an update from the Alzheimer's Association Global Biomarkers Consortium. Alzheimers Dement 9:137–140
32. Shi M, Caudle WM, Zhang J (2009) Biomarker discovery in neurodegenerative diseases: a proteomic approach. Neurobiol Dis 35:157–164
33. Zellner M, Veitinger M, Umlauf E (2009) The role of proteomics in dementia and Alzheimer's disease. Acta Neuropathol 118:181–195
34. Zolg W (2006) The proteomic search for diagnostic biomarkers: lost in translation? Mol Cell Proteomics 5:1720–1726
35. Tian Y, Stamova B, Jickling GC et al (2012) Effects of gender on gene expression in the blood of ischemic stroke patients. J Cereb Blood Flow Metab 32:780–791
36. Tang Y, Lu A, Ran R et al (2004) Human blood genomics: distinct profiles for gender, age and neurofibromatosis type 1. Brain Res Mol Brain Res 132:155–167
37. Liao IH, Corbett BA, Gilbert DL et al (2010) Blood gene expression correlated with tic severity in medicated and unmedicated patients with Tourette Syndrome. Pharmacogenomics 11:1733–1741
38. Tammen H, Schulte I, Hess R et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5:3414–3422

Chapter 33

A Practical Guide for Exploring Opportunities of Repurposing Drugs for CNS Diseases in Systems Biology

Hongkang Mei, Gang Feng, Jason Zhu, Simon Lin, Yang Qiu, Yue Wang, and Tian Xia

Abstract

Systems biology has shown its potential in facilitating pathway-focused therapy development for central nervous system (CNS) diseases. An integrated network can be utilized to explore the multiple disease mechanisms and to discover repositioning opportunities. This review covers current therapeutic gaps for CNS diseases and the role of systems biology in pharmaceutical industry. We conclude with a Multiple Level Network Modeling (MLNM) example to illustrate the great potential of systems biology for CNS diseases. The system focuses on the benefit and practical applications in pathway centric therapy and drug repositioning.

Key words Systems biology, Disease network, Multiple level network modeling (MLNM), Drug repositioning, Pharmacology

1 Introduction

Drug discovery is a longsome and complicated process. In modern pharmacology, it often starts with identifying a disease target through genetics or "omics" large scale profiling (e.g. transcriptome, proteome, metabolome profiling) or text mining, followed by investigating pathological functions or pathways in disease models [1]. Hopefully target manipulation with a small molecule or biologic may alter phenotypes in vitro/vivo and slow down the disease progression. Although successful in some cases, targeting a single gene faces significant attrition especially for central nervous system (CNS) diseases [2]. The etiology for a complex disease is rarely a single gene but usually multiple genes in various pathological pathways [3]. It may be unclear which pathway is dominant to induce the phenotype or whether multiple pathways may simultaneously play the role. Therefore, there is an urgent need for a new paradigm to deal with such complex diseases [4].

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5_33,

Recently, the concept of systems biology has been rapidly introduced into drug discovery and drug repositioning [1, 5–11]. In a nut shell, systems biology tries to solve a problem by integrating and examining the components and their interactions in a network. The information modelled by the network can include multiple levels of biological knowledge, ranging from genes, pathways, diseases, knockin and knockout phenotypes, to compounds and disease models at cellular level (cell survival, myelination, cell proliferation, synaptic function, neurogenesis etc). A shift from gene-centric to system-centric perspectives may suggest new alternative drug discovery approaches [1, 5–11].

The pharmaceutical industry is racing from target identification, validation, compound efficiency to safety improvement [12]. Here, we reviewed recent advances in network-based systematic disease analysis and pharmacology and show how systems biology can be applied to explore regulatory mechanisms of complex CNS diseases, to analyze the mechanistic associations between diseases, and to develop drug repositioning opportunities for CNS diseases.

1.1 Unmet Medical Needs and Complex Mechanisms of CNS Diseases

Central nervous system (CNS) diseases affect human motor, behavioral, learning and/or memory abilities, which can be impaired as a consequence of disease onset and progression. Dozens of CNS diseases have been discovered since 1800s, including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS) and encephalitis. In the USA, the number of patients with PD reaches 1–1.5 million, while in developing countries like China numbers are expected to increase exponentially [13]. The ones which only affect a small fraction of people (<200,000) are identified as rare/orphan CNS diseases (e.g. HD, ALS and ataxia disorders). Rare CNS diseases are largely genetic and affect an individual's entire life, but they are normally ignored because of low commercial value. The global market for CNS diseases therapeutics remains enormous and urgent. However, an astonishingly limited number of disease-modifying drugs for treatment of CNS diseases exist [13].

The development of therapies for CNS diseases entails significantly higher risks compared to other diseases [2, 14]. For example, amyloid-β (Aβ) and tauopathies are believed to have key roles in the observed clinical symptoms and pathology for AD. However, a recent series of failures of amyloid-β-targeted therapeutics in Phase III clinical trials indicate that the Aβ cascade models might be problematic [15]. Thus, it is critical to uncover underlying disease mechanisms and validate their characteristic properties in order to be successful. Thus, the leading therapy in PD is the use of dopamine agonists, however, their efficacy remains limited to individuals with motor symptoms and it will lead to dyskinesia after a few years of treatment. Neuroprotective-focused treatments for

non motor PD symptoms such as dementia and psychosis are the main unmet medical needs for PD [16].

