Andrew Emili · Jack Greenblatt
Shoshana Wodak *Editors*

# Systems Analysis of Chromatin-Related Protein Complexes in Cancer

Systems Analysis of Chromatin-Related Protein Complexes in Cancer

Andrew Emili • Jack Greenblatt
Shoshana Wodak
Editors

# Systems Analysis of Chromatin-Related Protein Complexes in Cancer

*Editors*
Andrew Emili
Donnelly Centre for Cellular and Biomolecular Research
Banting and Best Department of Medical Research
University of Toronto
Toronto, ON, Canada

Jack Greenblatt
Terrence Donnelly Centre for Cellular and Biomolecular Research
University of Toronto
Toronto, ON, Canada

Shoshana Wodak
Molecular Structure and Function Program
The Hospital for Sick Children
Toronto, ON, Canada

ISBN 978-1-4614-7930-7 ISBN 978-1-4614-7931-4 (eBook)
DOI 10.1007/978-1-4614-7931-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013944686

© Springer Science+Business Media New York 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

The chromatin fibres packaging our genes consist of the DNA double helix wrapped around histone proteins to form nucleosomes. Chromatin structure and function are modified by a highly specialized protein machinery that is responsible for epigenetic memory, which is defined as changes in gene accessibility and activity states that are stably inherited yet do not involve alterations in the DNA sequence of an organism. These include reversible covalent histone modifications generated enzymatically, for example site-specific methylation and demethylation marks, in a manner that alters the intrinsic stability of nucleosomes and compaction and accessibility of chromatin to the transcription apparatus. By creating a combinatorial "histone code", which is subsequently differentially recognized by effector chromatin protein complexes that physically associate with particular modified nucleosomes, chromatin can be dynamically altered to direct chromatin-based processes such as transcription and DNA repair in a locus-dependent contextual manner.

Most, if not all, chromatin modifying and remodelling enzymes are components of multi-subunit protein complexes, which often include physically associated non-coding RNAs. Controlled expression, assembly and activity of these chromatin protein complexes are essential for the proper execution of normal cellular behaviours, such as stem cell pluripotency or lineage-specific cellular differentiation, and when disrupted can cause the emergence of transformed phenotypes. Therefore, one key goal of the rapidly emerging field of epigenetics is to understand the molecular composition, biochemical regulation and physiological roles of these chromatin protein complexes, particularly in the accurate control of gene expression, DNA repair, genome stability, chromosome compaction and segregation. All of these are vital processes for normal human development, including stem cell renewal and the formation of cell types and tissues.

To elucidate the properties and biological roles of chromatin modifying complexes and the mechanisms underlying disruptions in chromosome structure and gene expression patterns that occur when chromatin protein interaction networks become perturbed in pathological states, the field of epigenetics draws on increasingly sophisticated analytical strategies. As a result, the technical expertise and

clinical experience needed to address the range and complexity of the biological problems involved can be overwhelming to the novice and expert alike, particularly as many scientists now entering the epigenetics field are trained in either traditional molecular biology, cancer biology, structural biology, drug discovery or high-throughput "omic" sciences, but are rarely familiar with all the relevant domains. This book was conceived to address this gap.

In this volume, leading international experts discuss recent progress in the application of both traditional and cutting-edge methods to explore the unique biology and complicated biochemistry of chromatin protein complexes, including the identification, functional evaluation and biomedical assessment of particular chromatin protein complexes in different epigenetic systems ranging from stem cell development to human cancer: *Bremner*, *Emili*, *Greenblatt* and *Wodak* review current knowledge of *chromatin protein networks and systems* in human; *Cagney*, *Coulombe*, *Figeys*, *Garcia*, *Moffat*, *Vermeulen* and *Washburn* describe *systematic interaction mapping efforts* aimed at documenting the networks of physical and functional interactions that occur among the components of the chromatin-related protein machinery; *Zhang* and *Mitchell* explore *global regulome studies* aimed at understanding the mechanistic aspects controlling chromatin states in normal and diseased states; *Copeland* and *Knapp* review progress in the *discovery and development of epigenetic drugs and small molecule chemical probes*; and, finally, *Stein* and *Wodak* illustrate how *disruption of chromatin protein complexes* is linked to common polygenic diseases, notably cancer, with major economic and social impact. Taken together, this collection of 16 invited chapters provides a holistic and authoritative overview into efforts to "crack" the epigenetic code, one of the most important challenges in biomedical research today.

Toronto, ON, Canada

Shoshana Wodak, Ph.D.
Jack Greenblatt, Ph.D.
Andrew Emili, Ph.D.

# Contents

**Networks of Histone Demethylases and Their Relevance to the Regulation of Chromatin Structure and Dynamics** ........................ 1
Edyta Marcon, Sandra Smiley, Andrei L. Turinsky, and Jack Greenblatt

**Histone Methyltransferase Complexes in Transcription, Development, and Cancer** ........................ 33
Jonathan B. Olsen, Jack Greenblatt, and Andrew Emili

**Chromatin Biology and Cancer Linked Through Protein–Protein Interactions** ........................ 49
Shuye Pu, Andrei L. Turinsky, and Shoshana Wodak

**Familial and Somatic Mutations of Histone-Modifying Enzymes in Cancer** ........................ 65
Chenyi Liu and Lincoln Stein

**ncRNA–Protein Interactions in Development and Disease from the Perspective of High-Throughput Studies** ........................ 87
Dorothy Yanling Zhao, Yue Li, Jack Greenblatt, and Zhaolei Zhang

**Identification of Chromatin-Binding Protein Complexes** ........................ 117
Alain Doucet and Daniel Figeys

**Exploring Chromatin Readers Using High-Accuracy Quantitative Mass Spectrometry-Based Proteomics** ........................ 133
Arne H. Smits and Michiel Vermeulen

**Proteomic Interrogation of Human Chromatin Protein States** ........................ 149
Natarajan V. Bhanu and Benjamin A. Garcia

**Quantitative Proteomics Characterization of Chromatin-Remodeling Complexes in Health and Disease** ........................ 177
Mahadevan Lakshminarasimhan and Michael P. Washburn

**Proteome Analysis of Chromatin Complexes in Differentiating Stem Cells** ........ 197
Ariane Watson and Gerard Cagney

**Structural Genomics and Drug Discovery for Chromatin-Related Protein Complexes Involved in Histone Tail Recognition** ........ 211
Panagis Filippakopoulos and Stefan Knapp

**Proteomic Analysis Reveals a Role for the GTPase RPAP4/GPN1 and the Cochaperone RPAP3 in Biogenesis of All Three Nuclear RNA Polymerases** ........ 227
Diane Forget, Philippe Cloutier, Céline Domecq, and Benoit Coulombe

**Targeting Chromatin Modifying Enzymes in Anticancer Drug Discovery** ........ 239
Victoria M. Richon, Mikel P. Moyer, and Robert A. Copeland

**Genome Organization in Cancer Cells** ........ 257
Harry Yu Zhou and Jennifer A. Mitchell

**Histone Chaperones, Epigenetics, and Cancer** ........ 277
Mina Rafiei and Rod Bremner

**Evaluating Chromatin Regulators in Cancer Cell Lines Through RNAi-Glasses** ........ 297
Troy Ketela, Kevin R. Brown, and Jason Moffat

**Index** ........ 309

# Contributors

**Natarajan V. Bhanu, Ph.D.** Department of Biochemistry and Biophysics, Epigenetics Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

**Rod Bremner, Ph.D.** Department of Ophthalmology and Vision Sciences, Toronto Western Hospital, Toronto, ON, Canada

**Kevin R. Brown, Ph.D.** Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Gerard Cagney, Ph.D., B.Sc.** University College Dublin, The Conway Institute, School of Biomolecular and Biomedical Science, Belfield, Dublin, Ireland

**Philippe Cloutier** Laboratory of Gene Transcription and Proteomics, Institut de recherches cliniques de Montréal, Montréal, QC, Canada

**Robert A. Copeland, Ph.D.** R&D Department, Epizyme, Inc., Cambridge, MA, USA

**Benoit Coulombe** Laboratory of Gene Transcription and Proteomics, Institut de recherches cliniques de Montréal, Montréal, QC, Canada

**Céline Domecq** Laboratory of Gene Transcription and Proteomics, Institut de recherches cliniques de Montréal, Montréal, QC, Canada

**Alain Doucet, Ph.D.** Department of Biochemistry Microbiology and Immunology, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

**Andrew Emili, Ph.D.** Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada

**Daniel Figeys, Ph.D.** Department of Biochemistry Microbiology and Immunology, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

**Panagis Filippakopoulos, Ph.D.** Nuffield Department of Clinical Medicine, Medicine Division, Oxford University, Oxfordshire, Oxford, UK

**Diane Forget** Laboratory of Gene Transcription and Proteomics, Institut de recherches cliniques de Montréal, Montréal, QC, Canada

**Benjamin A. Garcia, Ph.D.** Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

**Jack Greenblatt, Ph.D.** Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Troy Ketela, Ph.D.** Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Stefan Knapp, Ph.D.** Nuffield Department of Clinical Medicine, Medicine Division, Oxford University, Oxford, Oxfordshire, UK

**Mahadevan Lakshminarasimhan, Ph.D.** Stowers Institute for Medical Research, Kansas City, MO, USA

**Yue Li, B.Sc., M.Sc., Ph.D. Candidate** Department of Computer Science, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Chenyi Liu, M.Sc., Ph.D. Candidate** Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

**Edyta Marcon, Ph.D.** Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Jennifer A. Mitchell, Ph.D.** Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

**Jason Moffat, Ph.D.** Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

Banting and Best Department of Medical Research and Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

**Mikel P. Moyer, Ph.D.** R & D Department, Epizyme, Inc., Cambridge, MA, USA

**Jonathan B. Olsen, B.Sc., M.Sc.A** Department of Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Shuye Pu, M.D., M.Sc.** Molecular Structure and Function Program, The Hospital for Sick Children, Toronto, ON, Canada

**Mina Rafiei, M.Sc.** Laboratory of Medicine and Pathobiology, Genetics and Development Division, University of Toronto, Toronto, ON, Canada

**Victoria M. Richon, Ph.D.** Oncology Discovery and Preclinical Sciences, Global Oncology Division, Sanofi, Cambridge, MA, USA

**Sandra Smiley** Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Arne H. Smits, M.Sc.** Department of Molecular Cancer Research, University Medical Center Utrecht, Utrecht, The Netherlands

**Lincoln Stein, MD, Ph.D.** Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Informatics and Bio-Computing, Ontario Institute for Cancer Research, Toronto, ON, Canada

Bioinformatics and Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

**Andrei L. Turinsky, Ph.D., M.Sc.** Research Institute, Molecular Structure and Function Program, Hospital for Sick Children, Toronto, ON, Canada

**Michiel Vermeulen, Ph.D.** Department of Molecular Cancer Research, University Medical Center Utrecht, Utrecht, The Netherlands

**Michael P. Washburn, Ph.D.** Stowers Institute for Medical Research, Kansas City, MO, USA

**Ariane Watson, B.Sc.** University College Dublin, The Conway Institute, School of Biomolecular and Biomedical Science, Belfield, Dublin, Ireland

**Shoshana Wodak, Ph.D.** Molecular Structure and Function Program, The Hospital for Sick Children, Toronto, ON, Canada

**Dorothy Yanling Zhao, B.Sc., M.Sc., Ph.D. Candidate** Department of Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Zhaolei Zhang, B.Sc., Ph.D.** Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

**Harry Yu Zhou, B.Sc.** Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

# Networks of Histone Demethylases and Their Relevance to the Regulation of Chromatin Structure and Dynamics

**Edyta Marcon, Sandra Smiley, Andrei L. Turinsky, and Jack Greenblatt**

**Abstract** Regulation of chromatin structure and dynamics is crucial for gene expression, chromosome segregation, DNA replication, and DNA repair, effectively controlling all cellular processes. Such regulation is achieved by a multitude of chromatin modifying enzymes that cause changes in DNA accessibility. Modifications imposed by these enzymes include DNA methylation and various histone modifications. Often, an interplay of multiple mechanisms is necessary to properly regulate chromatin dynamics, leading to a defined functional outcome. Of the various histone modifications, methylation is the most complex and is reversed by the KDM1 and JMJC families of histone demethylases. Collectively, these enzymes can reverse all three histone methylation states, often acting on the same substrates, and yet having different functional outcomes. Thus far, substrates have been identified for 26/32 (80 %) of all known histone demethylases, but functional

E. Marcon, Ph.D. (✉)
Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Rm 930, 160 College Street, Toronto, ON, Canada, M5S 3E1
e-mail: edyta.marcon@utoronto.ca

S. Smiley
Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Rm 930, 160 College Street, Toronto, ON, Canada, M5S 3E1
e-mail: sandra.smiley@utoronto.ca

A.L. Turinsky, Ph.D, M.Sc.
Research Institute, Molecular Structure and Function Program, Hospital for Sick Children, Toronto, ON, Canada
e-mail: turinsky@sickkids.ca

J. Greenblatt, Ph.D.
Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Rm 940, Toronto, ON, Canada, M5S 3E1
e-mail: jack.greenblatt@utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_1,
© Springer Science+Business Media New York 2014

studies have lagged behind. Analysis of protein–protein interactions has greatly contributed to our understanding of the roles some of these proteins play in the regulation of chromatin structure and dynamics, sometimes explaining genetic associations previously established between aberrant expression of histone demethylases and certain human disorders. Here, we will discuss our current understanding of histone demethylases, emphasizing protein complexes, and their contributions to the function of histone demethylases, their connections to various human disorders and multiple types of cancer.

**Keywords** Histone demethylases • PHD (plant homeodomain) • Tudor • Chromo • JMJC (Jumonji C-terminal domain) • JMJN (Jumonji N-terminal domain) • Methylation • Demethylation • Histones

## Chromatin Structure and Function

The regulation of chromatin structure and dynamics serves as the foundation of epigenetics. Regions of open, more accessible chromatin and closed, more condensed chromatin are often interspersed and highly dynamic, allowing the local chromatin structure to be fine tuned in response to cell type, cell cycle stage, DNA damage, cellular signals, and external stimuli.

### *Regulation of Chromatin Conformation*

Chromatin regulation is controlled by two types of covalent modifications: DNA methylation and posttranslational modification of histones. The 5-methylcytosine mark imposed on specific CpG motifs in DNA is not inherently repressive; instead, these sites are recognized and bound by methyl-CpG-binding proteins, which recruit histone-modifying corepressor complexes to create closed chromatin and inhibit transcription.

Posttranslational histone modifications, primarily on their N-terminal tails, either directly affect interactions between histones and DNA or act as docking sites for other effector modules. For example, histone acetylation is mostly correlated with activation of gene expression, while histone methylation can be either activating or repressing, depending upon the context. To further complicate matters, combinations of marks on the same or adjacent histone tails create a "histone code" that fine tunes a precise chromatin state [1]. As well, histone variants that replace canonical histones in certain regions or situations also impact chromatin structure. Histone modifications and the exchange of histone variants are thought to be more dynamic than DNA methylation.

The human genome encodes large families of DNA methyltransferases and histone-modifying acetyltransferases, deacetylases, methylases, demethylases,

ubiquitinating and deubiquitinating enzymes, and kinases and phosphatases. Other protein families with over 200 members containing bromodomains, Tudor domains, chromodomains, methyl-binding domains, and/or PHD domains recognize particular patterns of histone modifications to recruit chromatin remodeling enzymes and influence chromatin function [2].

## *Studying Chromatin Dynamics*

The study of chromatin modifying enzymes has progressively moved from studying individual components towards elucidating protein complexes and pathways, as it became clear that protein partners play key roles in altering enzymatic activity, localization, and, consequently, protein function. Royer et al. [3] recently compared 13 datasets of yeast protein–protein interactions obtained through various methods (e.g., yeast two-hybrid (Y2H) and affinity purifications/mass spectrometry (AP–MS)) and concluded that AP–MS is a robust and highly accurate high-throughput method for the identification of protein complexes. Genome-wide AP–MS studies have already been performed for *E. coli* and yeast but are more difficult for mammals due to the larger genome size and technical difficulties [4–6]. Therefore, mammalian AP–MS studies have focused on subsets of proteins (e.g., mitotic complexes, deubiquitinating enzymes, and nuclear receptor coregulators), but even from this limited subset it is easy to see that studies of protein complexes have significantly contributed to the functional characterization of mammalian proteins [7–9].

Over the last few years, a project has been underway in the Greenblatt lab to define protein complexes containing known and bioinformatically predicted human chromatin-related proteins to further our understanding of epigenetic mechanisms [10–12]. Our ultimate goal is to produce a database of a comprehensive interactome of chromatin-associated proteins to facilitate further investigations of the regulation of chromatin structure, dynamics, and function. In this chapter, we focus mainly on histone demethylases in the regulation of chromatin function, emphasizing protein complexes and their roles in guiding various cellular processes, as well as their implications for human health and disease.