Many CNS diseases have complex etiologies. PD for example, with multiple genetic (up to 15 Parkinson's genetic genes or PARKs) and environmental contributing factors (e.g. aging, diet, infections, toxic environments) that lead to the degeneration of dopaminergic neurons in the substantia nigra. However, the causal factors or how they may alter the thalamocortical pathway that leads to dopamine neurons death are unknown [16]. Through "omics" profiling in PD models, multiple pathways have been found associated with PD, including "*programmed cell death*," "*oxidative stress and mitochondrial dysfunction*," "*protein degradation*," "*ion channel and neurotransmitter*," "*protein metabolisms and inflammation*" [17, 18]. Again, it is not clear the contribution of each pathway to various PD clinical symptoms. They could play a role simultaneously, sequentially, or they could reflect compensatory mechanisms in response to the disease impairment.

1.2 Therapeutic Challenge: Rethinking the Strategy for CNS Diseases

Although "omics" profiling provides data of correlations between gene/protein(s) and disease phenotype(s), it is hard to infer the causative association between gene and disease pathology. In addition, disease targets are cross-talking in the complex network. Thus, it is not straightforward to select and target just single genes in order to alter the disease process [12]. This phenomenon exists for most diseases, but especially for CNS diseases since they are polygenic, and affected by multiple environmental factors as well. The correspondence between in vitro or in vivo disease models (usually built upon a single disease mechanism) and the CNS clinical phenotypes is not guaranteed. Since the clinical symptoms are also heterogeneous, it is unclear which disease genes or pathways contribute to a specific symptom. On the other hand, single-gene manipulation often ends up with no obvious phenotype in model organisms, and drug candidates designed to target individual disease genes are experiencing high rates of attrition in clinical trials. Drug developers have realized the high risk of "one gene" dogma for CNS diseases [14]. They begin to shift from a "gene-centric" to a "network-centric" view due to the multifactorial, polypharmacological nature of many diseases [3]. The network presents regulatory associations of disease genes and may provide optimal gene(s) for therapeutic intervention through topological analysis and data integration. The efficacy could be enhanced through simultaneously manipulating multiple disease genes or pathways. Rational design and strategies could be the solution for CNS diseases in the context of pathological pathways, combined therapeutics, or drugs targeting multiple disease genes.

CNS diseases may share common molecular mechanisms despite heterogeneous pathological locations and clinical phenotypes. Thus, "*oxidative stress/mitochondrial dysfunction*," "*ubiquitin-proteasomal*

pathway" *and* "*protein metabolism pathway*" are all known to associate with AD, PD and HD. The commonality of CNS diseases provides a great opportunity to explore new indications and purposes of existing medicines in mechanisms associated to disease [19, 20]. Repositioning (or repurposing) existing drugs for new indications could reduce drug development risks and increase productivity [17, 21]. The strategy shows importance in developing therapies for CNS rare diseases, where the pharmaceutical industry research and development (R&D) risk keeps increasing due to increasing R&D costs [18, 21, 22]. Drug repositioning has gained monument recently and several repositioning approaches developed through systems biology approaches have been reported [20, 22]. It was proposed that: "*if two diseases share similar therapies, then other drugs that are currently used for only one of the two may also be therapeutic*" [23]. This statement could be risky without knowing the underlying disease mechanisms. Mechanism-focused repositioning strategies may constitute a more rational approach compared with structural and/or off-label prescriptions for FDA-approved drugs based approaches [22, 23].

2 Materials

2.1 Network Biology: Impact and Approach

Systems biology aims to understand complex behaviors of biological systems by investigating relationships among biological entities. Systems biology comprehensively integrates knowledge of multiple biological components, and of multiple functional perspectives with statistical, computational and mathematical analyses. This provides an analytical modeling capability from experiment design, data production, in silico modeling to biological discovery. In systems biology, the network concept is used to model a biological system. The nodes in the biological network denote all kinds of biological components. The edges characterize the relationships between two biological components (i.e. nodes) [10]. Numerous network-centric analytic approaches are being developed to analyze ever-increasing system-scale biological data (e.g. "omics" high-throughput data). These methods range from topological analyses to quantitative modeling. They successfully uncover many essential features of biological networks, as scale free, modularity and feedback effects, which shed light on regulation mechanisms and design principles of complex biological systems [10].

Furthermore, with biological network modeling philosophy, diseases are modeled in a *disease network* context [24]. Thus, in an effort to understand breast cancer, Pujana and colleagues reconstructed a cancer-associated molecular network with 118 genes connected by 866 potential functional associations, by integrating multiple "omics" data including co-expression and genetic interactions, and identified HMMR, a mobility receptor associated with

breast cancer susceptibility, interacting with the BRCA1 gene [25]. Other successful applications of network modeling have been documented in different types of cancers [26], pulmonary disease [27], AD [28], hepatitis [29], cardiovascular diseases [30, 31], asthma [32] and liver diseases [33]. These observations formulated a concept of "*disease module*." Biological components are non-randomly distributed, but act in concert in cellular networks to mediate cellular function or disorder contributing to a certain disease phenotype. The identification of disease related, "disease modules," would facilitate prediction of disease related genes, disclosing unknown functional implications of biological components in diseases and therefore uncover disease molecular mechanism [24]. Furthermore, the concept of *human disease network/human diseasome* was introduced: the nodes in the network represent human diseases and the edges denote disease-disease relationships [6].

2.2 Application of Network Approach to Pharmacology

The disease network modeling approach paves a new way for the transition from single-gene-based to mechanism-based pharmacology. Firstly, in a molecular interaction network context, it can help to analyze drugs and drug targets at a "system level," and therefore theoretically explore novel drug targets according to network theories, and characteristics of biological networks such as scale free, modularity and others [12]. In CNS disease studies, several promising applications of network systems biology in exploring candidate drug targets have been shown [34, 35].