## Roles of Histone Demethylases in Chromatin Structure and Function

Histone methylation is mediated by a family of 59 histone methyltransferases described elsewhere in this book. Methylation is one of the most complex of histone marks and has been demonstrated to occur on specific arginine and lysine residues on three of the four canonical histones (H3, H4, and H2B) [13]. Arginine residues can be monomethylated or dimethylated symmetrically or asymmetrically, whereas

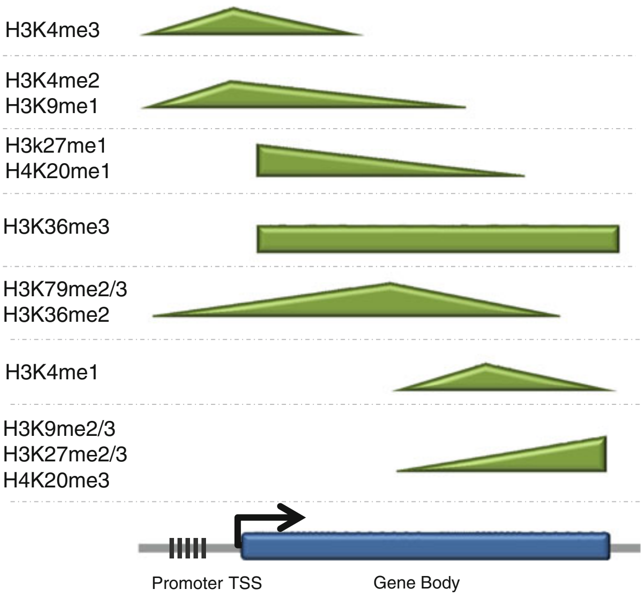



**Fig. 1** Histone methylation pattern most typically found on actively transcribed genes. H3K4me3, H3K4me2, and H3K9me1 are present in the promoter region, peaking at the transcription start site (TSS). H3K4me3 falls off quickly following the TSS, while H3K4me2 and H3K9me1 persist further towards the 3′ end of the gene. H3K27me1 and H4K20me1 are absent in the upstream (promoter) region, begin at the TSS and then tail off towards the 3′ end of the gene. H3K36me3 is also absent from the promoter region but is found along the downstream region transcribed by RNA polymerase II. H3K4me1, H3K9me2/3, H3K27me2/3, and H4K20me3 are all present in the 3′ end of the gene

lysine residues can accept up to three methyl groups, resulting in mono, di, or trimethylated states. Because methylation does not affect the charge on lysine or arginine residues, it does not alter chromatin structure directly, but instead, serves as a platform for the recruitment of other protein complexes. Methylation can facilitate either transcriptional activation or repression, depending on the cellular context and the specific nature of the modification. In general, methylation of histone 3 on lysine 4 (H3K4) or lysine 36 (H3K36) is a mark of active transcription, whereas methylation of either lysine 9 or lysine 27 (H3K9, H3K27) is associated with transcriptional silencing. These marks tend to be associated with specific genomic regions such as enhancers, transcription start sites, gene bodies, or 3′ UTRs (Fig. 1). Besides transcriptional control of gene expression, histone methylation is also involved in the regulation of other processes such as DNA replication and repair. For example, methylation of lysine 79 (H3K79) less studied than other methylation marks, is thought to have a role in the establishment of chromatin boundaries and DNA repair (Fig. 2) [14, 15].

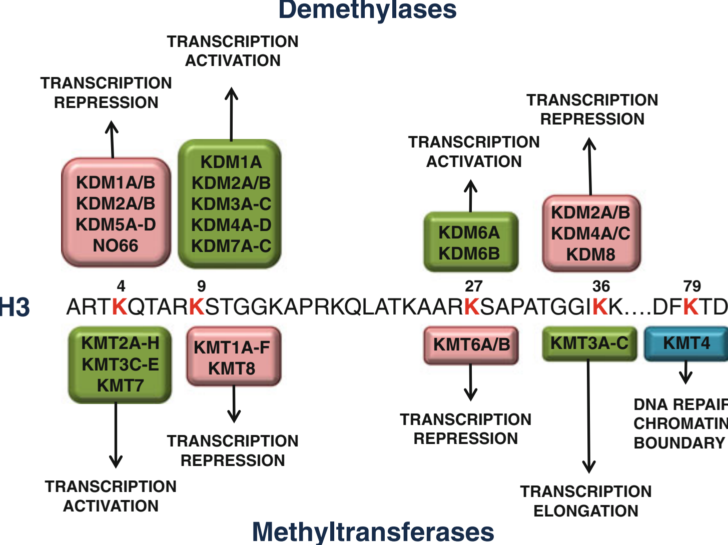



**Fig. 2** Methylation of lysine residues on the histone H3 tail can produce a variety of outcomes. Shown are the major methylated residues, the enzymes required to deposit or remove those marks and their corresponding cellular functions. Some KDMs can remove more than one type of histone mark with the specificity provided by the cellular environment, and may exert either activating or repressing effects (e.g., KDM4A demethylates both H3K9 and H3K36 with opposing outcomes) depending on the context. To date, no histone demethylase has a demonstrated role in removing H3K79 methylation

There are two distinct families of histone demethylases in mammalian cells, amine oxidases and oxygenases that are largely conserved throughout the eukaryotic kingdom. The human amine oxidase family has only two members (KDM1A/LSD1 and KDM1B/LSD2) and removes methyl groups in a flavine adenine dinucleotide-dependent reaction. They can demethylate mono and dimethyl lysines, but not trimethyl lysines, on histones and certain nonhistone proteins. In contrast, the oxygenase family members are capable of removing all three histone lysine methylation states in a Fe (II)- and α-ketogluterate-dependent process. The oxygenase family is referred to as the JUMONJI family due to the presence of a C-terminal catalytic JUMONJI domain (JMJC). The 32 known JUMONJI family members in the human genome are divided into groups based on sequence homologies and structural similarities (Fig. 3). Besides the JMJC domain, they can also possess ARID, Tudor, PHD, FBOX, and zinc-finger domains, as well as tetratricopeptide repeats and N-terminal JUMONJI (JMJN) domains. The ARID domain (AT-rich interaction domain) is a DNA-binding domain, whereas both the PHD (Plant Homeodomain) and Tudor domains recognize methylated histone residues, serving as recruitment platforms for other proteins. FBOX domains and tetratricopeptide repeats serve as protein interaction scaffolds, whereas zinc finger domains are versatile and can bind

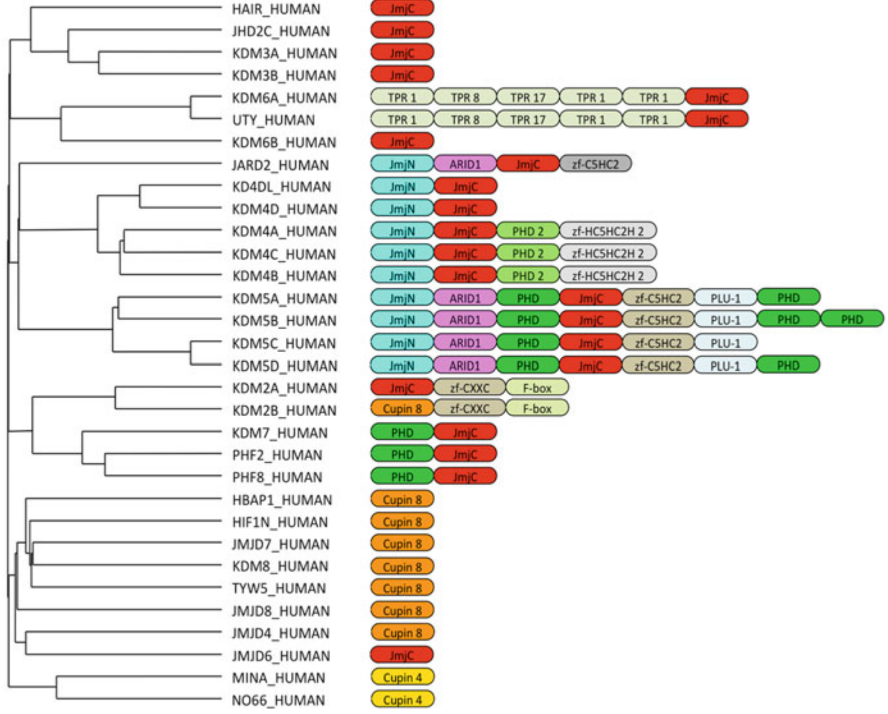



**Fig. 3** Alignment of all the KDMs from the human genome. The human genome has 32 KDMs based on the presence of the JMJC/cupin domain. The main feature of these proteins is that they use Fe(II) ions and a-ketogluterate as cofactors to remove methyl groups from histones. Some of these proteins also have the ability to hydroxylate protein substrates. In addition to the JMJC/cupin domain most possess other domains involved in guiding the proteins to their substrates, providing specificity. *JMJC* Jumonji C-terminal domain, *JMJN* Jumonji N-terminal domain, *ARID1* AT-rich interaction domain, *PHD* Plant homeodomain, *TPR* tetratricopeptide repeat, *ZF* zinc finger domain, F-BOX F-box sequence motif

DNA, RNA, or other proteins. While it is clear that JMJC domains possess catalytic activity, to date none of the JMJN domains has been demonstrated to have enzymatic activity. Recent evidence suggests that the JMJN domain in GIS1, the yeast homolog of human KDM4A physically interacts with the JMJC domain to control protein stability and transcriptional regulation [16]. The presence of these additional domains modifies the specificity of this group of enzymes. While members of the same group often act on similar substrates, they often differ in their temporal and spatial expression patterns, leading to different roles in the control of chromatin structure and transcriptional output.

Most JUMONJI proteins are able to remove methyl groups from particular methylated lysines on histones, but some are specific for arginine residues, nonhistone proteins (JMJD6), or act primarily as recruitment factors [17–20]. Several JUMONJI family members have not been investigated in detail, and their demethylase activities, substrates, and functional roles have not been described. Here, we review the roles of histone demethylases in the regulation of chromatin structure and dynamics, as well as their relationships to cellular processes and human diseases, with a specific focus on the data obtained through AP–MS analysis of protein complexes (Table 1).

## *KDM1 Family*

KDM1A and KDM1B are highly homologous, act on similar substrates, and have multiple cellular roles governed largely by the compositions of the complexes in which they reside. KDM1A can have diametrically opposite roles in that it demethylates two different lysine residues, one of which is a mark of active transcription (H3K4me1/me2), the other associated with transcriptional repression (H3K9me1/2) [21, 22]. Formation of a ligand-dependent complex between KDM1A with either androgen receptor (AR) or estrogen receptor alpha (ERα) leads to H3K9me1/2 demethylation at their respective target genes [22, 23]. This demethylation event produces hydrogen peroxide, leading to local DNA damage that, in turn, recruits DNA repair enzymes, themselves important for the transcriptional activation process at these loci. In contrast, transcriptional repression by KDM1A is mediated by its association with the corepressor complex, CoREST, with KDM1A demethylating the methylated H3K4, leading to transcriptional repression of CoREST target genes [22, 24–26]. Recently, KDM1A has also been shown to associate with another complex involved in transcriptional repression, the SWI/SNF chromatin-remodeling complex known as NURD [27]. Here, the metastasis-associated (MTA) protein plays a functional role analogous to that of CoREST, directing the demethylation of NURD target genes [27]. An added layer of regulation can be achieved through cross-talk with other histone modifications. For example, phosphorylation of histone H3 can alter the substrate specificity of KDM1A, inhibiting KDM1A-dependent demethylation of H3K4 [28].

Very recently, SFMBT1 was identified as another KDM1A interactor. SFMBT1 is a member of the MBT family of proteins. MBT domains bind mono and dimethylated lysine residues on histones and have been postulated to be components of polycomb repressive complexes (PRC) [29]. SFMBT1 was shown previously to function as a transcriptional repressor and this interaction suggests that SFMBT1 may exert its repressor activity through KDM1A [30]. It is clear from these results that KDM1A function is highly modulated by interactions with other proteins and complexes.

**Table 1** Human histone demethylases, their substrates, postulated functions, and their associations with human disease

| Name | Alternate name | Substrate | Function | Disease link |
|---|---|---|---|---|
| KDM1A | LSD1 | H3K4ME1/2, H3K9ME1/2 | Regulator of hematopoiesis | Breast and prostate cancer |
| KDM1B | LSD2 | H3K4ME1/2 | Genomic imprinting | – |
| KDM2A | JHDM1A | H3K36ME2, H3K4ME2/3 | Cell proliferation, senescence, rRNA tx, genomic stability | Prostate cancer |
| KDM2B | JHDM1B | H3K36ME2, H3K4ME2/3 | Cell proliferation and senescence | Liver and breast cancer |
| KDM3A | JMJD1A | H3K9ME1/2 | Metabolism, spermatogenesis | Kidney, renal, bladder, lung, colon, and prostate cancer |
| KDM3B | JMJD1B | H3K9ME1/2 | Cell proliferation | Leukemia, colorectal cancer |
| KDM3C | JMJD1C | H3K9ME1/2 | ES cells differentiation | Gastric cancer |
| KDM4A | JMJD2A | H3K9ME2/3,H3K36ME2/3 | Cell proliferation and replication, DNA damage | Breast cancer |
| KDM4B | JMJD2B | H3K9ME2/3,H3K36ME2/3 | DNA damage, genomic stability | Breast and colorectal cancer |
| KDM4C | JMJD2C | H3K9ME2/3,H3K36ME2/3 | ES cell renewal, development | Squamous cell carcinoma, prostate cancer and obesity |
| KDM4D | JMJD2D | H3K9ME2/3,H3K36ME2/3 | Cell proliferation and survival, spermatogenesis | – |
| KDM4DL | | – | – | – |
| KDM5A | JARID1A | H3K4ME2/3 | Proliferation and DNA repair | Leukemia, arthritis |
| KDM5B | JARID1B | H3K4ME2/3 | Cell cycle | Breast cancer |
| KDM5C | JARID1C | H3K4ME2/3 | Neuronal differentiation | X-lined mental retardation |

| | | | | |
|---|---|---|---|---|
| KDM5D | JARID1D | H3K4ME2/3 | Development | Prostate cancer |
| KDM6A | UTX | H3K27ME/2/3 | ES cells differentiation | Leukemia |
| KDM6B | JMJD3 | H3K27ME/2/3 | ES cells differentiation | Prostate cancer |
| KDM6C | UTY | – | – | Immunodeficiency |
| KDM7A | JHDM1D | H3K9ME1/2, H3K27ME1/2, H4K20me1 | Neuronal differentiation and brain development | – |
| KDM7B | PHF8 | H3K9ME1/2, H4K20ME1 | Cell cycle progression and brain development | Prostate cancer |
| KDM7C | PHF2 | H3K9ME1/2 | Liver and rDNA transcription | Esophageal squamous cell carcinoma |
| KDM8 | JMJD5 | H3K36ME2 | Cell cycle progression | – |
| JARID2 | | – | ES cells differentiation | T-cell acute leukemia |
| HR | | – | – | Hair deficiency, loss |
| HSPBAP1 | | – | Stress response | Epilepsy |
| HIF1AN | | ASPARAGINE HYDROXYLASE | Hypoxia response | Breast cancer, renal carcinoma |
| MINA | MINA53 | | Cell proliferation | Esophageal squamous cell carcinoma |
| NO66 | | H3K36ME2, H3K4ME3 | Osteoblast differentiation | – |
| JMJD4 | | – | – | – |
| JMJD6 | PTDSR | – | RNA splicing | – |
| JMJD7 | | – | – | – |
| JMJD8 | | – | – | – |
| TYW5 | | RNA HYDROXYLASE | | – |

Far less has been reported about the functional significance of the other amine oxidase family member, KDM1B. Like KDM1A, KDM1B can demethylate H3K4me1/2 and H3K9me1/2, again suggesting a dual role in both transcriptional repression and activation [31]. KDM1B has been shown to interact with RNA polymerase II elongation factors and demethylates H3K4me2 during transcription elongation. This leads to repression of target genes and plays a role in the establishment of genomic imprinting [32, 33]. In contrast, an interaction with NF-kB directs KDM1B to NF-kB target genes where it demethylates H3K9me2, thereby relieving the H3K9 silencing mark and leading to transcriptional activation [34]. Interestingly, KDM1B interacts with components of the DNA replication machinery and proteins involved in DNA repair, as well as nucleosome remodeling and histone modification complexes, indicating that KDM1B may be a member of multiple pathways [32]. KDM1B can also repress transcription through its N-terminal zinc finger domain independently of its histone demethylase activity [35].

## *JUMONJI Family*

### KDM2 Group

The KDM2 group is characterized by the presence of ARID, PHD, and FBOX domains in addition to the JMJC domain and contains two group members in mammals: KDM2A, the first JMJC family member to be identified; and its close homolog, KDM2B [36, 37]. Both KDM2A and KDM2B are able to demethylate H3K36me2 and H3K4me2/3, although the preference of KDM2B for either of these modifications is controversial. Nevertheless, both repress RNA polymerase I (RNA POLI) and RNA Polymerase II (RNA POLII) transcription and associate with the repressive polycomb group complexes (PRC) [36, 38, 39].

Despite high homology and similar substrate specificity of KDM2A and KDM2B, the two proteins have distinct cellular localizations and functions. KDM2A is found in regions of constitutive heterochromatin, including the pericentromeric satellite repeats. This localization is likely mediated through KDM2A binding to the repressive histone methylation mark, H3K9me3, and occurs preferentially in CpG islands that lack DNA methylation, an example of cross talk between histone and DNA marks [38, 40]. All three heterochromatin binding proteins (HP1) physically associate with KDM2A, and the ability of HP1 to localize to heterochromatin is reduced in KDM2A knockdowns, suggesting that KDM2A recruits these proteins [38]. In addition, KDM2A acts to repress transcription of pericentromeric satellite repeats, controlling centromeric integrity and genomic stability during mitosis [38, 41, 42]

There are contradictory reports as to the substrate for KDM2B. While some reports suggest that, similarly to KDM2A, KDM2B is mainly an H3K36me1/2 demethylase, other reports characterize it primarily as an H3K4me3 demethylase [36–38]. KDM2B has been shown to control transcriptional regulation of the

proliferation/senescence locus, p15/ink4b, through demethylation of H3K36me2 [37]. This finding provides a mechanism for a previous observation that the PRC2 polycomb complex regulates transcription of this locus, as KDM2B interacts with the EZH2 component of the PRC2 complex [43–45]. On the other hand, KDM2B has been shown to repress transcription of ribosomal genes and regulate cell morphology, chemokine expression, and apoptosis through demethylation of H3K4me3 [38, 39, 46]. Because the PHD domain of KDM2B is able to bind both H3K4me3 and H3K36me2, it is possible that KDM2B substrates are dictated mainly by the complexes associated with KDM2B during various cellular processes.

The KDM2 group is a good example of different very closely related enzymes acting on the same substrates in vitro or even in vivo, while exerting very different effects within the context of a cell. Subtle differences due to cellular localization, interacting cofactors, or temporal expression patterns are usually not accounted for by in vitro studies, which presumably is why conflicting results are obtained as to substrates and recognition of particular methylated histone residues.

## KDM3 Group

The KDM3 group consists of three members, KDM3A, KDM3B and KDM3C, characterized by the presence of JMJC and FBOX domains and the ability to demethylate H3K9me1/me2. As H3K9 methylation is usually repressive, this group mediates transcriptional activation as opposed to repression.

KDM3A has a variety of roles, including transcriptional activation of metabolic, spermatogenesis-related and androgen receptor-target genes, as well as the control of reprogramming and the hypoxia response [47–50]. During spermatogenesis, KDM3A physically interacts with the activated androgen receptor (AR), and they co-localize to AR target gene promoters, where KDM3A removes the repressive H3K9me3 mark to activate transcription [47]. Similarly, in the hypoxia response, KDM3A interacts physically with the hypoxia response factor, HIF-1, which recruits KDM3A to the promoter of the glucose transporter gene, GLUT3, where it demethylates H3K9me3 and activates transcription [50]. KDM3A-mediated demethylation of the HOXA1 promoter can drive cell division, and, in ES cells, KDM3A appears to play a role in the reactivation of ES cell-specific gene expression [49, 51].

KDM3B is still poorly understood but has been shown to regulate cell proliferation through an interaction with the colorectal cancer-related metastatic protein, PRL-3 [52]. KDM3C expression is controlled by the ES cell-specific POU5F1 (OCT3/OCT4) transcription factor and, consequently, KDM3C has been implicated in the control of gene expression during ES cell differentiation and in pancreatic islets [53]. KDM3C interacts with the NSD3 histone methyltransferase complex, leading to a coordinated regulation of mouse testes development [54]. The exact mechanisms through which members of the KDM3 group exert their multiple functions have not been determined, but direct control of gene expression by KDM3A on genes regulating development has been recently reported [55].

### KDM4 Group

The KDM4 group consists of five members, KDM4A/B/C/D and KDM4DL, each of which contains a JMJN domain in addition to a JMJC domain, two PHD domains, and two Tudor domains. KDM4A and KDM4C are capable of removing all three methylation states from H3K9 and H3K36, with higher activity on the trimethylated residues, while KDM4B and KDM4D are confined to H3K9me3 demethylation, and the substrates and functions for KDM4DL are unknown [56–58]. Both KDM4A and KDM4B can bind modified histones other than their substrates through their Tudor domains. While KDM4A can bind to methylated H4K20 and H3K4, KDM4B can bind only to methylated H4K20 [59–61]. Binding of KDM4A/KDM4B to H4K20 prevents P53BP1 recruitment to the same modified residue and thereby suppresses the DNA damage response. After the induction of DNA damage, KDM4A and KDM4B are degraded by the FBXO22-containing SCF E3 ubiquitin ligase, allowing P53BP1 to bind to the exposed methylated H4K20 and initiating the DNA damage response [59, 62].

Besides its role in the DNA damage response, the KDM4 group has additional cellular roles determined largely by its association partners. KDM4A acts both as a corepressor of E2F target genes through its interaction with HDAC1-3 and RB and as a coactivator of AR-target genes [63]. Overexpression of KDM4A leads to global changes in chromatin accessibility, accelerated cell cycle progression, and aberrant replication timing. These may be mediated through alterations in the localization of HP1γ, potentially explaining the linkage of KDM4 to multiple types of cancer [64].

KDM4B contributes to pericentromeric stability and chromosome segregation by demethylating H3K9me3 in pericentromeric chromatin [65]. It also directly targets the expression of cyclin-dependent kinase 6 (CDK6), which is essential for the G1/S transition, thereby contributing to the aberrant cell cycle progression phenotype of KDM4B mutants [66].

KDM4C, in concert with KDM1A, acts as a transcriptional activator for androgen receptor target genes and controls the self-renewal of ES cells by regulating the transcription of pluripotency-specific transcription factors [53, 67]. In mouse development it may regulate the expression of the proliferation-related transcription factors, MYC and KLF4 [68]. KDM4D has a number of disparate roles. It binds directly to p53 and activates p53 target genes, presumably through the removal of repressive H3K9 methylation marks on the promoter regions of these genes. The same study also found that KDM4D can act in an opposing manner, in a p53-independent pathway, to stimulate cell proliferation and survival, illustrating the complex balance of in vivo functions [69]. KDM4D also appears to be responsible for regulation of spermatogenesis via the activation of androgen-responsive genes and for demethylation of repressive H3K9 methylation marks surrounding the enhancers of tissue-specific genes; although this enhances expression, it is not sufficient for gene activation [70–72].

**KDM5 Group**

The KDM5 group members each contain five conserved domains, including both JMJN and JMJC domains, as well as ARID, two PHD, and Zinc finger domains. This family has four members (KDM5A through D) with KDM5A and KDM5B being located on the autosomes, whereas KDM5C and KDM5D are located on the X and Y chromosomes, respectively. All KDM5 family members have been shown to specifically recognize and demethylate H3K4me2/3 during the cell cycle and differentiation and, as such, are generally involved in transcriptional repression. KDM5A interacts with the SIN3B complex and acts in an RB-dependent manner to silence RB target genes enabling senescence. It can also repress genes regulated by the Notch pathway [73–75]. Moreover, KDM5A is closely connected with the PRC2 complex. The PRC2–KDM5A complex imposes transcriptional silencing on target genes by coordinated trimethylation of H3K27me3 and demethylation of H3K4me3 [76]. Recently, KDM5A has also been implicated in the DNA damage response where, upon ionizing radiation-induced double-strand break formation, KDM5A accumulates at the sites of DNA damage. Other silencing marks, such as H3K27me3, are incorporated during the DNA damage response, and so it is plausible that KDM5A involvement in DNA repair is mediated through a polycomb-related mechanism [77].

KDM5B is a transcriptional repressor with roles in neural differentiation, senescence, and cellular proliferation [73, 78–80]. Through an association with the transcription factors MYC and TFAP2C, KDM5B acts to downregulate the cell cycle gene CDKN1A, thereby promoting cell cycle progression [79]. Similarly, in senescence, KDM5B associates with RB, promoting repression of RB-dependent cell cycle genes [81]. In contrast with the described functions of the other KDM5 family members, KDM5B is able to function as a transcriptional activator during self-renewal. This seemingly contradictory finding stems from the observation that KDM5B is recruited by the histone acetyltransferase complex member MORF4L1 to the bodies of actively transcribed genes where demethylation of H3K4 inhibits cryptic initiation of transcription, a process that interferes with efficient RNA polymerase elongation [82].

KDM5C has a role in transcriptional repression of a subset of neuronal genes through its association with REST, a transcription factor required for silencing of neuronal genes in non-neuronal tissues, histone acetyltransferases, HDAC1-2, and the histone methyltransferase EHMT2 [83]. KDM5C also interacts with PCNA through a PCNA interaction motif in its sequence, and this interaction is necessary for KDM5C association with chromatin [84]. The polycomb-like protein RING6A associates with KDM5D and promotes the demethylase activity of KDM5D in vitro. In vivo, KDM5D appears to target RING6A to developmentally controlled genes where they act in concert to repress transcription [85]. KDM5D might also play a role in spermatogenesis, more specifically in the regulation of meiosis, as it forms a complex with MSH5 and appears to target MSH5 to condensed chromatin during meiotic prophase [86]. Therefore, it is obvious that, even though KDM5 enzymes all have identical substrates, they perform very different roles within

cellular networks and their roles are defined by both their expression patterns and the complexes with which they are associated.

### KDM6 Group

The KDM6 group (KDM6A-C) is characterized by the presence of several tetratricopeptide repeats, in addition to the JMJC-domain, KDM6A and KDM6B are both able to demethylate H3K27me2 and me3 but, to date, no demethylase activity has been detected for KDM6C [87–89]. Because H3K27me2/3 are repressive marks established mostly at gene promoters and in coding regions, KDM6A and KDM6B are transcriptional coactivators involved in multiple processes, including cell cycle progression, differentiation, development, and the inflammatory response. KDM6A co-purifies with the H3K4 methyltransferase complexes that contain the mixed-lineage leukemia proteins, MLL1-3, resulting in H3K27me3 demethylation along with H3K4me3 methylation and leading to the activation of gene expression (e.g., the HOX gene cluster) [90–92]. Such coordinated activation of gene expression by complexes exhibiting different but reinforcing activities, MLL-KDM6A and PRC2-KDM5A complexes, illustrates how an assembly of protein complexes with functionally synergistic activities can be an efficient way to control chromatin structure and dynamics.

Similarly to KDM6A, KDM6B is an important activator of HOX genes and bivalent promoters (promoters in ES cells that are marked by both activating and repressive marks) [85, 87, 93, 94]. However, in spite of extensive homology with KDM6A, KDM6B does not appear to associate with MLL; instead, it associates with KDM7A, an H3K9me1/2, H3K27me1/2, and H4K20me1 histone demethylase, along with proteins regulating transcription elongation [95]. KDM6B also has another role in that it cooperates with KDM4B to control the differentiation potential of human bone marrow mesenchymal stem/stromal cells, again illustrating collaborative efforts within the KDM family [96].

As mentioned previously, KDM6C has no detectable demethylase activity, and so it is unclear what role, if any, it does play. However, protein–protein interaction data from the Greenblatt lab indicates that, similarly to KDM6A, KDM6C co-purifies with components of the MLL complexes (unpublished data). It is possible, therefore, that it does have demethylation activity that requires the presence of previously uncharacterized cofactors, or rather functions as a recruitment/stabilization factor rather than a bona fide histone demethylase.

### KDM7 Group

The KDM7 group includes three members (KDM7A-C) and is characterized by the presence of PHD and JMJC domains. KDM7A is an H3K9me1/2 and H3K27me1/2 demethylase, while KDM7B demethylates both H3K9me1/2 and H4K20me1, and

KDM7C seems to use exclusively H3K9me1/2 as a substrate [36, 97–99]. KDM7A is a positive transcriptional regulator that localizes to the nucleolus, where it is involved in the transcription of rDNA [97, 100, 101]. It is also important for activation of neuronal-specific genes and brain development and directly controls the transcription of FGF4, an oncogenic growth factor [102, 103].

KDM7B, in addition to its ability to demethylate H3K9me1/2, was the first demethylase found to act on H4K20me1, a modification important for cell cycle progression, neural differentiation, and brain development [98, 99]. KDM7B interacts directly with the CTD of RNA polymerase II and serves as a general coactivator, present at many active genes [104]. In addition, KDM7B acts in concert with activating factors E2F1, HCFC1, and SETD1A to demethylate H4K20me1 onE2F1-target genes [99].

KDM7C is a transcriptional activator of HNFA in liver and rDNA genes. Interestingly, unmodified KDM7C is inactive and requires PKA-dependent phosphorylation to bind, demethylate, and form a complex with ARID5B. This activated KDM7C/ARID5B complex is then recruited to promoters, where it demethylates H3K9me2, leading to transcriptional activation [105]. All members of KDM7 family are able to bind H3K4me3 marks but do not use them as a substrate. Instead, binding to the trimethylated lysine 4 on histone H3 increases the enzymatic activity of KDM7 family members towards H3K9me2, illustrating the influence of other domains within the same protein on the catalytic activity of histone demethylases [99, 106].

### JARID2

JARID2 possesses both ARID and zinc finger domains in addition to JMJN and JMJC domains. However, it does not exhibit any detectable demethylase activity, acting instead as a recruitment factor for other chromatin-modifying complexes. JARID2 is a component of polycomb repressive complex 2 (PRC2), where it is important for targeting PRC2 to its target genes and modulating its activity [18, 107]. JARID2 also interacts with the histone methyltransferase SETDB1 to control the levels of H3K9me2/3 at the NOTCH locus, thereby regulating NOTCH expression [108]. Through its recruiting ability, JARID2 has been implicated in transcriptional regulation and ES cell differentiation.

### JMJC-Only Demethylases

There are other JUMONJI family members which possess a JMJC domain but contain no other recognizable domains and thus are not grouped into specific families (Fig. 3). JMJC-only members are, by far, more obscure and less studied, with few exceptions. KDM8, an H3K36me2 demethylase, is known to participate in cell cycle progression and circadian systems [109–111]. NO66 is a histone demethylase with specificity for H3K4me3 and H3K36me2. Together with SP7/OSX, it regulates osteoblast differentiation by demethylating H3K4 and H3K36 and inhibiting

SP7/OSX-mediated promoter activation. It may also function in replication and remodeling of heterochromatic regions [112, 113]. JMJD6 is the only JMJC-containing protein which has been demonstrated to be capable of demethylating arginine residues on histones H3 and H4 in vitro, but it is not clear whether this is its primary in vivo function as it also demethylates nonhistone substrates [17]. JMJD6 is required for organogenesis, hematopoietic differentiation, and regulation of cytokine responses [114–116]. For other JMJC-only group members, such as HSPBAP1, HIF1AN, HR, MINA53, JMJD4, JMJD7, JMJD8, and TYW5, no demethylase activities, histone substrates, or binding sites have been identified. Several of them have known cellular roles, although the mechanisms of their actions remain unclear. HSPBAP1 might have a role in the cellular stress response [72]. HIF1AN plays a role in the response to hypoxic conditions by hydroxylation of asparagines residues within HIF, suppressing its transcriptional activity [19, 117]. HR could act as transcription factor regulating cell growth, possibly through the hyperactivation of WNT signaling pathways [54, 118]. MINA53 was identified in a screen searching for MYC targets, is present in the nucleolus and, as is the case for MYC, is involved in cell proliferation [119]. TYW5 is a tRNA hydroxylase that acts in the biosynthesis of a hypermodified nucleoside, hydroxywybutosine, which is essential for correct phenylalanine codon translation [20]. JMJD7-PLA2G4B is a read-through protein encoding a calcium-dependent phospholipase, while nothing is known about JMJD4 or JMJD8 aside from the fact that they contain C-terminal JMJC domains [120].

## *Demethylation of Nonhistone Substrates*

Even though most histone demethylases work specifically on histone substrates, some are able to remove methyl marks from nonhistone proteins, thereby modifying their activities. The tumor suppressor protein p53 is one of the most extensively studied methylated proteins, as it can be methylated on several residues to various extents. The degree of methylation and the methylated residue determine p53 activity and its roles during cell cycle progression, DNA repair, and apoptosis. Several histone methyltransferases, including KMT3C, KMT5A, and KMT7, methylate p53 on different residues and in response to different environmental stimuli. However, KDM1A is the only demethylase that has, to date, been shown to demethylate p53, preventing p53 interaction with P53BP1, and thereby inducing apoptosis [27, 121, 122].

KDM1A can also relieve methylation on E2F1 and DNMT1. E2F1 is a transcription factor with a role in cell cycle progression and apoptosis, and its target genes include several pro-apoptotic factors in DNA damage-induced apoptosis. Methylation of E2F1 by KTM7 results in transcriptional activation, while demethylation by KDM1A destabilizes E2F1, inhibiting DNA damage-induced cell death, and promoting DNA repair [123]. DNA methyltransferase 1 (DNMT1) is also

methylated by KMT7, with effects that are opposite to that on E2F1. In this case, demethylation by KDM1A, rather than methylation by KMT7, stabilizes DNMT1, ultimately regulating DNA methylation [124]. This connection between KDM1A and DNMT1 then provides an important link between DNA methylation and histone modification in the regulation of chromatin dynamics. Furthermore, it seems KDM1A and KMT7 have opposing roles in the regulation of the methylation status of nonhistone substrates.

Only a few members of the JUMONJI family of lysine demethylases have so far been shown to demethylate nonhistone substrates. KDM2A demethylates the p65 subunit of the NF-κB, a regulator of immune and inflammatory responses, reversing the mark imposed by KMT3B, and leading to the inhibition of NF-κB signaling [125]. Interestingly, KDM2A expression is itself driven by NF-κB signaling in a feedback loop. KDM4A-C can demethylate trimethylated lysine peptides in several nonhistone proteins in vitro, including WIZ, CDYL1, CBS, and EHMT1 [126]. All of these proteins are found in chromatin-related complexes regulating transcription, illustrating yet another example of a cross talk within chromatin modification machinery.

KDM7C demethylation of the nonhistone protein, ARID5B, is necessary for its function as a histone demethylase cofactor. KDM7C is activated upon phosphorylation by a protein kinase (PKA) and can then demethylate ARID5B. Only upon ARID5B demethylation can the ARID5B–KDM7C complex be targeted to H3K9ME2 at promoters of KDM7C target genes [105]. Even though JMJD6 is able to demethylate arginine residues on histones in vitro, histone demethylation does not seem to be a primary function of JMJD6 in vivo. Instead, it functions as a lysyl hydroxylase for the splicing factor U2AF65, thereby regulating its pre-mRNA splicing activity [127]. There are several other putative histone demethylases for which no substrates or possible functions have been identified but that can potentially act on nonhistone substrates (Table 1).

## The Role of Chromatin in Human Disease

The traditional carcinogenesis model proceeded from initiation (exposure to a carcinogen leading to a change in nucleotide sequence) through promotion (additional exposure to a carcinogen leading to enhanced cell division) to progression (malignant tumor formation). However, there are many carcinogens that do not lead to DNA sequence changes but, instead, affect chromatin structure. This altered chromatin structure can result in changes in gene expression or chromosome instability, leading to an imbalance among apoptosis, proliferation, and differentiation. Even in the absence of specific carcinogens, the same effects can be induced by mutations or defects in many of the chromatin-modifying enzymes, including DNA repair proteins.