Secondly, network systems biology can help in the design of drug combinations [36, 37]. Combination therapeutic designs can, in principle, target multiple molecular sites in a disease module and introduce synergistic effects [38, 39]. Thus, four key genes (ECG-R, COX2, MMP1 and MMP2) were identified implicated in triggering breast cancer metastasis through a protein interaction network modeling. Furthermore, a combinatorial design of drugs targeting the genes showed a promising effect on suppressing the growth of primary tumors [38].

Thirdly, network systems biology equips drug repositioning/repurposing with a solid theory foundation. The theory of drug repositioning is based on: (a) different diseases may share common molecular mechanism(s); (b) a drug may have multiple molecular targets. The concept of drug repositioning is not novel, but unfortunately most of successful findings of drug repositioning came from serendipity [40]. The occurrence and application of network systems biology is rapidly changing this. On the basis of the aforementioned approaches, network systems biology can be expected to play a crucial role in drug repositioning, with progressive incorporation of new rational exercises and strategies.

Pharmaceutical companies are increasingly developing new research pipelines for repositioning purposes, such as Pfizer, Roche, Merck, Eli Lilly and Novartis [41]. Academia are also

making great efforts on proposing computational and synthetic analytical frameworks [42–45], and on developing public available data resources for drug repurposing [46, 47]. For example, a network-based approach was developed to predict novel binding targets of existing drugs with a drug-target association network and to identify potential candidates of drug repositioning based on respective ligand structural similarities [48]. By combining multiple levels data sources, the PROMISCUOUS database [46] and Ondex data integration platform [49] are devoted to a systematic discovery of drug repositioning.

Network systems biology are also making an impact on understanding drugs side effects and predicting potential adverse events of drug candidates. Network systems biology allows for predicting drug off-targets by screening drug-protein interaction maps and databases. This unveils drug functional activities in cellular networks and tissue contexts, and can include patient-specific genomic variations, which highly influence drug specificity, efficiency and toxicology in individuals. Using this philosophy, several studies have performed network-based approach to systematically investigate adverse effects [50–53]. For example, Bourne's lab built a protein ligand binding network to predict drug off-targets and study unknown molecular mechanisms underlying pathophysiology of hypertension and side effects induced by Torcetrpib, a cholesteryl ester transfer protein (CETP) inhibitor used to treat high cholesterol levels [53].

3 Methods

We now outline a practical guide consisting of data mining, integration, analytic inference and visualization to illustrate the application of network-based method to drug discovery. It presents: (1) A multi-level integrative network modeling and, (2) Interpretation of the network in disease and pharmaceutical context. The pipeline shows a proof of concept understanding disease mechanisms and predicting drug repositioning specific to CNS common and rare diseases.

3.1 Step 1: Building and Visualizing a Multi-level Integrative Network

Based on existing systems biology resources, the pipeline presents an analytic flow: (1) data collection; (2) network integration and inference; (3) results interpretation and visualization. As a proof of concept, we focus on a total of 47 diseases related to CNS, psychiatry, and psychology. The list includes diseases with enormous unmet medical need such as AD, PA, HD, and rare/orphan CNS diseases such as ALS, Friedreich's ataxia, muscular dystrophy and spinocerebellar ataxias. Some psychiatric diseases such as autism, schizophrenia and depression are also included because,

Table 1
Systems biology resources applied in this work

Systems biology resources	Web address	Content applied in the study	Description
MetaCore	http://www.genego.com/metacore.php	Pathway, potency compound	Platform for data mining and pathway analysis (license required)
Biocarta	http://www.biocarta.com	Pathway	Pathways providing organic and biochemical products
GenMapp	http://www.genmapp.org/	Pathway	Maps representing pathways and groupings of genes
Pipeline, information	http://sites.informahealthcare.com/pipeline/	FDA approved drug and experimental medicine	Drug intelligence service (license required)
R Language	http://www.r-project.org/	Statistical inference	The R project for statistical computing
Mesh	http://www.nlm.nih.gov/mesh/	Disease ontology	Medical subject headings (Mesh)
Cytoscape	http://www.cytoscape.org/	Visualization platform	Open source platform. Network analysis and visualization
Partek genomics suite	http://www.partek.com/partekgs	Visualization platform	Statistics and visualization platform for omics data

Detailed table showing relevant resources (e.g. MeSH. MetaCore, Pipeline) used to collect different levels of data, including disease-related genes, pathways, FDA-approved drugs and late stage compounds

although they are mental and behavior disorders, their underlying mechanisms might be similar to those of CNS diseases. In the data collection step, we used MeSH. MetaCore, Pipeline and other systems biology resources (Table 1) to collect different levels of data, including disease-related genes, pathways, FDA-approved drugs and late stage compounds. In the next step, we built an integrative CNS-disease-related network based on the approaches described by Li and Agarwal [5] and Feng et al. [54]. In the network, disease nodes will also link to medicine nodes which represent drug/experimental medicines associated to the diseases (Table 1). Gene and pathway nodes (called gene or pathway clusters, respectively) represent all disease genes or pathways commonly shared by two associated diseases. A compound node denotes a cluster of potent compounds associated with disease targets. In the visualziation step, an application interface was developed based on Cytoscape 2.8 [55] (Table 1) to visualize and interpret the results (Fig. 1).