The best studied examples of cancer susceptibility genes are BRCA1/2, and mutations in these DNA repair genes are responsible for 2–10 % of all breast cancers and 5–10 % of all ovarian cancers worldwide. The inheritance of BRCA1/2 mutations increases the risk of breast cancer by 50–80 %, the risk of contralateral breast cancer by 60 %, and the risk of ovarian cancer by 15–25 % (reviewed in [128]). The incomplete penetrance of the cancer phenotype illustrates that not all individuals with a particular genetic mutation will develop breast cancer and, conversely, patients with the same tumor type may have very divergent genetic or epigenetic changes. Presumably, the specific combination of genetic mutations, genetic interactions between gene products, differences in epigenetic programming, and exposure to environmental factors can all influence cancer predisposition and manifestation.

So far, mutations in a number of different chromatin-related proteins have been linked to serious developmental disorders. Alpha-thalassemia/mental retardation syndrome results from loss of function mutations in the chromatin-remodeling enzyme, ATRX, and may be due to changes in DNA methylation patterns [129, 130]. ICF syndrome, a rare autosomal recessive disorder characterized by immunodeficiency, instability of pericentromeric heterochromatin, mental retardation, and developmental defects, is linked to mutations within DNMT3B, a DNA methyltransferase [131]. Similarly, MECP2 contains a methyl-CpG recognition domain, and loss of function mutations in this protein cause Rett syndrome, the most frequent cause of mental retardation in females [132]. Rubinstein–Taybi Syndrome (RSTS) is characterized by congenital malformation and mental retardation stemming largely from the mutations in the histone acetylase CBP [133]. EGF-stimulated phosphorylation of histone H3 on serine 10 by the serine–threonine PKA is a leading cause of Coffin–Lowry Syndrome [134]. In addition to these more prevalent and better studied disorders, many more have been linked to the aberrant expression of chromatin modifying enzymes, especially many types of cancer.

The link between chromatin modifications and human disease is strong, and listing all of the currently known connections here would be impossible. Enrichment profiles of all known or predicted chromatin modification enzymes with links to human disease demonstrate the wide-ranging effects chromatin can have on many aspects of gene regulation and chromosome stability (Fig. 4a). Here, neoplasms or abnormal tissue masses show highest enrichment, but there are other highly enriched categories. Figure 4b shows hierarchical view of the most enriched subsets and illustrates how broad categories can be partitioned into smaller categories or single disorders.

From this, it is obvious that often members of the same complex are implicated in the same disease allowing prediction of disease-related genes and consequently predictions of novel therapeutic targets. Thus, identification of histone demethylases associated with disease and their potential partners can be advantageous in the identification of novel disease genes and therapeutic targets.

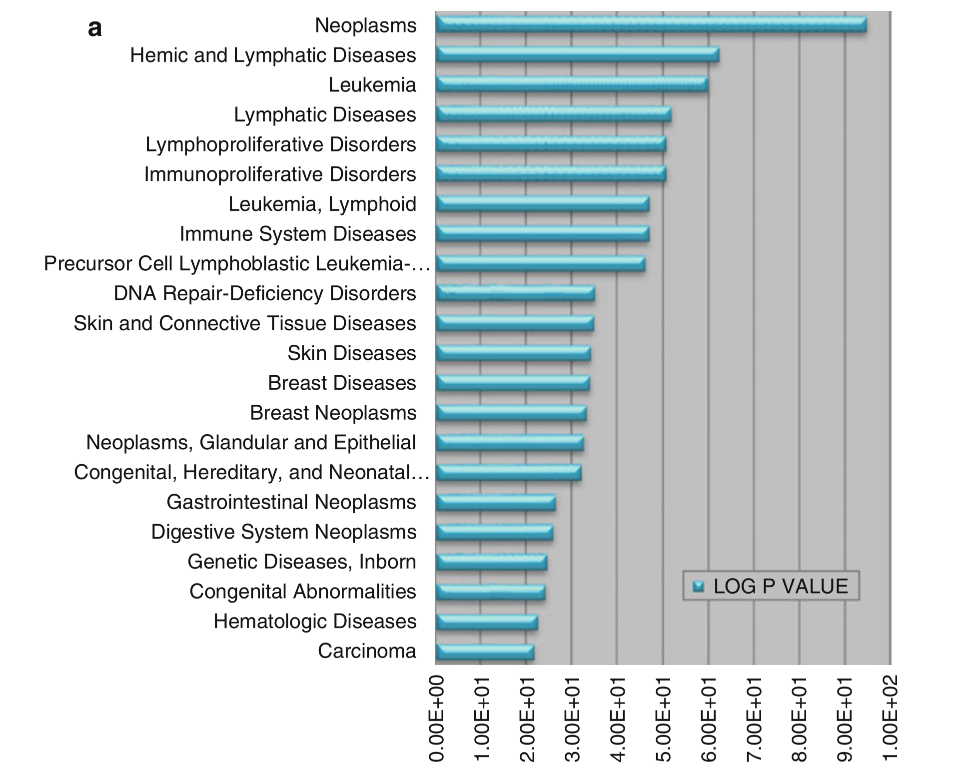



**Fig. 4** Disease associations of chromatin-related proteins. (**a**) Enrichment of human disorders among chromatin-related proteins according to the *p* value. (**b**) A hierarchical illustration of the most enriched categories. Disease associations and hierarchies were built on 753 either known or hypothetical chromatin modifiers, which had 4,489 neighbors in BIOGRID and high confidence iRefWeb, using a combination of three disease sources (GAD, gene2Mesh, PharmGKB) with *p*-value of 0.05 or better after Benjamini–Hochberg correction

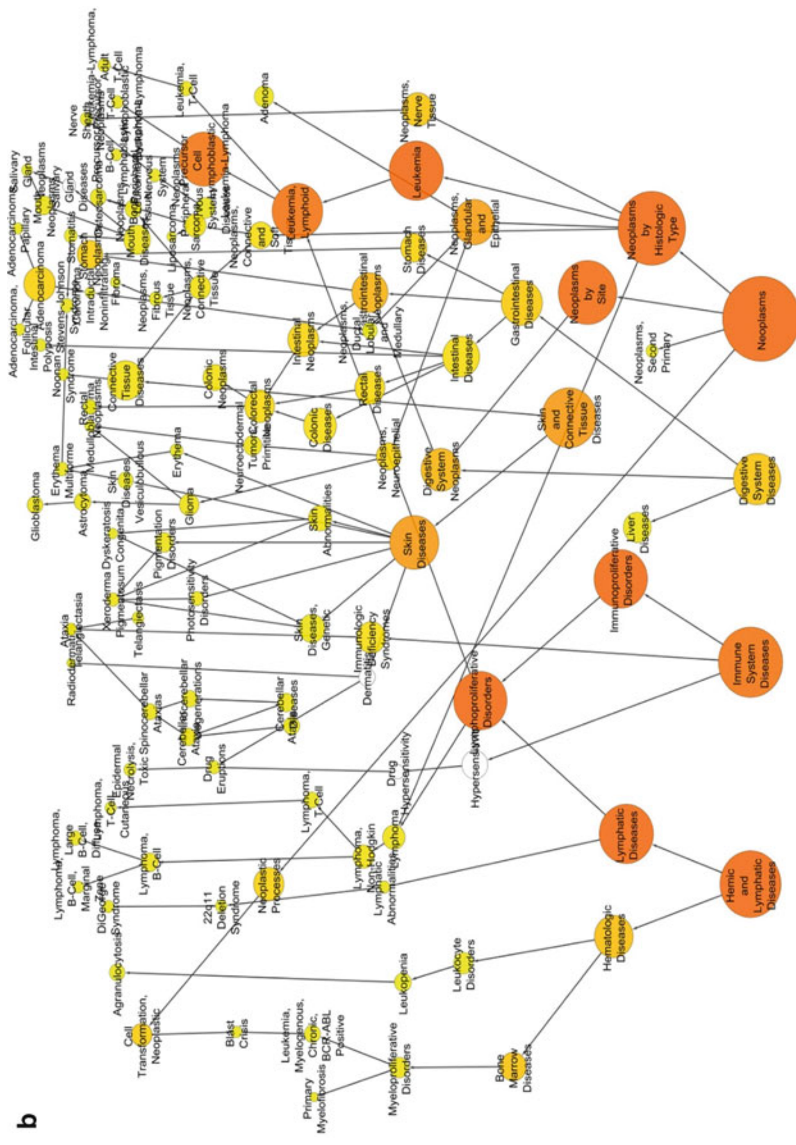



Fig. 4 (continued)

### *Histone Demethylases in Human Disease*

As with other chromatin regulatory molecules, aberrant expression of several histone demethylases has been linked to a variety of human disorders, various developmental defects and a large number of cancers (Table 1). Echoing their divergent roles in the cell, including both repressive and activating functions, histone demethylases can act as either proto-oncogenes or tumor suppressors. As is the case for their roles in transcriptional regulation, several KDMs can act in concert to affect different stages of the tumorigenic process. Below, we try to summarize known associations between KDMs and various human cancers.

KDM1A, a potential proto-oncogene, is overexpressed in prostate cancer, breast cancer, and neuroblastoma [135–137]. KDM4C and KDM7B are both also overexpressed in prostate cancer but, while KDM1A and KDM4C are both involved in proliferation, KDM7B has been implicated in cell migration and invasion, illustrating how different demethylases can contribute to the formation of the same type of tumor [67, 138]. Other demethylases, including KDM2A, KDM4A, KDM5B, and KDM6B, also contribute to prostate cancer manifestation and progression when overexpressed [70, 139]. In the case of KDM2A, this appears to occur through its role in the maintenance of the heterochromatic state and genomic stability, while KDM4A, KDM5B, and KDM6B exert their effects through the activation of androgen receptor target genes [42]. Another member of KDM5 family, KDM5D, may also promote prostate cancer development, as it is deleted in 52 % of cases [140]. However, the direct link has not being established. KDM4C has been implicated in the development of carcinomas other than prostate cancer, including esophageal carcinoma, medulloblastoma, and lymphoma [141–143]. It seems that KDM4C also contributes to obesity through the transcriptional repression of PPARγ, a nuclear receptor responsible for the regulation of adipose cell differentiation [144].

Depending upon the context, KDM2B can act as either a proto-oncogene or tumor suppressor. Overexpression of KDM2B has been linked to breast cancer and lymphomas, while depletion of KDM2B has been observed in brain and liver cancers [38, 145]. The tumorigenic potential of KDM2B is due to its effects on both proliferation and DNA repair.

Overexpression of KDM3A has been observed in bladder and lung cancers, presumably through its activation of HOXA1 expression that, in turn, drives the expression of genes in the p44/p42 mitogen-activated protein kinase pathway, leading to oncogenic transformation [51]. KDM3B, together with the CBP-containing histone acetyltransferase complex, plays a role during acute lymphoblastic leukemia by targeting the activation of the leukomegenic LMO2 gene. KDM3B together with KDM4B has been implicated in colorectal cancer formation and progression, presumably through their interaction with PRL-3, a metastatic gene in colorectal cancer [52]. However, while KDM3B seems to function as a tumor suppressor, KDM4B is a potential oncoprotein. In addition to colorectal cancer, KDM4B has also been linked to breast cancer through its role in the establishment of pericentromeric heterochromatin and thus the maintenance of genomic stability [65]. KDM4A and

KDM4D are both potential oncogenes, with KDM4A linked to breast cancer formation through the stimulation of ERα activity, a function dependent on the catalytic activity of KDM4A [146]. KDM4D also has pro-proliferative potential through its interaction with p53 [69].

Collectively, the KDM5 family has been linked to multiple human disorders. KDM5A interacts with many proteins previously correlated with tumorigenesis, including TBP, LMO2, MYC, SIN3/HDACs, and RBP-J, and is a tumor suppressor [139, 147]. Translocations between KDM5A and a nuclear pore protein, NUP98, can lead to acute myeloid leukemia [148]. Mutations within KDM5A confer susceptibility to ankylosing spondylitis, a form of chronic, inflammatory arthritis [149]. Overexpression of KDM5B is often observed in breast, testis, and bladder cancer presumably through its interactions with two developmentally regulated transcription factors, BF-1 and PAX9 [150–152]. Through an interaction with KDM1A and components of NURD complex, KDM5B can suppress angiogenesis and metastasis in breast cancer cells [153]. KDM5C has been linked to X-linked mental retardation syndrome and epilepsy, but it can also regulate the HPV e2 tumor suppressor protein, a leading cause of cervical cancer [83, 154].

A few other KDMs have also been linked to tumor development or progression. KDM7B, in addition to its involvement in prostate cancer, has been associated with X-linked mental retardation [106]. KDM6A is a tumor suppressor gene, as mutations have been detected in multiple myelomas and leukemia [139, 155, 156]. HIF1AN may be linked to survival in invasive breast cancer through the regulation of HIF-1α, while mutations within MINA53 might contribute to esophageal squamous cell carcinoma, gastric cancer, and lung cancer [157–160]. Some of the less well-studied KDMs have also been linked to human disease. Intractable epilepsy, or epilepsy unresponsive to anticonvulsants, is associated with mutations in HSPBAP1, while mutations within HR lead to various forms of hair loss, including depletion of eyebrows, eyelashes, and body hair [161, 162].

## Conclusions

Although current understanding of the relationship between chromatin modification and human disease is based mainly on genetic associations, recent advances in AP–MS-based analyses of protein complexes are beginning to clarify not only the critical role histone demethylases play in regulating chromatin structure and function but also the interplay between different chromatin modifying pathways, such as DNA methylation and histone modifications, with human disease. AP–MS analyses of protein complexes have been so valuable in assigning functional identity to chromatin-related enzymes mainly because these enzymes act in the context of large macromolecular complexes that are not only essential for the activation of enzymatic activity but also provide recognition specificity. Often, such large protein complexes contain seemingly opposing enzymatic functions but, nevertheless, lead to synergistic outcomes for tighter transcriptional regulation. For example, the

H3K4me3 demethylase, KDM5A, interacts with a PRC2 complex harboring H3K27 methyltransferase activity. By coordinated demethylation of H3K4me3 and methylation of H3K27 this enzyme complex functions in transcriptional repression. In contrast, the KDM6A/MLL complex functions antagonistically to the KDM5A/PRC2 complex. In this case, KDM6A demethylates H3K27me3 while at the same time, MLL methylates H3K4, leading to transcriptional activation [87, 92]. As well, the KDM3C/NSD3 complex formation leads to transcriptional activation by KDM3C-dependent demethylation of H3K9me1/2 and NSD3-dependent methylation of H3K4 [54]. KDMs are also found in complexes with enzyme activities other than histone methylation. For example, KDM2B is associated with the PRC1 complex, which connects the removal of H3K36me2 and H3K4me2/3 with monoubiquitylation of H2A on lysine 119, leading to transcriptional silencing. Such synergistic roles have been also observed between histone demethylases and reader proteins that recognize methylated residues and act as recruitment platforms for additional chromatin modifiers. For example, HP1α, a chromodomain containing protein, not only recruits the histone methyltransferase SUV420H2 to heterochromatin, but also binds KDM4A leading to regulation of the methylation status of H3K36 [163, 164].

Because of the complicated nature of chromatin structure and function and the extensive amount of cross talk between different modifications and enzymes and their functional outcomes, chromatin has to be analyzed in terms of a large epigenetic network map within which the complexes perform their functions. Such a network map is not static, and different complexes can share subunits depending on intracellular and extracellular signals. The generation and analysis of such network maps would help in the identification of novel complex components and in functional predictions, providing a global view of chromatin dynamics. More importantly, the analysis of protein complexes provides an important means of identifying and characterizing novel human disease-related genes and has the potential of identifying potential new therapeutic targets. In fact, as is the case for other types of chromatin-modifying enzymes (e.g., HDACs or KMTs), several KDMs are already under development as potential therapeutic targets via the identification of KDM-specific inhibitors [165, 166].

## References


1. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403(6765):41–5.
2. Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. Cell Res. 2011;21(4):564–78.
3. Royer L, Reimann M, Stewart AF, Schroeder M. Network compression as a quality measure for protein interaction networks. PLoS One. 2012;7(6):e35729.
4. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440(7084):637–43.
5. Babu M, Vlasblom J, Pu S, Guo X, Graham C, Bean BD, et al. Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. Nature. 2012;489(7414):585–9.

6. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature. 2005;433(7025):531–7.
7. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Heriche JK, Sykora MM, et al. Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science. 2010;328(5978):593–9.
8. Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. Cell. 2009;138(2):389–403.
9. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, et al. Analysis of the human endogenous coregulator complexome. Cell. 2011;145(5):787–99.
10. Mak AB, Ni Z, Hewel JA, Chen GI, Zhong G, Karamboulas K, et al. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. Mol Cell Proteomics. 2010;9(5):811–23.
11. Ni Z, Olsen JB, Guo X, Zhong G, Ruan ED, Marcon E, et al. Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. Transcription. 2011;2(5):237–42.
12. Ni Z, Olsen JB, Emili A, Greenblatt JF. Identification of mammalian protein complexes by lentiviral-based affinity purification and mass spectrometry. Methods Mol Biol. 2011;781:31–45.
13. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet. 2012;13(5):343–57.
14. Wakeman TP, Wang Q, Feng J, Wang X-F. Bat3 facilitates H3K79 dimethylation by DOT1L and promotes DNA damage-induced 53BP1 foci at G1/G2 cell-cycle phases. EMBO J. 2012;31(9):2169–81.
15. Altaf M, Utley RT, Lacoste N, Tan S, Briggs SD, Côté J. Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin. Mol Cell. 2007;28(6):1002–14.
16. Quan Z, Oliver SG, Zhang N. JmjN interacts with JmjC to ensure selective proteolysis of Gis1 by the proteasome. Microbiology. 2011;157(Pt 9):2694–701.
17. Chang B, Chen Y, Zhao Y, Bruick RK. JMJD6 is a histone arginine demethylase. Science. 2007;318(5849):444–7.
18. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell. 2009;139(7):1290–302.
19. Mahon PC, Hirota K, Semenza GL. FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. Genes Dev. 2001;15(20):2675–86.
20. Noma A, Ishitani R, Kato M, Nagao A, Nureki O, Suzuki T. Expanding role of the jumonji C domain as an RNA hydroxylase. J Biol Chem. 2010;285(45):34503–7.
21. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell. 2004;119(7):941–53.
22. Metzger E, Wissmann M, Yin N, Muller JM, Schneider R, Peters AH, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature. 2005;437(7057):436–9.
23. Perillo B, Ombra MN, Bertoni A, Cuozzo C, Sacchetti S, Sasso A, et al. DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. Science. 2008;319(5860):202–6.
24. Lee MG, Wynder C, Cooch N, Shiekhattar R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature. 2005;437(7057):432–5.
25. Shi YJ, Matson C, Lan F, Iwase S, Baba T, Shi Y. Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell. 2005;19(6):857–64.
26. Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, et al. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell. 2006;23(3):377–87.

27. Wang Y, Zhang H, Chen Y, Sun Y, Yang F, Yu W, et al. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell. 2009;138(4):660–72.
28. Metzger E, Imhof A, Patel D, Kahl P, Hoffmeyer K, Friedrichs N, et al. Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4. Nature. 2010;464(7289):792–6.
29. Lin S, Shen H, Li J-L, Tang S, Gu Y, Chen Z, et al. Proteomic and functional analyses reveal the chromatin reader SFMBT1's role in regulating epigenetic silencing and the myogenic gene program. J Biol Chem. 2013;288(9):6238–47.
30. Wu S, Trievel RC, Rice JC. Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. FEBS Lett. 2007;581(17):3289–96.
31. Karytinos A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, et al. A novel mammalian flavin-dependent histone demethylase. J Biol Chem. 2009;284(26):17775–82.
32. Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, et al. Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Mol Cell. 2010;39(2):222–33.
33. Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, et al. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature. 2009;461(7262):415–8.
34. Van Essen D, Zhu Y, Saccani S. A feed-forward circuit controlling inducible NF-kappaB target gene activation by promoter histone demethylation. Mol Cell. 2010;39(5):750–60.
35. Yang Z, Jiang J, Stewart DM, Qi S, Yamane K, Li J, et al. AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. Cell Res. 2010;20(3):276–87.
36. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, et al. Histone demethylation by a family of JmjC domain-containing proteins. Nature. 2006;439(7078):811–6.
37. He J, Kallin EM, Tsukada Y, Zhang Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nat Struct Mol Biol. 2008;15(11):1169–75.
38. Frescas D, Guardavaccaro D, Bassermann F, Koyama-Nasu R, Pagano M. JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature. 2007;450(7167):309–13.
39. Janzer A, Stamm K, Becker A, Zimmer A, Buettner R, Kirfel J. The H3K4me3 Histone Demethylase Fbxl10 Is a Regulator of Chemokine Expression, Cellular Morphology, and the Metabolome of Fibroblasts. J Biol Chem. 2012;287(37):30984–92.
40. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010;143(3):470–84.
41. Tanaka Y, Okamoto K, Teye K, Umata T, Yamagiwa N, Suto Y, et al. JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. EMBO J. 2010;29(9):1510–22.
42. Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, et al. KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle. 2008;7(22):3539–47.
43. Bracken AP, Kleine-Kohlbrecher D, Dietrich N, Pasini D, Gargiulo G, Beekman C, et al. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev. 2007;21(5):525–30.
44. Kotake Y, Cao R, Viatour P, Sage J, Zhang Y, Xiong Y. pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK-4alpha tumor suppressor gene. Genes Dev. 2007;21(1):49–54.
45. Tzatsos A, Pfau R, Kampranis SC, Tsichlis PN. Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. Proc Natl Acad Sci USA. 2009;106(8):2641–6.
46. Ge R, Wang Z, Zeng Q, Xu X, Olumi AF. F-box protein 10, an NF-kappaB-dependent anti-apoptotic protein, regulates TRAIL-induced apoptosis through modulating c-Fos/c-FLIP pathway. Cell Death Differ. 2011;18(7):1184–95.

47. Okada Y, Scott G, Ray MK, Mishina Y, Zhang Y. Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. Nature. 2007;450(7166):119–23.
48. Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, et al. PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. Mol Cell. 2007;25(6):801–12.
49. Ma DK, Chiang CH, Ponnusamy K, Ming GL, Song H. G9a and Jhdm2a regulate embryonic stem cell fusion-induced reprogramming of adult neural stem cells. Stem Cells. 2008;26(8):2131–41.
50. Krieg AJ, Rankin EB, Chan D, Razorenova O, Fernandez S, Giaccia AJ. Regulation of the histone demethylase JMJD1A by hypoxia-inducible factor 1 alpha enhances hypoxic gene expression and tumor growth. Mol Cell Biol. 2010;30(1):344–53.
51. Cho HS, Toyokawa G, Daigo Y, Hayami S, Masuda K, Ikawa N, et al. The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. Int J Cancer. 2012;131(3):E179–89.
52. Liu Y, Zheng P, Ji T, Liu X, Yao S, Cheng X, et al. An epigenetic role for PRL-3 as a regulator of H3K9 methylation in colorectal cancer. Gut. 2013;62(4):571–81.
53. Katoh Y, Katoh M. Comparative integromics on JMJD2A, JMJD2B and JMJD2C: preferential expression of JMJD2C in undifferentiated ES cells. Int J Mol Med. 2007;20(2):269–73.
54. Kim SM, Kim JY, Choe NW, Cho IH, Kim JR, Kim DW, et al. Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. Nucleic Acids Res. 2010;38(19):6389–403.
55. Herzog M, Josseaux E, Dedeurwaerder S, Calonne E, Volkmar M, Fuks F. The histone demethylase Kdm3a is essential to progression through differentiation. Nucleic Acids Res. 2012;40(15):7219–32.
56. Cloos PA, Christensen J, Agger K, Maiolica A, Rappsilber J, Antal T, et al. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature. 2006;442(7100):307–11.
57. Fodor BD, Kubicek S, Yonezawa M, O'Sullivan RJ, Sengupta R, Perez-Burgos L, et al. Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. Genes Dev. 2006;20(12):1557–62.
58. Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature. 2006;442(7100):312–6.
59. Mallette FA, Mattiroli F, Cui G, Young LC, Hendzel MJ, Mer G, et al. RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. EMBO J. 2012;31(8):1865–78.
60. Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006;312(5774):748–51.
61. Ozboyaci M, Gursoy A, Erman B, Keskin O. Molecular recognition of H3/H4 histone tails by the tudor domains of JMJD2A: a comparative molecular dynamics simulations study. PLoS One. 2011;6(3):e14765.
62. Tan MK, Lim HJ, Harper JW. SCF(FBXO22) regulates histone H3 lysine 9 and 36 methylation levels by targeting histone demethylase KDM4A for ubiquitin-mediated proteasomal degradation. Mol Cell Biol. 2011;31(18):3687–99.
63. Gray SG, Iglesias AH, Lizcano F, Villanueva R, Camelo S, Jingu H, et al. Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. J Biol Chem. 2005;280(31):28507–18.
64. Black JC, Allen A, Van Rechem C, Forbes E, Longworth M, Tschop K, et al. Conserved antagonism between JMJD2A/KDM4A and HP1gamma during cell cycle progression. Mol Cell. 2010;40(5):736–48.
65. Slee RB, Steiner CM, Herbert BS, Vance GH, Hickey RJ, Schwarz T, et al. Cancer-associated alteration of pericentromeric heterochromatin may contribute to chromosome instability. Oncogene. 2012;31(27):3244–53.

66. Toyokawa G, Cho HS, Iwai Y, Yoshimatsu M, Takawa M, Hayami S, et al. The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. Cancer Prev Res (Phila). 2011;4(12):2051–61.
67. Wissmann M, Yin N, Muller JM, Greschik H, Fodor BD, Jenuwein T, et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nat Cell Biol. 2007;9(3):347–53.
68. Wang J, Zhang M, Zhang Y, Kou Z, Han Z, Chen DY, et al. The histone demethylase JMJD2C is stage-specifically expressed in preimplantation mouse embryos and is required for embryonic development. Biol Reprod. 2010;82(1):105–11.
69. Kim TD, Oh S, Shin S, Janknecht R. Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D. PLoS One. 2012;7(4):e34618.
70. Shin S, Janknecht R. Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. Biochem Biophys Res Commun. 2007;359(3):742–6.
71. Iwamori N, Zhao M, Meistrich ML, Matzuk MM. The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility. Biol Reprod. 2011;84(6):1225–34.
72. Xi ZQ, Sun JJ, Wang XF, Li MW, Liu XZ, Wang LY, et al. HSPBAP1 is found extensively in the anterior temporal neocortex of patients with intractable epilepsy. Synapse. 2007;61(9): 741–7.
73. Chicas A, Kapoor A, Wang X, Aksoy O, Evertts AG, Zhang MQ, et al. H3K4 demethylation by Jarid1a and Jarid1b contributes to retinoblastoma-mediated gene silencing during cellular senescence. Proc Natl Acad Sci USA. 2012;109(23):8971–6.
74. Liefke R, Oswald F, Alvarado C, Ferres-Marco D, Mittler G, Rodriguez P, et al. Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. Genes Dev. 2010;24(6):590–601.
75. Van Oevelen C, Wang J, Asp P, Yan Q, Kaelin Jr WG, Kluger Y, et al. A role for mammalian Sin3 in permanent gene silencing. Mol Cell. 2008;32(3):359–70.
76. Pasini D, Hansen KH, Christensen J, Agger K, Cloos PA, Helin K. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev. 2008;22(10):1345–55.
77. Seiler DM, Rouquette J, Schmid VJ, Strickfaden H, Ottmann C, Drexler GA, et al. Double-strand break-induced transcriptional silencing is associated with loss of trimethylation at H3K4. Chromosome Res. 2011;19(7):883–99.
78. Schmitz SU, Albert M, Malatesta M, Morey L, Johansen JV, Bak M, et al. Jarid1b targets genes regulating development and is involved in neural differentiation. EMBO J. 2011;30(22):4586–600.
79. Wong PP, Miranda F, Chan KV, Berlato C, Hurst HC, Scibetta AG. Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor p21(cip) (CDKN1A). Mol Cell Biol. 2012;32(9):1633–44.
80. Catchpole S, Spencer-Dene B, Hall D, Santangelo S, Rosewell I, Guenatri M, et al. PLU-1/JARID1B/KDM5B is required for embryonic survival and contributes to cell proliferation in the mammary gland and in ER+breast cancer cells. Int J Oncol. 2011;38(5):1267–77.
81. Nijwening JH, Geutjes EJ, Bernards R, Beijersbergen RL. The histone demethylase Jarid1b (Kdm5b) is a novel component of the Rb pathway and associates with E2f-target genes in MEFs during senescence. PLoS One. 2011;6(9):e25235.
82. Xie L, Pelz C, Wang W, Bashar A, Varlamova O, Shadle S, et al. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J. 2011;30(8):1473–84.
83. Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, et al. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature. 2007;447(7144):601–5.
84. Liang Z, Diamond M, Smith JA, Schnell M, Daniel R. Proliferating cell nuclear antigen is required for loading of the SMCX/KMD5C histone demethylase onto chromatin. Epigenetics Chromatin. 2011;4(1):18.

85. Lee MG, Norman J, Shilatifard A, Shiekhattar R. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. Cell. 2007;128(5):877–87.
86. Akimoto C, Kitagawa H, Matsumoto T, Kato S. Spermatogenesis-specific association of SMCY and MSH5. Genes Cells. 2008;13(6):623–33.
87. Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature. 2007;449(7163):731–4.
88. Hong S, Cho YW, Yu LR, Yu H, Veenstra TD, Ge K. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proc Natl Acad Sci USA. 2007;104(47):18439–44.
89. Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen S, et al. A histone H3 lysine 27 demethylase regulates animal posterior development. Nature. 2007;449(7163):689–94.
90. Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, et al. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol. 2007;27(5):1889–903.
91. Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem. 2007;282(28):20395–406.
92. Lee MG, Villa R, Trojer P, Norman J, Yan KP, Reinberg D, et al. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science. 2007;318(5849):447–50.
93. Burgold T, Spreafico F, De Santa F, Totaro MG, Prosperini E, Natoli G, et al. The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. PLoS One. 2008;3(8):e3034.
94. Kim SW, Yoon SJ, Chuong E, Oyolu C, Wills AE, Gupta R, et al. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Dev Biol. 2011;357(2):492–504.
95. Chen S, Ma J, Wu F, Xiong LJ, Ma H, Xu W, et al. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. Genes Dev. 2012;26(12):1364–75.
96. Ye L, Fan Z, Yu B, Chang J, Al Hezaimi K, Zhou X, et al. Histone demethylases KDM4B and KDM6B promotes osteogenic differentiation of human MSCs. Cell Stem Cell. 2012;11(1):50–61.
97. Wen H, Li J, Song T, Lu M, Kan PY, Lee MG, et al. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. J Biol Chem. 2010;285(13):9322–6.
98. Qi HH, Sarkissian M, Hu GQ, Wang Z, Bhattacharjee A, Gordon DB, et al. Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. Nature. 2010;466(7305):503–7.
99. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, Liu WT, et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature. 2010;466(7305):508–12.
100. Feng W, Yonezawa M, Ye J, Jenuwein T, Grummt I. PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat Struct Mol Biol. 2010;17(4):445–50.
101. Zhu Z, Wang Y, Li X, Xu L, Wang X, Sun T, et al. PHF8 is a histone H3K9me2 demethylase regulating rRNA synthesis. Cell Res. 2010;20(7):794–801.
102. Huang C, Xiang Y, Wang Y, Li X, Xu L, Zhu Z, et al. Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4. Cell Res. 2010;20(2):154–65.
103. Huang C, Chen J, Zhang T, Zhu Q, Xiang Y, Chen CD, et al. The dual histone demethylase KDM7A promotes neural induction in early chick embryos. Dev Dyn. 2010;239(12):3350–7.

104. Fortschegger K, De Graaf P, Outchkourov NS, Van Schaik FM, Timmers HT, Shiekhattar R. PHF8 targets histone methylation and RNA polymerase II to activate transcription. Mol Cell Biol. 2010;30(13):3286–98.
105. Baba A, Ohtake F, Okuno Y, Yokota K, Okada M, Imai Y, et al. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. Nat Cell Biol. 2011;13(6):668–75.
106. Kleine-Kohlbrecher D, Christensen J, Vandamme J, Abarrategui I, Bak M, Tommerup N, et al. A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. Mol Cell. 2010;38(2):165–78.
107. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell. 2009;139(7):1303–14.
108. Mysliwiec MR, Carlson CD, Tietjen J, Hung H, Ansari AZ, Lee Y. Jarid2 (Jumonji, AT rich interactive domain 2) regulates NOTCH1 expression via histone modification in the developing heart. J Biol Chem. 2012;287(2):1235–41.
109. Hsia DA, Tepper CG, Pochampalli MR, Hsia EY, Izumiya C, Huerta SB, et al. KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. Proc Natl Acad Sci USA. 2010;107(21):9671–6.
110. Jones MA, Harmer S. JMJD5 Functions in concert with TOC1 in the arabidopsis circadian system. Plant Signal Behav. 2011;6(3):445–8.
111. Ishimura A, Minehata K, Terashima M, Kondoh G, Hara T, Suzuki T. Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. Development. 2012;139(4):749–59.
112. Sinha KM, Yasuda H, Coombes MM, Dent SY, De Crombrugghe B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. EMBO J. 2010;29(1):68–79.
113. Eilbracht J, Reichenzeller M, Hergt M, Schnolzer M, Heid H, Stohr M, et al. NO66, a highly conserved dual location protein in the nucleolus and in a special type of synchronously replicating chromatin. Mol Biol Cell. 2004;15(4):1816–32.
114. Boeckel JN, Guarani V, Koyanagi M, Roexe T, Lengeling A, Schermuly RT, et al. Jumonji domain-containing protein 6 (Jmjd6) is required for angiogenic sprouting and regulates splicing of VEGF-receptor 1. Proc Natl Acad Sci USA. 2011;108(8):3276–81.
115. Lee YF, Miller LD, Chan XB, Black MA, Pang B, Ong CW, et al. JMJD6 is a driver of cellular proliferation and motility and a marker of poor prognosis in breast cancer. Breast Cancer Res. 2012;14(3):R85.
116. Del Rizzo PA, Krishnan S, Trievel RC. Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function. Mol Cell Biol. 2012;32(19):4044–52.
117. Lando D, Peet DJ, Gorman JJ, Whelan DA, Whitelaw ML, Bruick RK. FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. Genes Dev. 2002;16(12):1466–71.
118. Ahmad W, Zlotogorski A, Panteleyev AA, Lam H, Ahmad M, Faiyaz ul Haque M, et al. Genomic organization of the human hairless gene (HR) and identification of a mutation underlying congenital atrichia in an Arab Palestinian family. Genomics. 1999;56(2):141–8.
119. Tsuneoka M, Koda Y, Soejima M, Teye K, Kimura H. A novel myc target gene, mina53, that is involved in cell proliferation. J Biol Chem. 2002;277(38):35450–9.
120. Song C, Chang XJ, Bean KM, Proia MS, Knopf JL, Kriz RW. Molecular characterization of cytosolic phospholipase A2-beta. J Biol Chem. 1999;274(24):17063–7.
121. Huang J, Sengupta R, Espejo AB, Lee MG, Dorsey JA, Richter M, et al. p53 is regulated by the lysine demethylase LSD1. Nature. 2007;449(7158):105–8.
122. Kontaki H, Talianidis I. Lysine methylation regulates E2F1-induced cell death. Mol Cell. 2010;39(1):152–60.
123. Xie Q, Bai Y, Wu J, Sun Y, Wang Y, Zhang Y, et al. Methylation-mediated regulation of E2F1 in DNA damage-induced cell death. J Recept Signal Transduct Res. 2011;31(2):139–46.

124. Chen T, Li E. Establishment and maintenance of DNA methylation patterns in mammals. Curr Top Microbiol Immunol. 2006;301:179–201.
125. Lu T, Jackson MW, Wang B, Yang M, Chance MR, Miyagi M, et al. Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. Proc Natl Acad Sci USA. 2010;107(1):46–51.
126. Ponnaluri VK, Vavilala DT, Putty S, Gutheil WG, Mukherji M. Identification of non-histone substrates for JMJD2A-C histone demethylases. Biochem Biophys Res Commun. 2009;390(2):280–4.
127. Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, et al. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. Science. 2009;325(5936): 90–3.
128. Goldberg JI, Borgen PI. Breast cancer susceptibility testing: past, present and future. Expert Rev Anticancer Ther. 2006;6(8):1205–14.
129. Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell. 1995;80(6):837–45.
130. Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, et al. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet. 2000;24(4):368–71.
131. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci USA. 1999;96(25):14412–7.
132. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23(2):185–8.
133. Cho KS, Elizondo LI, Boerkoel CF. Advances in chromatin remodeling and human disease. Curr Opin Genet Dev. 2004;14(3):308–15.
134. Delaunoy J, Abidi F, Zeniou M, Jacquot S, Merienne K, Pannetier S, et al. Mutations in the X-linked RSK2 gene (RPS6KA3) in patients with Coffin-Lowry syndrome. Hum Mutat. 2001;17(2):103–16.
135. Kahl P, Gullotti L, Heukamp LC, Wolf S, Friedrichs N, Vorreuther R, et al. Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. Cancer Res. 2006;66(23):11341–7.
136. Schulte JH, Lim S, Schramm A, Friedrichs N, Koster J, Versteeg R, et al. Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy. Cancer Res. 2009;69(5):2065–71.
137. Janzer A, Lim S, Fronhoffs F, Niazy N, Buettner R, Kirfel J. Lysine-specific demethylase 1 (LSD1) and histone deacetylase 1 (HDAC1) synergistically repress proinflammatory cytokines and classical complement pathway components. Biochem Biophys Res Commun. 2012;421(4):665–70.
138. Bjorkman M, Ostling P, Harma V, Virtanen J, Mpindi JP, Rantala J, et al. Systematic knockdown of epigenetic enzymes identifies a novel histone demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion. Oncogene. 2012;31(29):3444–56.
139. Xiang Y, Zhu Z, Han G, Ye X, Xu B, Peng Z, et al. JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. Proc Natl Acad Sci USA. 2007;104(49):19226–31.
140. Perinchery G, Sasaki M, Angan A, Kumar V, Carroll P, Dahiya R. Deletion of Y-chromosome specific genes in human prostate cancer. J Urol. 2000;163(4):1339–42.
141. Ehrbrecht A, Muller U, Wolter M, Hoischen A, Koch A, Radlwimmer B, et al. Comprehensive genomic analysis of desmoplastic medulloblastomas: identification of novel amplified genes and separate evaluation of the different histological components. J Pathol. 2006;208(4): 554–63.

142. Vinatzer U, Gollinger M, Mullauer L, Raderer M, Chott A, Streubel B. Mucosa-associated lymphoid tissue lymphoma: novel translocations including rearrangements of ODZ2, JMJD2C, and CNN3. Clin Cancer Res. 2008;14(20):6426–31.
143. Yang ZQ, Imoto I, Fukuda Y, Pimkhaokham A, Shimada Y, Imamura M, et al. Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. Cancer Res. 2000;60(17):4735–9.
144. Lizcano F, Romero C, Vargas D. Regulation of adipogenesis by nuclear receptor PPARgamma is modulated by the histone demethylase JMJD2C. Genet Mol Biol. 2011;34(1):19–24.
145. Magerl C, Ellinger J, Braunschweig T, Kremmer E, Koch LK, Holler T, et al. H3K4 dimethylation in hepatocellular carcinoma is rare compared with other hepatobiliary and gastrointestinal carcinomas and correlates with expression of the methylase Ash2 and the demethylase LSD1. Hum Pathol. 2010;41(2):181–9.
146. Berry WL, Shin S, Lightfoot SA, Janknecht R. Oncogenic features of the JMJD2A histone demethylase in breast cancer. Int J Oncol. 2012;41(5):1701–6.
147. Zeng J, Ge Z, Wang L, Li Q, Wang N, Bjorkholm M, et al. The histone demethylase RBP2 Is overexpressed in gastric cancer and its inhibition triggers senescence of cancer cells. Gastroenterology. 2010;138(3):981–92.
148. Van Zutven LJ, Onen E, Velthuizen SC, Van Drunen E, Von Bergh AR, Van den Heuvel-Eibrink MM, et al. Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. Genes Chromosomes Cancer. 2006;45(5):437–46.
149. Pointon JJ, Harvey D, Karaderi T, Appleton LH, Farrar C, Wordsworth BP. The histone demethylase JARID1A is associated with susceptibility to ankylosing spondylitis. Genes Immun. 2011;12(5):395–8.
150. Lu PJ, Sundquist K, Baeckstrom D, Poulsom R, Hanby A, Meier-Ewert S, et al. A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. J Biol Chem. 1999;274(22):15633–45.
151. Barrett A, Madsen B, Copier J, Lu PJ, Cooper L, Scibetta AG, et al. PLU-1 nuclear protein, which is upregulated in breast cancer, shows restricted expression in normal human adult tissues: a new cancer/testis antigen? Int J Cancer. 2002;101(6):581–8.
152. Tan K, Shaw AL, Madsen B, Jensen K, Taylor-Papadimitriou J, Freemont PS. Human PLU-1 Has transcriptional repression properties and interacts with the developmental transcription factors BF-1 and PAX9. J Biol Chem. 2003;278(23):20507–13.
153. Li Q, Shi L, Gui B, Yu W, Wang J, Zhang D, et al. Binding of the JmjC demethylase JARID1B to LSD1/NuRD suppresses angiogenesis and metastasis in breast cancer cells by repressing chemokine CCL14. Cancer Res. 2011;71(21):6899–908.
154. Smith JA, White EA, Sowa ME, Powell ML, Ottinger M, Harper JW, et al. Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. Proc Natl Acad Sci USA. 2010;107(8):3752–7.
155. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463(7279):360–3.
156. Van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009;41(5):521–3.
157. Tan EY, Campo L, Han C, Turley H, Pezzella F, Gatter KC, et al. Cytoplasmic location of factor-inhibiting hypoxia-inducible factor is associated with an enhanced hypoxic response and a shorter survival in invasive breast cancer. Breast Cancer Res. 2007;9(6):R89.
158. Tsuneoka M, Fujita H, Arima N, Teye K, Okamura T, Inutsuka H, et al. Mina53 as a potential prognostic factor for esophageal squamous cell carcinoma. Clin Cancer Res. 2004;10(21): 7347–56.
159. Zhang Q, Hu CM, Yuan YS, He CH, Zhao Q, Liu NZ. Expression of Mina53 and its significance in gastric carcinoma. Int J Biol Markers. 2008;23(2):83–8.

160. Komiya K, Sueoka-Aragane N, Sato A, Hisatomi T, Sakuragi T, Mitsuoka M, et al. Mina53, a novel c-Myc target gene, is frequently expressed in lung cancers and exerts oncogenic property in NIH/3T3 cells. J Cancer Res Clin Oncol. 2010;136(3):465–73.
161. Xi ZQ, Xiao F, Yuan J, Wang XF, Wang L, Quan FY, et al. Gene expression analysis on anterior temporal neocortex of patients with intractable epilepsy. Synapse. 2009;63(11):1017–28.
162. Hillmer AM, Kruse R, Betz RC, Schumacher J, Heyn U, Propping P, et al. Variant 1859G → A (Arg620Gln) of the "hairless" gene: absence of association with papular atrichia or androgenic alopecia. Am J Hum Genet. 2001;69(1):235–7.
163. Souza PP, Volkel P, Trinel D, Vandamme J, Rosnoblet C, Heliot L, et al. The histone methyltransferase SUV420H2 and Heterochromatin Proteins HP1 interact but show different dynamic behaviours. BMC Cell Biol. 2009;10:41.
164. Lin CH, Li B, Swanson S, Zhang Y, Florens L, Washburn MP, et al. Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Mol Cell. 2008;32(5):696–706.
165. Liang Y, Quenelle D, Vogel JL, Mascaro C, Ortega A, Kristie M. A novel selective LSD1/KDM1A inhibitor epigenetically blocks Herpes Simplex Virus lytic replication and reactivation from latency. MBio. 2013;4(1):e00558-12.
166. Sayegh J, Cao J, Zou MR, Morales A, Blair LP, Norcia M, et al. Identification of small molecule inhibitors of Jumonji AT-Rich Interactive Domain 1B (JARID1B) histone demethylase by a sensitive high-throughput screen. J Biol Chem. 2013;288(13):9408–17.

# Histone Methyltransferase Complexes in Transcription, Development, and Cancer

**Jonathan B. Olsen, Jack Greenblatt, and Andrew Emili**

**Abstract** Dynamic regulation of the mammalian epigenome enables precise control of the developmental gene expression programs that direct stem and progenitor cell proliferation, self-renewal, and differentiation. Among the post-translational modifications that occur on chromatin, histone methylation is a key epigenetic mark with central roles in virtually all DNA-templated processes, including gene transcription by RNA polymerase II (RNAPII). Histone methylation is catalyzed by various histone methyltransferase enzymes, which typically operate within the context of conserved macromolecular complexes. Characterization of the composition and function of histone methyltransferase complexes is critical to understanding the molecular and epigenetic underpinning of cell fate decisions during development. Aberrant histone methylation is frequently observed at the onset and progression of the disease state, originating either directly by inactivating or activating causal mutations that drive pathogenesis or indirectly as facilitators

J.B. Olsen, B.Sc., M.Sc.A. (✉)
Department of Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Rm 930, Toronto, ON, Canada, M5S 3E1
e-mail: jonathan.olsen@utoronto.ca

J. Greenblatt, Ph.D
Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Rm 930, Toronto, ON, Canada, M5S 3E1
e-mail: jack.greenblatt@utoronto.ca

A. Emili, Ph.D
Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, 160 College St., Rm 940, Toronto, ON, Canada, M5S 3E1
e-mail: andrew.emili@utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_2,
© Springer Science+Business Media New York 2014

that perpetuate cancer-related pathways. Here, we review the molecular biology of diverse, often conserved, multicomponent histone methyltransferase complexes with emphasis on the biochemical and physiological roles of these complexes in transcription regulation and chromatin architecture in normal development and human diseases such as cancer.

**Keywords** Chromatin • Histone • Methylation • Methyltransferase • Polycomb • Trithorax • Protein complex • Transcription • Epigenetics • Noncoding RNA

## Introduction

Though it has been known since the 1960s that histone proteins are subject to posttranslational modification [1], the potential impact of these alterations on chromatin structure and function only began to come to light when the crystal structure of the nucleosome was published in 1997 [2]. The main protein components of the nucleosome—an octamer of histone proteins—consists of two key structural features: (1) inter-histone protein interactions via hydrophobic bridges between internal globular domains and (2) long, flexible, basic N-terminal tails that extend outwards from the nucleosomal core. These insights gave rise to mechanistic paradigms that are now widely accepted in chromatin biology; specifically, that histone tail modifications, either alone or in combination, regulate chromatin structure and accessibility, both by altering internucleosomal interactions and by recruiting specific chromatin-modifying (enzymatic) complexes via molecular recognition of modified histone tails [3]. Accordingly, the dynamic regulation of histone posttranslational modification is now seen as a common unifying epigenetic mechanism underlying essentially all DNA-templated processes, including transcription, DNA compaction, replication, and repair.

Notable among the constellation of conserved modifications, which map to at least two-thirds of modifiable tail residues [4], is methylation of ε-amino groups of lysine and ω-guanidino groups of arginine residues by histone methyltransferase enzymes. Relative to other histone posttranslational modifications, lysine and arginine methylation comprise a higher degree of structural complexity; in that these basic side chains can accept multiple methyl groups. Specifically, lysine residues can be enzymatically monomethylated (me1), dimethylated (me2), or trimethylated (me3) and arginine residues monomethylated, symmetrically dimethylated (me2s), and asymmetrically dimethylated (me2a). In contrast to modifications that neutralize or induce changes in the intrinsic histone charge state (e.g., lysine acetylation by lysine acetyltransferases and serine phosphorylation by protein kinases, respectively), lysine and arginine methylation increase the hydrophobic character of the modified side chain without affecting the overall charge. At least in some instances, histone methylation itself can restructure local nucleosomal surfaces or change the orientation of tail regions in vitro [5].

Histone methyltransferases—the so-called writers of histone methylation marks—often exhibit exquisite specificity both in terms of the degree of methylation they generate as well as the recognition sites (e.g., surrounding amino acid sequence of target residues) they modify. This precision in catalytic activity is often reciprocated by diverse recognition repertoires of highly evolved families of methyl-lysine and methyl-arginine-binding domains—the "readers" of histone methylation—often found in large effector multiprotein complexes. These include specific recognition by certain PHD fingers and WD40 repeats, as well as by various modules of the royal superfamily, including tudor domains, chromodomains, chromo barrels, and MBT domains. These effector chromatin-related complexes then either carry out other essential chromatin alterations or else directly regulate various chromatin-based processes (e.g., transcription or heterochromatin formation). Histone methylation is also generally reversible by demethylase enzymes—the "erasers" of histone methylation—although both histone methylation and demethylation can be coordinated to specific methylation patterns on the same or neighboring histones.

The discovery of histone methyltransferases and our understanding of their importance in developmental gene expression have roots in early genetic screens in *Drosophila* aimed at identifying regulators of the position effect variegation (PEV) phenomenon wherein the expression of genes inserted near heterochromatin is silenced. These pioneering studies found suppressors and enhancers of variegation (Su(var) and E(var), respectively) [6]. Other mutants, characterized by defects in body segmentation due to aberrant homeotic (Hox) gene expression, encoded Polycomb group (PcG) and Trithorax group (TrxG) proteins, which repress and promote, respectively, transcription [7]. Genes identified in such screens (e.g., Suppressor of variegation 3–9 Su(var)3-9, Enhancer of zeste (Ez) and Trithorax (Trx)) often encoded proteins with a common 130 amino acid motif known as the SET domain [8]. In 2000, it was demonstrated that the SET domain of Drosophila Suv39h (and its homologs) possesses intrinsic site-specific histone methyltransferase activity [9]. Since then, extensive structural and functional analyses have attributed specific enzymatic functions to a number of proteins harboring lysine and arginine methyltransferase activity. Moreover, macromolecular complexes encompassing histone methyltransferases have been identified in virtually all eukaryotic organisms studied and have been implicated in directing cell fate decisions in different physiological contexts.

Histone methylation is now among the better characterized epigenetic marks known to enable the establishment and maintenance of precise cell- and tissue-specific gene expression programs that are essential for proper metazoan development and cell lineage "memory." Over the past decade, important discoveries and rapid technological advances have impacted our understanding of the dynamic landscape of mammalian epigenomes and its regulation by histone methyltransferase complexes. Here, we review the functional significance of key multicomponent histone methyltransferase complexes in the regulation and proper execution of early developmental gene expression programs, highlighting central paradigms regarding

the composition and activities of annotated histone methyltransferase complexes and providing examples of the relationship between aberrant histone methylation to cancer and other diseases.

## Human Histone Methyltransferases

A number of studies indicate that the functional domains responsible for catalyzing histone methylation are conserved and, in many cases, expanded in mammalian phyla, generally reflecting the size and complexity of the genomes involved. The SET domain family (responsible for all mono-, di-, and tri-methylation of various histone lysine residues, except for H3K79) exhibits a near five-fold expansion in mammals relative to yeast. The variety and number of domains that recognize methylation marks in particular are remarkably increased in humans (PHD fingers: fivefold; royal superfamily domains: 9–14 times; the MBT domain: absent in yeast) [10]. The striking enhancements in the "readers" and "writers" of histone methylation likely reflect a greater need for precise histone methylation patterns to ensure proper gene regulation during development and to define tissue-specific gene expression profiles.

The human genome encodes an estimated 60 enzymes with demonstrated or predicted abilities to methylate histone, lysine, and arginine residues, and these enzymes represent an emerging class of clinically important drug targets [11]. The majority of these enzymes are lysine methyltransferases that contain a characteristic SET domain or an homologous domain bearing significant sequence similarity—except for DOT1L [12], which has a catalytic domain more reminiscent of those found in arginine and unrelated small molecule methyltransferases. Histone methyltransferases make use of a common cofactor, S-adenosyl-L-methionine (SAM), as the methyl donor. Subfamilies of histone methyltransferases, for which up-to-date reviews are available, comprise groups of enzymes with similar protein domain architectures and evolutionary history, and include the following: (1) suppressor of variegation 3–9 (SUV39) [13], (2) SET1/mixed lineage leukemia (MLL) [14], (3) SET2/nuclear receptor SET domain-containing (NSD) [15], (4) retinoblastoma-interacting zinc finger (RIZ)/PR domain-containing (PRDM) [16], (5) SET and MYND domain-containing (MYND), (6) enhancer of zeste (EZ) [17], (7) suppressor of variegation 4–20 (SUV420) [18].