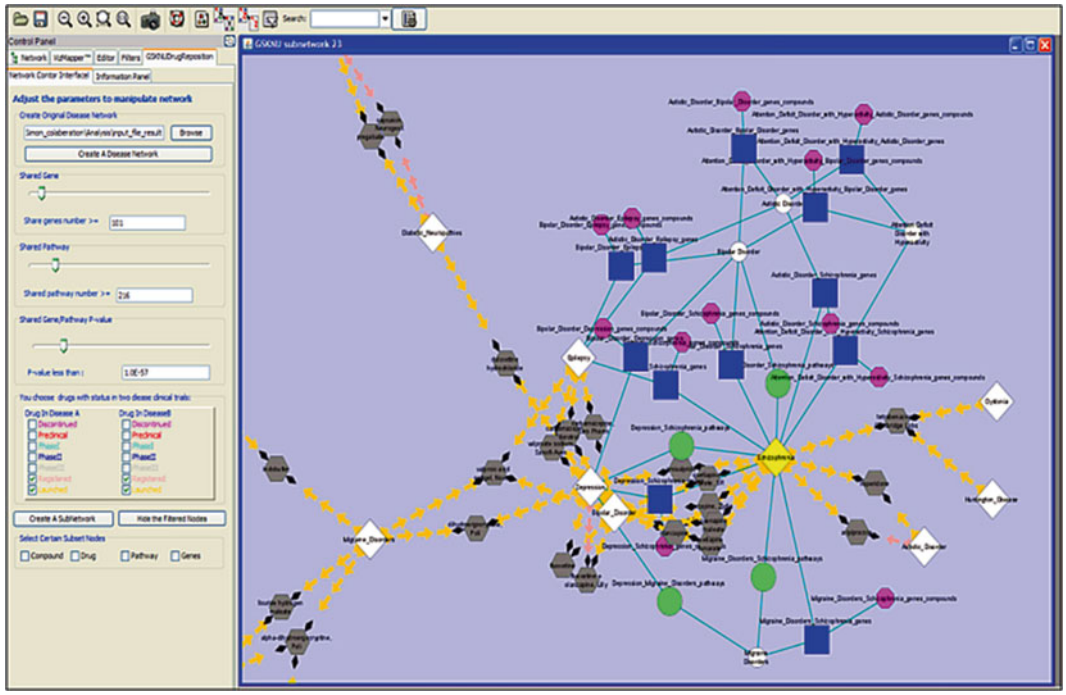


Fig. 1 CNS disease network. Case example. The figure shows the structure of the CNS multiple level network (e.g. gene nodes, pathway nodes, compound nodes and edges) based on different controlling parameters including shared gene number, shared pathways number, and disease association with *p*-values. A Cytoscape plug-in software was developed to perform this visualization [5, 54, 55]

3.2 Step 2: Understanding the Disease Mechanism and Repositioning Opportunities for CNS Rare Diseases Using the Integrative Network

After network and hierarchical clustering analysis, the results (Fig. 2 and Table 2) show several disease and pathway clusters among 632 pathways associated with 47 CNS diseases, which provides a global view of CNS diseases in a systematic context. Based on the hypothesis that related diseases may share similar modules/pathways, we suggest: (1) disease indications clustered in the results may share common medicines; (2) mechanism differentiation across diseases may offer opportunity to distinguish diseases at a pathway level; (3) It could open the way to rational therapy development for complex diseases through targeting multiple mechanisms.

In the results (Fig. 2 and Table 2), several disease clusters can be highlighted, where diseases share a set of enriched pathways. For instance, PD appears clustered with restless legs syndrome, basal ganglia disease, progress supranuclear palsy, multiple system atrophy and Friedreich's ataxia. The cluster is enriched in 35 pathways such as "*Proteolysis. Ubiquitin-proteasomal proteolysis*," "*Amino acid metabolism. Tryptophan/phenylalanine*," "*Methionine metabolism*," "*Transport synaptic vesicle exocytosis*," "*Development neurogenesis: synaptogenesis*," "*Stress induction of HSP regulation*,"

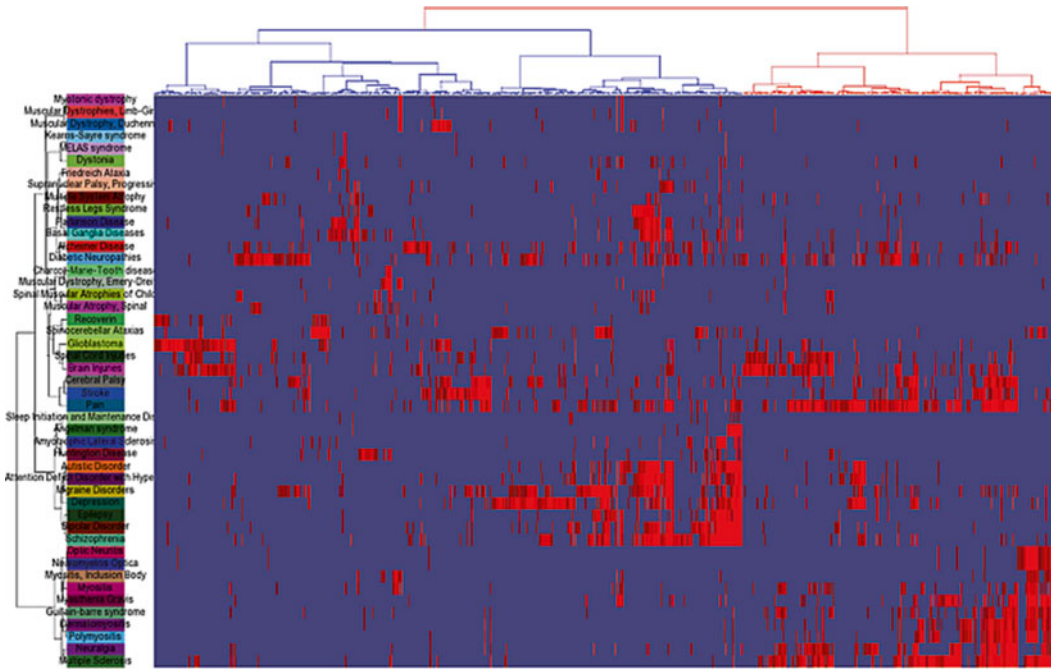


Fig. 2 A heat map showing hierarchical clustering for 47 CNS diseases and 632 disease modules. A heat map, generated using Partek Genomics Suite, shows the hierarchical clustering result for the 47 CNS diseases (*y* axis) and 632 disease modules (*x* axis) studied. Each disease cluster has a unique mechanistic profile. This heat map provides a global view of CNS diseases and shared pathways

among others, which are known to be relevant to PD [17, 18]. However, the mechanistic results for the other diseases in the same cluster are rarely documented. Therefore, the results show potential to explore molecular mechanisms and drug repositioning opportunities based on information on the shared pathways. For example, rotigotine, an FDA-approved drug for treating both PD and restless legs syndrome (Table 2a), may support this hypothesis (please refer to Fig. 2 and Table 2 for other examples). Disease pairs not appearing in the same cluster also provide potential repositioning opportunities. For instance, PD and epilepsy are shown in different clusters, but a medicine, safinamide (Merck KGaA) for PD (currently in phase III) appears linked also to epilepsy and restless legs syndrome (currently in Phase II). Safinamide functions as a monoamine oxidase B (MAOb) inhibitor, and MAOb also plays a key role in common pathways shared between PD and epilepsy such as "*Gamma-aminobutyrate (GABA) biosynthesis and metabolism*," "*Histidine-glutamate-glutamine and proline metabolism*," *and* "*Calcium transport.*" On the other hand, clustered CNS diseases may have their own unique mechanisms, showing the differences between diseases with similar symptoms. We find PD modules include for example: "*Oxidative*

Table 2
Pairs of CNS diseases connected in the network

(*A*)

CNS disease 1	*CNS disease 2*	*FDA approved drug*	*Drug stage*
Alzheimer's disease	Brain injuries	Cerebrolysin	Launched
Alzheimer's disease	Migraine disorders	Alpha-dihydroergocryptine	Launched
Alzheimer's disease	Parkinson's disease	Alpha-dihydroergocryptine	Launched
Amyotrophic lateral sclerosis	Huntington's disease	Riluzole	Launched
Autistic disorder	Schizophrenia	Risperidone	Launched
Bipolar disorder	Epilepsy	Carbamazepine, Microtrol	Launched
Bipolar disorder	Epilepsy	Valproic acid softgel	Launched
Bipolar disorder	Epilepsy	Valproate sodium	Launched
Bipolar disorder	Epilepsy	Carbamazepine	Launched
Bipolar disorder	Depression	Fluoxetine	Launched
Bipolar disorder	Depression	Olanzapine, Zydis	Launched
Bipolar disorder	Migraine disorders	Valproic acid softgel	Launched
Bipolar disorder	Schizophrenia	Olanzapine, Zydis	Launched
Bipolar disorder	Schizophrenia	Quetiapine fumarate, SR	Launched
Bipolar disorder	Schizophrenia	Quetiapine fumarate	Launched
Bipolar disorder	Schizophrenia	Olanzapine	Launched
Bipolar disorder	Schizophrenia	Asenapine fumarate	Launched
Depression	Diabetic neuropathies	Duloxetine hydrochloride	Launched
Depression	Schizophrenia	Amisulpride	Launched
Depression	Schizophrenia	Olanzapine	Launched
Depression	Migraine disorders	Dihydroergocryptine	Launched
Diabetic neuropathies	Neuralgia	Pregabalin	Launched
Dystonia	Schizophrenia	Tetrabenazine	Launched
Epilepsy	Migraine disorders	Valproic acid softgel	Launched
Epilepsy	Sleep initiation and maintenance disorders	Sodium phenobarbitone, SSP; chloral hydrate, SSP	Launched
Huntington's disease	Schizophrenia	Tetrabenazine	Launched
Migraine disorders	Stroke	Indobufen	Launched
Migraine disorders	Parkinson's disease	Alpha-dihydroergocryptine	Launched
Migraine disorders	Parkinson's disease	Lisuride hydrogen maleate	Launched

Parkinson's disease	Restless legs syndrome	Pramipexole	Launched		
Parkinson's disease	Restless legs syndrome	Ropinirole hydrochloride	Launched		
Parkinson's disease	Restless legs syndrome	Rotigotine	Launched		
Spinal cord injuries	Stroke	Monosialoganglioside GM-1, Amar	Launched		

(*B*)

CNS disease 1	*FDA approved drug for disease 1*	*Drug stage*	*CNS disease 2*	*FDA approved drug for disease 2*	*Medicine stage*
Epilepsy	Gabapentin	Launched	Amyotrophic lateral sclerosis	No	Discontinued
Attention deficit disorder with hyperactivity	Atomoxetine	Launched	Depression	No	Discontinued
Epilepsy	Oxcarbazepine	Launched	Restless legs syndrome	No	No development reported
Myasthenia gravis	Tacrolimus	Launched	Stroke	No	No development reported
Parkinson's disease	Pramipexole	Launched	Schizophrenia	No	No development reported
Restless legs syndrome	Pramipexole	Launched	Schizophrenia	No	No development reported
Schizophrenia	Paliperidone	Launched	Sleep initiation and maintenance disorders	No	No development reported
Alzheimer's disease	Donepezil hydrochloride	Launched	Attention deficit disorder with hyperactivity	No	Phase II
Alzheimer's disease	Memantine hydrochloride	Launched	Autistic disorder	No	Phase II
Alzheimer's disease	Memantine hydrochloride	Launched	Neuralgia	No	Phase II
Alzheimer's disease	Donepezil hydrochloride	Launched	Schizophrenia	No	Phase II
Alzheimer's disease	Memantine hydrochloride	Launched	Diabetic neuropathies	No	Phase III
Diabetic neuropathies	Pregabalin	Launched	Restless legs syndrome	No	Phase III
Multiple sclerosis	Nabiximols	Launched	Spinal cord injuries	No	Phase III
Neuralgia	Nabiximols	Launched	Spinal cord injuries	No	Phase III
Bipolar disorder	Olanzapine	Launched	Alzheimer's disease	No	No development reported
Schizophrenia	Olanzapine	Launched	Alzheimer's disease	No	No development reported