## Histone Methylation and Transcription

The initial connection between histone methylation and active transcription within euchromatin was discovered in the ciliated protozoan *Tetrahymena*, where H3K4 methylation was observed in the transcriptionally active macronucleus but not in

the silent micronucleus [19]. The importance of H3K4 methylation, as well as many other histone methylation marks, to transcription is reflected in recent ChIP-chip and ongoing ChIP-sequencing studies that have revealed genome-wide patterns of methylation marks correlating with transcriptional states and the presence of *cis-regulatory* elements, including enhancer and promoter regions, exons and introns, and various phases of the transcription cycle (e.g., initiation, elongation) [20]. Certain aspects regarding the influence of histone methylation in the regulation of transcription, particularly RNA stalling and the processivity of RNA polymerase II (RNAPII), have been well studied, although many mechanistic details remain unresolved.

Active genes are characterized by histone hyperacetylation (e.g., on various lysine residues of histones H3 and H4) in their promoter regions, as well as trimethylation on H3K4, H3K36, and H3K79 and ubiquitination on H2BK120. The mechanisms leading to the generation of H3K4me3, H3K36me3, and H3K79me3 are co-transcriptional—that is, the histone methyltransferases responsible for these modifications associate physically with RNAPII via the C-terminal domain (CTD) of its large subunit, RPB1. Unique among polymerases, eukaryotic RNAPII contains a large CTD consisting of multiple, sometimes degenerate repeats (52 in humans) of the heptapeptide sequences YSPTSPS. The CTD controls transcription initiation, elongation, and termination and couples transcription to histone modification and various aspects of mRNA processing (e.g., capping, splicing, polyadenylation) [21]. The CTD repeats can be phosphorylated on S2, S5, and S7 [22, 23]. Phosphorylation of S5 is carried out by the CDK subunit of the general transcription factor TFIIH and is strongest near the promoter, although some S5 phosphorylation persists further downstream. Entry into productive elongation by RNAPII is accompanied by phosphorylation of the CTD on S2. S2 phosphorylation is carried out by the CDK9 subunit of the enzyme P-TEFb [24], which is important for release from the general transcription factors located in the promoter.

Phosphorylation of the RNAPII CTD is closely coupled with histone methylation in the transcribed region. CTD phosphorylation at repeat residue S5 is critical for targeting the yeast Set1 (COMPASS) complex to promoter regions, where it di- and tri-methylates histone H3 on lysine 4, as well as analogous complexes in *Drosophila* and human [14]. In addition to S5 phosphorylation, H3K4me3 requires upstream H2B ubiquitination and the PAF complex, which associates with both RNAPII and Set1/COMPASS. Once RNAPII escapes the promoter, H3K79 becomes di- and tri-methylated by the histone methyltransferase DOT1L, which resides in a "super elongation complex" that also contains the CTD S2 kinase P-TEFb [25]. Phosphorylation of the CTD at S2 is in turn recognized by the H3K36 methyltransferase SET2, leading to formation of H3K36me3 throughout the transcribed region [26]. This methylation mark is recognized by the histone deacetylation complex Rpd3S, which assembles deacetylated nucleosomes in the wake of elongating RNAPII and thus helps prevent cryptic transcription initiation [27]. Hence, stepwise physical interactions within a gene body emerge from the establishment of differential histone methylation marks.

## Histone Methylation and Heterochromatin

"Silent" heterochromatin is typically characterized by high levels of H3K9 and H3K27 methylation, as well as DNA methylation, low levels of H3K4 methylation, and generally low levels of histone acetylation. There are two major forms of heterochromatin: facultative and constitutive. Facultative heterochromatin encompasses developmentally regulated genes that are silenced, and constitutive heterochromatin comprises various repetitive sequences located at centromeric and telomeric regions [28]. The generation and maintenance of facultative heterochromatin requires trimethylation of H3K27 by the conserved EZH2 methyltransferase complex known as Polycomb repressive complex 2 (PRC2) [29]. Constitutive heterochromatin requires di- and tri-methylated H3K9 by SUV39H1 and SUV39H2, as well as nucleosomes with pre-monomethylated H3K9 via PRDM3 and PRDM16 [30]. Thus, the partitioning of eukaryotic genomes into heterochromatic regions diverse sets of histone methylation marks by various histone methyltransferase complexes.

## Polycomb Methyltransferase Complexes

PcG and TrxG proteins operate biochemically as distinct, yet evolutionary conserved enzymatic complexes that together ensure the proper spatial and temporal expression of master regulator Hox genes during development [31]. The Hox genes, which are situated contiguously as clusters, encode homeobox transcription factors that regulate diverse cellular signaling pathways involved in development and disease [32]. The PcG histone methyltransferase complex PRC2 generates the H3K27 methylation mark. Several aspects of PRC2 function remain under intense investigation, including the exact composition and functions of mammalian PcG complexes in light of vastly expanded number of paralog subunits, the physical associations that direct and specify PcG to particular loci to facilitate chromosomal reorganization, and the mechanisms whereby PcG complexes prevent elongation by RNAPII at target genes.

PRC2 is composed of four subunits: EZH2/EZH1, SUZ12, EED, and RBBP4/RBBP7. The inner workings of these core components and their contribution to PRC2 function have been extensively characterized [33]. EZH1 and EZH2 harbor SET domains, yet depend critically on inputs from other core PRC2 subunits, particularly EED and SUZ12, for maximal catalytic activity. The WD40 repeat protein EED functions as an allosteric stimulator of PRC2 when bound to H3K27me3 peptides [34], which is thought to enable the maintenance and spreading of silent chromatin [35]. SUZ12 mediates PRC2 complex assembly and also acts as an allosteric activator of EZH2 enzymatic activity [36]. Additional complex components regulate PRC2 activity, including the zinc-finger protein AEBP2, the multiple PCL homologs (PCL1, PCL2, and PCL3), and JARID2 [29]. Table 1 contains a brief summary of PRC2 components and their putative functions.

**Table 1** Common and subtype-specific components of human SET1/MLL complexes and select functions

| Subunit | SET1/MLL complex | Function |
|---|---|---|
| ASH2L | SETD1A/SETD1B, MLL1/MLL2, MLL3/MLL4 | Stimulates methyltransferase activity |
| RBBP5 | SETD1A/SETD1B, MLL1/MLL2, MLL3/MLL4 | Stimulates methyltransferase activity |
| WDR5 | SETD1A/SETD1B, MLL1/MLL2, MLL3/MLL4 | Stimulates methyltransferase activity; Binds unmethylated and methylated H3K4 peptides |
| DPY30 | SETD1A/SETD1B, MLL1/MLL2, MLL3/MLL4 | Stimulates methyltransferase activity |
| WDR82 | SETD1A/SETD1B | Binds S5 phosphorylated RNAPII |
| CXXC1 | SETD1A/SETD1B | Binds unmethylated CpG islands |
| MENIN | MLL1/MLL2 | Targets MLL1 to HOX loci |
| HCFC1 | MLL3/MLL4 | |
| UTX | MLL3/MLL4 | H3K27 demethylase |
| PTIP | MLL3/MLL4 | |
| PA1 | MLL3/MLL4 | |
| NCOA6 | MLL3/MLL4 | Nuclear receptor coactivator |

The expansion of PRC2 complex components (and various PcG components in general) in mammalian systems is thought to enable the assembly of more specialized, yet functionally related complexes. In the case of PRC2, the paralogs EZH1 and EZH2 form mutually exclusive PRC2 complexes of similar global composition. Redundancy in EZH1 and EZH2 function has been reported, such as in ES cells [37] and hair follicle homeostasis and wound repair [38]. However, EZH1 and EZH2 also function nonredundantly in certain cellular contexts, such as during myogenic differentiation where EZH2–PRC2 strongly colocalizes with H3K27me3 versus EZH1–PCR2 with H3K4me3 [39]. Interestingly, EZH1–PRC2 plays a role in stimulating transcription elongation in this context, though the mechanistic basis of this regulation and why EZH1 but not EZH2 is used preferentially remains unclear.

## Trithorax Methyltransferase Complexes

Most TrxG proteins exert their function as part of large multiprotein complexes that regulate transcription via histone methylation or chromatin remodeling. Our current understanding of TrxG histone methyltransferase complexes stems from the biochemical purification and functional characterization of the yeast H3K4 methyltransferase SET1. In *Saccharomyces cerevisiae*, SET1 resides in an 8-member complex known as COMPASS, where, as with PRC2, certain other core COMPASS subunits (e.g., Cps50 and Cps30) are essential for complex methyltransferase activity [14, 40]. COMPASS is recruited to initiated or promoter-proximal paused forms of RNAPII through recognition of phosphorylated CTD at S5 by the Cps35/Swd2 subunit, also known as Wdr82 [41].

**Table 2** Components of human PRC2 and select functions

| Subunit | Function |
|---|---|
| EZH1/EZH2 | Generation of H3K27me2/me3 |
| SUZ12 | Allosteric activator of methyltransferase activity; binding H3K4me3 or H3K36me3 peptides reduces activation |
| EED | Binds H3K27me3 peptides; allosteric activator of methyltransferase activity |
| RBBP4/RBBP7 | Histone chaperones; binds unmodified residues 1–10 of H3 |
| PCL1/PCL2/PCL3 | Binds H3K36me3 peptides via tudor domains |
| AEBP2 | Allosteric regulator of methyltransferase activity; targeting to specific DNA sites |
| JARID2 | Co-recruited with PRC2 to target genes; possible targeting factor (C+G rich) |

Human cells express at least six SET1-related proteins (SETD1A, SETD1B, MLL1, MLL2, MLL3, and MLL4) that form COMPASS-like complexes. Each of these related complexes contains at its core a SET1 homolog as well as four additional core cofactors, ASH2L, RBBP5, WDR5, and DPY30. In contrast to yeast SET1, human SET1 homologs are much larger and contain a diverse array of protein domains, imparting unique functional characteristics. In addition, each COMPASS-like complex contains unique subunits. For instance, SETD1A and SETD1B, which are most closely related to yeast SET1, contain the homolog of the CTD S5-binding subunit Cps35/Swd2/Wdr82, which mediates recruitment to the phosphorylated CTD at S5 [41]. SETD1A and SETD1B COMPASS complexes also uniquely contain the subunit CXXC1, which binds unmethylated CG-rich DNA regions known as CpG islands, thereby regulating the global positioning of H3K4me3 in ES cells [42]. The MLL1 and MLL2 COMPASS-like complexes, which are closely related to *Drosophila* Trx, uniquely contain the tumor suppressor Menin, which is implicated in targeting MLL1 to the promoters of Hox gene loci during hematopoiesis [43], while MLL3 and MLL4 contain the H3K27 demethylase UTX, which helps to reverse gene silencing by PRC2 at HOX gene loci [44]. A summary of the SET1/MLL complex components and their reported functions is found in Table 2.

Understanding the unique composition of human COMPASS-like complexes has shed critical insights regarding the function of these complexes in transcription, development, and disease. This is particularly true for MLL1 because reciprocal chromosomal translocations involving MLL1 and a variety of fusion partners cause human myeloid malignancies such as acute myeloid and lymphoid leukemias [45]. The MLL1 protein is normally cleaved by the protease Taspase-1 into N- and C-terminal fragments that then stably re-associate. The C-terminal region harbors the SET domain as well as a binding interface for the other core COMPASS subunits. The N-terminal fragment is responsible for targeting MLL1 activity to HOX loci via physical association with targeting cofactors such as Menin [43].

In normal hematopoietic progenitor cells, the MLL1 complex maintains the appropriate expression levels of HOX genes, which are essential for self-renewal

and proliferation. However, reciprocal chromosomal translocation creates an MLL1 oncoprotein lacking its native C-terminal enzymatic region. Among the many fusion proteins reported are the elongation factor ELL and other components of a "super elongation complex" that contains the CTD S2 kinase P-TEFb and the H3K79 methyltransferase DOT1L [46]. This oncogenic fusion results in hyperactive HOX gene expression, though the underlying mechanisms mediating this hyperactivity are not fully understood. Importantly, essential roles for DOT1L-mediated H3K79 methylation in HOX gene upregulation in leukemias have been well established, pointing to inhibition of DOT1L enzymatic activity as a promising therapeutic strategy [47, 48]. Thus, aberrant recruitment of DOT1L and H3K79 methylation to HOX gene loci constitutes at least one mechanism at play in MLL1-related leukemias.

## Noncoding RNA and Histone Methylation

Unlike gene activation, which is thought to be targeted to specific loci mostly by site-specific DNA-binding proteins, gene silencing is sometimes, if not usually, targeted by protein complexes containing noncoding RNAs (ncRNAs) as integral components [49]. Indeed, the number of diverse ncRNA transcripts in mammalian genomes putatively outnumbers those of protein-coding transcripts. Small ncRNAs, such as miRNAs, piRNAs, and siRNAs, generated from larger precursors, are incorporated into RITS (RNA-induced transcription silencing) complexes [50], which are thought to be recruited to nascent transcripts via base-pairing interactions [51], leading to polycomb-dependent silencing [52]. In contrast, long intragenic ncRNA (lincRNAs) can directly interact with components of histone methyltransferase complexes and appear to drive them to specific loci. For example, the recruitment of PRC2 to the inactive X chromosome requires recognition of stem–loop structures situated in the lincRNA Xist by the EZH2 subunits of PRC2 [53]. Xist is transcribed by both the active and inactive X chromosomes; however, the antisense transcript of Xist, known as Tsix, is transcribed from the active X chromosome and precludes Xist-PRC2 association through a currently unknown mechanism. The lincRNA ANRIL recruits PRC2 and PRC1 to the INK4b-ARF-INK4a tumor suppressor locus, which plays a central role in cell cycle regulation, senescence, and stress-induced apoptosis [54]. The HOX antisense intergenic RNA (HOTAIR) recruits PRC2 as well as the H3K4 demethylase LSD1 to the silenced HOXD cluster [55]. In addition to targeting silencing complexes to specific chromosomal regions, lincRNAs such as HOTTIP, which emanates from the 5′ end of the HOXA locus, can recruit TrxG complexes across HOXA through direct binding of the Wdr5 subunit, resulting in the maintained activation state of the locus [56].

Additional classes of RNA molecules originating at *cis*-regulatory elements may also play prominent roles in histone methyltransferase recruitment. Enhancer RNAs, which are transcribed from enhancer elements, may open chromatin structure and thereby facilitate enhancer–promoter looping [57]. Short RNAs associated with Polycomb target gene promoters generate stem–loop structures recognized by the

SUZ12 subunit of PRC2 and are thought to facilitate PRC2 recruitment [58]. Systematic characterization of protein–ncRNA interactions and the mechanisms whereby such complexes contribute to the regulation of gene expression and development is currently an area of immense research interest.

## Differential Complex Assembly and Atypical Function

Much evidence has come to light of late that, despite conventional models, histone methyltransferases can affect transcription in disparate manners as subunits of diverse protein complexes depending on cellular context. For example, in castration-resistant prostate cancer cells, EZH2 is phosphorylated on residue S21, which leads to the assembly of a PRC2/H3K27me3-independent transcriptional coactivator complex bound to the androgen receptor that drives an oncogenic gene expression program [59]. Intriguingly, the catalytic activity of EZH2 is required for assembly of this complex, possibly pointing to methylation of non-histone substrates in complex formation. In addition, a transcriptional corepressor complex containing the H3K9 methyltransferase G9a and the H3K4 demethylase Jarid1a is recruited to the β-globin locus during erythropoiesis where it silences the expression of embryonic E(y)-globin gene [60]; yet G9a also forms a distinct complex with Mediator and acts as a coactivator of the adult β(maj)-globin gene [61]. Along these lines, G9a can physically associate with nuclear receptor complexes (e.g., glucocorticoid receptor) to seemingly recruit coactivators such as the H3 arginine methyltransferase CARM1 and the histone acetyltransferase p300 [62]. Collectively, there is a growing body of evidence demonstrating that differential complex assembly can lead to diverse roles of the histone methyltransferases.

## Histone Methylation, Pluripotency, and Differentiation

In ES cells, site-specific DNA-binding transcription factors establish a regulatory network that creates and maintains a pluripotent state, characterized by extensive open chromatin with relatively little heterochromatin [63]. Oct4 and Sox2 are critical transcription factors driving pluripotency [64] and when overexpressed can reprogram human fibroblasts to a pluripotent state [65], whereas Nanog has a secondary role in maintaining pluripotency [66, 67]. These and other master regulators, like Klf4 and Ronin, are thought to recruit multiple chromatin-modifying complexes [68] to establish a cooperative feedback network of both positive and negative regulation that ultimately stabilizes stem cell self-renewal [69].

There is considerable evidence that establishing and maintaining the pluripotent state, as well as exiting pluripotency to various differentiated states, involve extensive chromatin modification. Reprogramming mouse somatic cells to pluripotency can be aided by overexpressing Myc [70], an oncoprotein that globally influences

open chromatin structure, in part by stimulating expression of the histone acetyltransferase GCN5 [71]. Moreover, efficiently creating pluripotent cells and bypassing the need for Myc and Klf4 can be achieved by treating fibroblasts with either an inhibitor specific for the H3K9 methyltransferases G9a (and its paralog GLP) or the HDAC inhibitor valproic acid [65], implying that a transient reversal of silencing is important for inducing pluripotency. Conversely, maintenance of the pluripotent state depends on repression of developmental genes, and this is achieved by the pluripotency transcription factors in conjunction with PcG complexes PRC1 and PRC2 that generate the methylated H3K27 silencing mark and ubiquitinated H2A [31]. The pluripotency transcription factors activate the expression of genes required to maintain pluripotency, including themselves and the H3K9 demethylases Jmjd1a and Jmjd2c [72] that then activate additional cofactors like Tcl1 and Nanog [73].