(continued)

Table 2
(continued)

Autistic disorder	Risperidone	Launched	Attention deficit disorder with hyperactivity	No	No development reported
Schizophrenia	Risperidone	Launched	Attention deficit disorder with hyperactivity	No	No development reported
Parkinson's Disease	L-dihydroxyphenylserine	Launched	Attention deficit disorder with hyperactivity	No	Phase II
Schizophrenia	Aripiprazole	Launched	Attention deficit disorder with hyperactivity	No	Phase II
Parkinson's Disease	Zonisamide	Launched	Bipolar disorder	No	Phase II
Stroke	Ibudilast	Launched	Multiple sclerosis	No	Phase II
Attention deficit disorder with hyperactivity	Atomoxetine	Launched	Alzheimer's disease	No	Phase III
Stroke	Atorvastatin	Launched	Alzheimer's disease	No	Phase III
Parkinson's disease	Pramipexole	Launched	Depression	No	Phase III
Restless legs syndrome	Pramipexole	Launched	Depression	No	Phase III
Multiple sclerosis	Nabiximols	Launched	Diabetic neuropathies	No	Phase III
Stroke	Atorvastatin	Launched	Multiple sclerosis	No	Phase III

FDA-approved drugs and new trials. (A) Pairs of CNS diseases sharing the same FDA-approved drug. (B) Disease pairs in the network of which one disease has a FDA-approved drug while the other has only medicines at clinical stage

phosphorylation," "*Acetaminophen metabolism*," "*Axonal guidance" and* "*Glutathione metabolism*," while restless legs syndrome includes different disease modules such as "*Nitric oxide signaling pathway*," "*Vitamins mediators and cofactors metabolism*," "*Nitric oxide biosynthesis and transport*." The differences in disease modules suggest that their etiology may differ.

Further analysis of the results also show that fifteen pairs of CNS related diseases share the same FDA-approved drug (Table 2a). For example, FDA-approved drug lisuride (Hoffmann-La Roche) is not only approved to prevent migraine attacks but also is an anti-Parkinson's disease drug (please refer to Fig. 2 and Table 2 for other examples).

For over 200 disease pairs in the network, one disease in the pairs has an FDA-approved drug while the other has no marketed therapy or experimental medicine. This could be a repositioning opportunity for pharmaceutical companies. For 29 disease pairs, one disease has an FDA-approved drug while the other has a medicine in clinical trials (Table 2b). For example, zonisamide (Elan) has been approved for PD and it is in phase II trial for bipolar and migraine disorder (Table 2b). Another example, gabapentin (Pfizer), has been approved for epilepsy, but it is reported as "discontinued" for amyotrophic lateral sclerosis. Gabapentin is an agonist for GABA receptor and calcium channel; Epilepsy and ALS share 21 pathways including "*Beta-alanine metabolism and transport*," "*Glycine pathways and transport" and* "*Calcium/manganese/sodium transport*." It is understandable that a therapy for epilepsy would possibly work for ALS as well. The details for the "discontinued" status are unknown. It could be due to disease specific efficiency, compound toxicity, business or marketing reasons. Instead of discoveries by "serendipitous observation," rational repositioning design through network systems biology constitutes a more solid, mechanistic-based approach. As a top challenge for repositioning, we need to carefully define the criteria to select the alternative indications towards new therapies. It might be the most important factor to thoroughly understand alternative disease treatments [56]. We believe our network approaches can overcome these difficulties towards new discoveries and relevant breakthroughs.

4 Notes

Network systems biology is gaining momentum in human health research, but it is confronted with the need for careful integration, analysis and subsequent visualization of all kinds of available data, with enormous size and dimension, and abundance of information produced by a variety of high throughput platforms, biological experiments, clinical trials, and computational inference studies. It

is obviously challenging, because of the heterogeneity of knowledge contained in datasets produced for many different purposes, presented in different formats, and managed in disparate locations. Therefore, devising integrative, knowledge-oriented analytical frameworks is a prerequisite for developing system-based pharmacology. We herein reviewed current advances on knowledge integration, mining and inference, and discussed a drug repositioning analysis focusing on CNS diseases, taking advantage of these techniques.

Efforts have yet to be made towards integrating biomedical data and knowledge in standardized, curated databases, with standard analytic software tools and analytical methods: (1) Various standard data formats are proposed to facilitate exchange, query and management between diverse data sources, including Open Biological Ontology (OBO) foundry initiated to create a controlled vocabulary community, to integrate diverse biological and medical information [57]; model information standards to represent network modeling information such as the Systems Biology Markup Language (SBML) [58]; and the Ontology Web Language (OWL) and Resource Description Framework (RDF) Based on Semantic Web technologies released to enable computational systems to manage and communicate pertinent biological information in internet environment (please refer to review [59] for details); (2) Computational biologists to develop software platforms [60] able to integrate multiple-purpose applications and visualization tools such as Cytoscape [61], which provide highly flexible environment to integrate a wide range of functional modules and powerful visualization interface; (3) Analytic approaches based on statistics, machine learning and computational simulation to address integration hurdles and interpret experimental data that would increase our understanding of cellular systems, diseases, and drug function. For instance, Shao's lab recently developed statistical frames to analyze human disease related genes by integrating global phenotype and protein interaction data, and to predict drug targets by combining similarities of drug phenotypes and drug chemical structures extracted from FDA and DrugBank [62–64]. Also, disease-drug associations have been statistically modeled by integrating gene expression and protein interaction data obtained from NCBI Gene Expression Omnibus (GEO) and Human Protein Reference Database (HPRD) [8, 65].

Systems biology has been proven to have a significant role in drug discovery [27]. The traditional "one target for one disease" approach has resulted in a high failure rate for complex diseases [14, 66], while pathway oriented therapeutic development including repositioning begins to show its promise [56, 67]. Our pipeline exemplifies its value in understanding CNS diseases mechanisms and drug repositioning opportunities. Future endeavors to develop disease modifying cures for neurodegenerative and neuroinflammatory

diseases can begin to identify targets in multiple core mechanisms, or investigate drug combinations [3]. The availability of "omics" data, computational algorithms and software platforms applied to comprehensive systems biology approaches will continue to play a critical role in pharmacology [12].

Acknowledgments

We gratefully acknowledge Pan Du, Chenbing Guan, Yong Li and Peter Woollard for sharing the dataset and scientific insights with us. Also thanks to Minhua Zhang to help us with the manuscript.

References

1. Yang Y, Adelstein SJ, Kassis AI (2009) Target discovery from data mining approaches. Drug Discov Today 14:147–154
2. Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–715
3. Azmi AS, Wang Z, Philip PA et al (2010) Proof of concept: network and systems biology approaches aid in the discovery of potent anticancer drug combinations. Mol Cancer Ther 9:3137–3144
4. Stein A, Pache RA, Bernado P et al (2009) Dynamic interactions of proteins in complex networks: a more structured view. FEBS J 276:5390–5405
5. Li Y, Agarwal P (2009) A pathway-based view of human diseases and disease relationships. PLoS One 4:e4346
6. Goh KI, Cusick ME, Valle D et al (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690
7. Loscalzo J, Kohane I, Barabasi AL (2007) Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. Mol Syst Biol 3:124
8. Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. PLoS One 4:e6536
9. Stegmaier P, Krull M, Voss N et al (2010) Molecular mechanistic associations of human diseases. BMC Syst Biol 4:124
10. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. Cell 144:986–998
11. Hidalgo CA, Blumm N, Barabasi AL, Christakis NA (2009) A dynamic network approach for the study of human phenotypes. PLoS Comput Biol 5:e1000353
12. Arrell DK, Terzic A (2010) Network systems biology for drug discovery. Clin Pharmacol Ther 88:120–125
13. Finkbeiner S (2010) Bridging the valley of death of therapeutics for neurodegeneration. Nat Med 16:1227–1232
14. Flordellis CS, Manolis AS, Paris H, Karabinis A (2006) Rethinking target discovery in polygenic diseases. Curr Top Med Chem 6: 1791–1798
15. Brunden KR, Trojanowski JQ, Lee VM (2009) Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies. Nat Rev Drug Discov 8:783–793
16. Decision Resources (2010) Target product profiles 2009: physician insights on key attributes: Parkinson's disease (report). Chapter 1. Etiology and pathophysiology. Decision Resources Inc http://decisionresources.com/Products-and-Services/; http://www.businesswire.com/news/home/20100311005839/en/Research-Markets-Target-Product-Profiles-2009-Physician#.U77e0rFHSf0
17. Simunovic F, Yi M, Wang Y et al (2009) Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. Brain 132:1795–1809
18. Zhang X, Zhou JY, Chin MH et al (2010) Region-specific protein abundance changes in the brain of MPTP-induced Parkinson's disease mouse model. J Proteome Res 9: 1496–1509
19. Skovronsky DM, Lee VM, Trojanowski JQ (2006) Neurodegenerative diseases: new concepts of pathogenesis and their therapeutic implications. Annu Rev Pathol 1:151–170
20. Qu XA, Gudivada RC, Jegga AG et al (2009) Inferring novel disease indications for known

drugs by semantically linking drug action and disease mechanism relationships. BMC Bioinformatics 10(Suppl 5):S4

21. Haupt VJ, Schroeder M (2011) Old friends in new guise: repositioning of known drugs with structural bioinformatics. Brief Bioinform 12: 312–326
22. Ha S, Seo YJ, Kwon MS et al (2008) IDMap: facilitating the detection of potential leads with therapeutic targets. Bioinformatics 24: 1413–1415
23. Chiang AP, Butte AJ (2009) Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. Clin Pharmacol Ther 86:507–510
24. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12:56–68
25. Pujana MA, Han JD, Starita LM et al (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 39:1338–1349
26. Omenn GS (2011) An omics perspective on cancer research. Expert Rev Proteomics 8: 147–148
27. Auffray C, Adcock IM, Chung KF et al (2010) An integrative systems biology approach to understanding pulmonary diseases. Chest 137:1410–1416
28. Soler-Lopez M, Zanzoni A, Lluis R et al (2011) Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. Genome Res 21:364–376
29. de Chassey B, Navratil V, Tafforeau L et al (2008) Hepatitis C virus infection protein network. Mol Syst Biol 4:230
30. Diez D, Wheelock AM, Goto S et al (2010) The use of network analyses for elucidating mechanisms in cardiovascular disease. Mol Biosyst 6:289–304
31. Tanaka T, Oka T, Shimada Y et al (2008) Pharmacogenomics of cardiovascular pharmacology: pharmacogenomic network of cardiovascular disease models. J Pharmacol Sci 107:8–14
32. Hwang S, Son SW, Kim SC et al (2008) A protein interaction network associated with asthma. J Theor Biol 252:722–731
33. Fedeles SV, Tian X, Gallagher AR et al (2011) A genetic interaction network of five genes for human polycystic kidney and liver diseases defines polycystin-1 as the central determinant of cyst formation. Nat Genet 43:639–647
34. Kaimal V, Sardana D, Bardes EE et al (2011) Integrative systems biology approaches to identify and prioritize disease and drug candidate genes. Methods Mol Biol 700:241–259
35. Elmer GI, Kafkafi N (2009) Drug discovery in psychiatric illness: mining for gold. Schizophr Bull 35:287–292
36. Pujol A, Mosca R, Farres J, Aloy P (2010) Unveiling the role of network and systems biology in drug discovery. Trends Pharmacol Sci 31:115–123
37. Wu Z, Zhao XM, Chen L (2010) A systems biology approach to identify effective cocktail drugs. BMC Syst Biol 4(Suppl 2):S7
38. Fliri AF, Loging WT, Volkmann RA (2010) Cause-effect relationships in medicine: a protein network perspective. Trends Pharmacol Sci 31:547–555
39. Geva-Zatorsky N, Dekel E, Cohen AA et al (2010) Protein dynamics in drug combinations: a linear superposition of individual-drug responses. Cell 140:643–651
40. Sardana D, Zhu C, Zhang M et al (2011) Drug repositioning for orphan diseases. Brief Bioinform 12:346–356
41. Dimond PF (2010) Drug repositioning gains in popularity. Gen Eng Biotechnol News 30; http://www.genengnews.com/gen-articles/drug-repositioning-gains-in-popularity/3263/
42. Deftereos SN, Andronis C, Friedla EJ, Persidis A (2011) Drug repurposing and adverse event prediction using high-throughput literature analysis. Wiley Interdiscip Rev Syst Biol Med 3:323–334
43. Iorio F, Bosotti R, Scacheri E et al (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci U S A 107:14621–14626
44. Kotelnikova E, Yuryev A, Mazo I, Daraselia N (2010) Computational approaches for drug repositioning and combination therapy design. J Bioinform Comput Biol 8:593–606
45. Dubus E, Ijjaali I, Barberan O, Petitet F (2009) Drug repositioning using in silico compound profiling. Future Med Chem 1:1723–1736
46. von Eichborn J, Murgueitio MS, Dunkel M et al (2011) PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res 39:D1060–D1066
47. Luo H, Chen J, Shi L et al (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. Nucleic Acids Res 39(Web Server issue):W492–W498
48. Keiser MJ, Setola V, Irwin JJ et al (2009) Predicting new molecular targets for known drugs. Nature 462:175–181
49. Cockell SJ, Weile J, Lord P et al (2010) An integrated dataset for in silico drug discovery. J Integr Bioinform 7:116
50. Wallach I, Jaitly N, Lilien R (2010) A structure-based approach for mapping adverse

drug reactions to the perturbation of underlying biological pathways. PLoS One 5:e12063

51. Yang L, Chen J, He L (2009) Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. PLoS Comput Biol 5:e1000441
52. Campillos M, Kuhn M, Gavin AC et al (2008) Drug target identification using side-effect similarity. Science 321:263–266
53. Xie L, Evangelidis T, Bourne PE (2011) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. PLoS Comput Biol 7:e1002037
54. Feng G, Du P, Krett NL et al (2010) A collection of bioconductor methods to visualize gene-list annotations. BMC Res Notes 3:10
55. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
56. Capell K (2009) The FDA O.K.'s Novartis Ilaris. Business Week. http://www.businessweek.com/bwdaily/dnflash/content/jun2009/db20090618_060721.htm
57. Smith B, Ashburner M, Rosse C et al (2007) The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255
58. Brazma A, Krestyaninova M, Sarkans U (2006) Standards for systems biology. Nat Rev Genet 7:593–605
59. Antezana E, Kuiper M, Mironov V (2009) Biological knowledge management: the emerging role of the Semantic Web technologies. Brief Bioinform 10:392–407
60. Klipp E, Liebermeister W, Helbig A et al (2007) Systems biology standards – the community speaks. Nat Biotechnol 25:390–391
61. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504
62. Wu X, Jiang R, Zhang MQ, Li S (2008) Network-based global inference of human disease genes. Mol Syst Biol 4:189
63. Zhao S, Li S (2010) Network-based relating pharmacological and genomic spaces for drug target identification. PLoS One 5: e11764
64. Yao X, Hao H, Li Y, Li S (2011) Modularity-based credible prediction of disease genes and detection of disease subtypes on the phenotype-gene heterogeneous network. BMC Syst Biol 5:79
65. Suthram S, Dudley JT, Chiang AP et al (2010) Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 6:e1000662
66. Mullard A (2011) 2010 FDA drug approvals. Nat Rev Drug Discov 10:82–85
67. Rajasethupathy P, Vayttaden SJ, Bhalla US (2005) Systems modeling: a pathway to drug discovery. Curr Opin Chem Biol 9:400–406

INDEX

A

Juan I. Castrillo and Stephen G. Oliver (eds.), *Systems Biology of Alzheimer's Disease*, Methods in Molecular Biology, vol. 1303, DOI 10.1007/978-1-4939-2627-5, © Springer Science+Business Media New York 2016

B

C

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

Y