Long ncRNAs also have important roles in controlling the expression of pluripotency genes through associations specific chromatin-modifying complexes: the expression of multiple long ncRNAs correlates with ES cell pluripotency or differentiation [74], and many ncRNAs bind to chromatin-modifying complexes that repress (e.g., PRC2) or activate (e.g., MLL1) transcription [75]. Several ncRNAs required for the maintenance of pluripotency as well as for differentiation have been recently identified, many of which physically associate with particular chromatin-related protein complexes, including those related to histone methylation [76]. Considerable work still needs to be done to reveal the molecular mechanisms driving the assembly, targeting and function of these and other histone methyltransferase–lincRNA complexes.

A large number of developmental genes and other genes in ES cells are “bivalent” in the sense that they jointly have methylation marks for both repression (H3K27me2/me3) and activation (H3K4me2/me3) states [77]. Upon differentiation of ES cells and silencing of the master regulator Oct4, some bivalent genes resolve into stably inactive genes with H3K27me3 and H3K9 methylation marks after the H3K4 methylation is removed by the demethylase RBP2 (JARID1A) [78]. Others resolve into active genes with H3K4me3 marks upon removal of H3K27 by the demethylases JMJD3 and UTX, the latter of which is recruited along with the H3K4 methyltransferase complex MLL [44].

Long-term silencing of genes that promote the pluripotent/progenitor state requires multifaceted heterochromatin-like silencing systems. For example, inactivation of the Oct4 locus, and other factors that contribute to pluripotency (e.g., Nanog, Dnmt3L), during ES cell differentiation involves H3K9 methylation by a G9a-containing protein complex that also coordinates the recruitment of histone deacetylation and DNA methyltransferase enzymes [79]. ES cells lacking G9a are able to undergo retinoic acid-induced differentiation but maintain the ability to readily revert back to a pluripotency ground state. Analogous roles for G9a in other silencing pathways ensure silencing of other genes critical for pluripotency (e.g., Nanog, Dnmt3L). The exact composition of this G9a-containig complex, as well as the mechanisms underlying its recruitment to the promoters of pluripotency genes such as Oct4 is not fully understood.

## Summary and Future Directions

It is now evident that histone methyltransferase complexes play critical roles in shaping the mammalian epigenome in diverse cellular and physiological contexts. Indeed, efforts to profile mammalian epigenomes in several development and disease contexts are ongoing and will likely reveal informative insights regarding the epigenetic underpinnings of multicellular development, cell fate, and cancer. We have outlined key concepts and emerging paradigms regarding the composition and function of well-established histone methyltransferase complexes; however, further research mapping the protein–protein, protein–RNA, and protein–DNA interaction networks are needed to provide a more comprehensive landscape of the molecular mechanisms that regulate targeting of these enzymes. Moreover, proteomic investigations aimed at characterizing assembly of chromatin complexes and possible methylation of non-histone substrates are of interest. Given the importance of this class of enzyme as an emerging druggable target of therapeutic importance, continuing research conducted along these lines may very well form the basis of future clinical strategies.

## References

1. Allfrey V, Faulkner R, Mirsky A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA. 1964;51:786–94.
2. Luger K, Mäder A, Richmond R, Sargent D, Richmond T. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251–60.
3. Bannister A, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381–95.
4. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
5. Lu X, Simon M, Chodaparambil J, Hansen J, Shokat K, Luger K, et al. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol. 2008;15(10):1122–4.
6. Reuter G, Spierer P. Position effect variegation and chromatin proteins. Bioessays. 1992;14(9): 605–12.
7. Ringrose L, Paro R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet. 2004;38:413–43.
8. Jenuwein T, Laible G, Dorn R, Reuter G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci. 1998;54(1):80–93.
9. Rea S, Eisenhaber F, O'Carroll D, Strahl B, Sun Z, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000;406(6796):593–9.
10. On T, Xiong X, Pu S, Turinsky A, Gong Y, Emili A, et al. The evolutionary landscape of the chromatin modification machinery reveals lineage specific gains, expansions, and losses. Proteins. 2010;78(9):2075–89.
11. Copeland R, Solomon M, Richon V. Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov. 2009;8(9):724–32.
12. Nguyen A, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. Genes Dev. 2011;25(13):1345–58.
13. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev. 2011;25(8):781–8.

14. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012;81:65–95.
15. Wagner E, Carpenter P. Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol. 2012;13(2):115–26.
16. Hohenauer T, Moore A. The Prdm family: expanding roles in stem cells and development. Development. 2012;139(13):2267–82.
17. Morey L, Helin K. Polycomb group protein-mediated repression of transcription. Trends Biochem Sci. 2010;35(6):323–32.
18. Yang H, Mizzen C. The multiple facets of histone H4-lysine 20 methylation. Biochem Cell Biol. 2009;87(1):151–61.
19. Strahl B, Ohba R, Cook R, Allis C. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci USA. 1999;96(26):14967–72.
20. Black J, Van R, Whetstine J. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol Cell. 2012;48(4):491–507.
21. Buratowski S. Progression through the RNA polymerase II CTD cycle. Mol Cell. 2009;36(4):541–6.
22. Kim M, Suh H, Cho E, Buratowski S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J Biol Chem. 2009;284(39):26421–6.
23. Komarnitsky P, Cho E, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 2000;14(19): 2452–60.
24. Peterlin B, Price D. Controlling the elongation phase of transcription with P-TEFb. Mol Cell. 2006;23(3):297–305.
25. Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol. 2012;13(9):543–7.
26. Krogan N, Kim M, Tong A, Golshani A, Cagney G, Canadien V, et al. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol. 2003;23(12):4207–18.
27. Carrozza M, Li B, Florens L, Suganuma T, Swanson S, Lee K, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005;123(4):581–92.
28. Grewal S, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8(1):35–46.
29. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011;469(7330):343–9.
30. Pinheiro I, Margueron R, Shukeir N, Eisold M, Fritzsch C, Richter F, et al. Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell. 2012;150(5):948–60.
31. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. Cell. 2007;128(4):735–45.
32. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. Nat Rev Cancer. 2010;10(5):361–71.
33. O'Meara M, Simon J. Inner workings and regulatory inputs that control Polycomb repressive complex 2. Chromosoma. 2012;121(3):221–34.
34. Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci USA. 2010;107(45):19266–71.
35. Ahn S, Keogh M, Buratowski S. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J. 2009;28(3):205–12.
36. Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell. 2004;15(1):57–67.
37. Shen X, Liu Y, Hsu Y, Fujiwara Y, Kim J, Mao X, et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell. 2008;32(4):491–502.

38. Ezhkova E, Lien W, Stokes N, Pasolli H, Silva J, Fuchs E, et al. EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. Genes Dev. 2011;25(5):485–98.
39. Mousavi K, Zare H, Wang A, Sartorelli V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol Cell. 2012;45(2):255–62.
40. Cosgrove M, Patel A. Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. FEBS J. 2010;277(8):1832–42.
41. Lee J, Skalnik D. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol Cell Biol. 2008;28(2):609–18.
42. Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, et al. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. 2012;26(15):1714–28.
43. Hughes C, Rozenblatt-Rosen O, Milne T, Copeland T, Levine S, Lee J, et al. Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Mol Cell. 2004;13(4):587–97.
44. Agger K, Cloos P, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature. 2007;449(7163):731–4.
45. Marschalek R. Mechanisms of leukemogenesis by MLL fusion proteins. Br J Haematol. 2011;152(2):141–54.
46. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. Genes Dev. 2011;25(7):661–72.
47. Nguyen A, Taranova O, He J, Zhang Y. DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. Blood. 2011;117(25):6912–22.
48. Bernt K, Zhu N, Sinha A, Vempati S, Faber J, Krivtsov A, et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell. 2011;20(1):66–78.
49. Lee J. Epigenetic regulation by long noncoding RNAs. Science. 2012;338(6113):1435–9.
50. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal S, et al. RNAi-mediated targeting of heterochromatin by the RITS complex. Science. 2004;303(5658):672–6.
51. Bühler M, Verdel A, Moazed D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell. 2006;125(5):873–86.
52. Kim D, Villeneuve L, Morris K, Rossi J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol. 2006;13(9):793–7.
53. Jeon Y, Lee J. YY1 tethers Xist RNA to the inactive X nucleation center. Cell. 2011;146(1): 119–33.
54. Aguilo F, Zhou M, Walsh M. Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer Res. 2011;71(16):5365–9.
55. Tsai M, Manor O, Wan Y, Mosammaparast N, Wang J, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689–93.
56. Wang K, Yang Y, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120–4.
57. Orom U, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. Trends Genet. 2011;27(10):433–9.
58. Kanhere A, Viiri K, Araújo C, Rasaiyaah J, Bouwman R, Whyte W, et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell. 2010;38(5):675–88.
59. Xu K, Wu Z, Groner A, He H, Cai C, Lis R, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science. 2012;338(6113):1465–9.
60. Chaturvedi C, Somasundaram B, Singh K, Carpenedo R, Stanford W, Dilworth F, et al. Maintenance of gene silencing by the coordinate action of the H3K9 methyltransferase G9a/KMT1C and the H3K4 demethylase Jarid1a/KDM5A. Proc Natl Acad Sci USA. 2012;109(46): 18845–50.

61. Shankar S, Bahirvani A, Rao V, Bharathy N, Ow J, Taneja R, et al. G9a, a multipotent regulator of gene expression. Epigenetics. 2013;8(1):16–22.
62. Bittencourt D, Wu D, Jeong K, Gerke D, Herviou L, Ianculescu I, et al. G9a functions as a molecular scaffold for assembly of transcriptional coactivators on a subset of glucocorticoid receptor target genes. Proc Natl Acad Sci USA. 2012;109(48):19673–8.
63. Meshorer E, Misteli T. Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol. 2006;7(7):540–6.
64. Avilion A, Nicolis S, Pevny L, Perez L, Vivian N, Lovell-Badge R, et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 2003;17(1):126–40.
65. Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen A, et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol. 2008;26(7):795–7.
66. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, et al. Nanog safeguards pluripotency and mediates germline development. Nature. 2007;450(7173):1230–4.
67. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 2003;113(5):643–55.
68. Dejosez M, Krumenacker J, Zitur L, Passeri M, Chu L, Songyang Z, et al. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell. 2008;133(7):1162–74.
69. Boyer L, Lee T, Cole M, Johnstone S, Levine S, Zucker J, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122(6):947–56.
70. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76.
71. Knoepfler P, Zhang X, Cheng P, Gafken P, McMahon S, Eisenman R, et al. Myc influences global chromatin structure. EMBO J. 2006;25(12):2723–34.
72. Loh Y, Wu Q, Chew J, Vega V, Zhang W, Chen X, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet. 2006;38(4):431–40.
73. Loh Y, Zhang W, Chen X, George J, Ng H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev. 2007;21(20):2545–57.
74. Dinger M, Amaral P, Mercer T, Pang K, Bruce S, Gardiner B, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 2008;18(9):1433–45.
75. Khalil A, Guttman M, Huarte M, Garber M, Raj A, Rivea M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA. 2009;106(28):11667–72.
76. Guttman M, Donaghey J, Carey B, Garber M, Grenier J, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295–300.
77. Bernstein B, Mikkelsen T, Xie X, Kamal M, Huebert D, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315–26.
78. Pasini D, Hansen K, Christensen J, Agger K, Cloos P, Helin K, et al. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev. 2008;22(10):1345–55.
79. Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, et al. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol. 2006;8(2):188–94.

# Chromatin Biology and Cancer Linked Through Protein–Protein Interactions

**Shuye Pu, Andrei L. Turinsky, and Shoshana Wodak**

**Abstract** Up-to-date human protein–protein interaction (PPI) networks for chromatin modification (CM) proteins are constructed and analyzed to explore the functional link between cancer and chromatin-modifying enzymes (CME), such as histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (HMT), histone demethylases (HDM), and DNA-modifying enzymes (DME, including DNA methyltransferases and methylcytosine dioxygenases). In a high-confidence human CM network, extensive interactions (physical associations) are found among CMEs, indicating that CMEs regulate and cooperate with each other to produce complex epigenetic marks. Our results also show that neighbors (interaction partners) of CMEs are enriched not only with proteins involved in transcription (transcription factors and cofactors) but also with proteins coded by oncogenes, tumor suppressor genes, and cancer genes. It is highly likely that products of oncogenes and tumor suppressor genes control gene expression at least in part by regulating the activities of CMEs and that dys-regulation of CMEs plays an important role in tumorigenesis. In addition to drugs targeting CMEs and chromatin readers, drugs targeting process-specific regulators (activators, inhibitors, and recruiters) of CMEs may provide effective and selective alternatives for epigenetic cancer therapy. Identification and characterization of CME regulators should be a top priority in epigenetics and cancer research.

S. Pu, M.D., M.Sc. (✉) • S. Wodak, Ph.D. (✉)
Molecular Structure and Function Program, The Hospital for Sick Children,
555 University Avenue, Toronto, ON, Canada, M5G 1X8
e-mail: shuyepu@sickkids.ca; shoshana@sickkids.ca

A.L. Turinsky, Ph.D., M.Sc.
Research Institute, Molecular Structure and Function Program,
Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada, M5G 1X8
e-mail: turinsky@wodaklab.org

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_3,
© Springer Science+Business Media New York 2014

**Keywords** Chromatin modification • Protein–protein interaction • Transcription • Tumor suppressor • Oncogene • Cancer • Epigenetic therapy

## Introduction

Chromatin, consisting mainly of DNA and nuclear proteins, is the material substance of the eukaryotic genome and epigenome [1]. The fundamental organizational unit of chromatin is the nucleosome, which contains 147 base pairs of DNA wrapped around a complex consisting of eight histone protein subunits (two copies of each of the histones H2A, H2B, H3, and H4). Chromatin modification (CM) is referred to here as comprising the processes of covalent modification of DNA and histones, chromatin remodeling that alters relative position of nucleosome with respect to DNA, and exchange of histone variants in and out of nucleosomes [2]. CM plays a fundamental role in the modulation of DNA-related processes, such as transcription, DNA replication and repair, as well as chromosome organization [2]. CM processes are dynamic (more so for histone modifications than DNA methylation) and highly regulated. Aberrant regulation of genes involved in CM processes, or mutations of these genes, lead to a wide range of pathological conditions, including cancer [3–9]. So far, however, we have limited understanding of CM processes, under both normal and disease conditions. We also know little of the roles of CM proteins in initiation and maintenance of malignancy.

Most cellular processes are carried out by groups of interacting proteins. Information on protein–protein interactions (PPI) has therefore been instrumental in elucidating the function of protein coding genes [10]. Recent progress in high-throughput techniques has enabled large-scale characterization of PPI in yeast [11–15], worm [16], fly [17], and human [18, 19] yielding increasingly more global views of the protein interactions networks in these organisms. In parallel, a growing number of specialized databases have been curating PPI data from publications reporting results from the fast growing number of focused small-scale experiments and making these data available to the scientific community [20–27].

In this chapter, the current status of the human protein interaction network involving CM proteins—the CM PPI network—is reviewed, and this network is analyzed in order to shed light on the functional role of CM proteins in epigenetics and tumorigenesis. The role of noncoding RNAs in chromatin modifications and gene regulation [28] is examined in chapter “Identification of Chromatin-Binding Protein Complexes” of this book.

Up to 16 types of histone modifications have been identified so far [29–31]. Of these, histone methylation and acetylation as well as DNA methylation are the CM processes most commonly disrupted in cancer [30]. Our analysis therefore focuses on protein interactions involving the five types of chromatin-modifying enzymes (CME). These include histone methyltransferases (HMT), demethylases (HDM), histone acetyltransferases (HAT), deacetylases (HDAC), and DNA-modifying enzymes (DME). PPI involving these CME are charted, and the neighborhood of

each of the five enzyme types in the PPI network is functionally characterized. The functional context is derived by analyzing the over-representation—or enrichment—in specific Gene Ontology [32] terms among interaction partners of CME. To explore potential mechanistic links of CMEs with cancer, we analyze the enrichment in oncogenes, tumor suppressors, and other cancer-related genes. Finally, epigenetic therapies targeting these CMEs are briefly discussed.

## Protein-Coding Genes Involved in CM Processes

The protein-coding human genes involved in CM processes are denoted here as "CM genes." We consider the set of human CM genes from the *D*isease *An*notated *C*hromatin *E*pigenetics *R*esource (DAnCER) database [33] (http://wodaklab.org/dancer/). DAnCER contains rich information on CM genes for five model organisms including human. These genes are classified into two categories: the so-called Confirmed CM genes are those with published evidence on their CM role, based on laboratory experiments; whereas "Putative" CM gene currently lack experimental evidences, but their CM role was predicted by bioinformatics methods using sequence homology to, or similar domain composition as, "Confirmed" CM genes [34, 35].

There are currently 637 "Confirmed" and 1,444 "Putative" human CM genes in the DAnCER database. The proteins coded by these CM genes are mainly composed of writers, erasers, and readers of epigenetic marks in DNA and histones [29, 36]. A subset of the CM proteins, the CME, mainly functions as writers and erasers, although some have reader capability as well.

## Construction of PPI Networks Involving CM Proteins

In order to analyze human CM proteins in the context of their protein interaction landscape, we built a human PPI network on the basis of data from public sources. Only data on experimentally characterized physical interactions were considered. These interactions may be direct pairwise interactions determined by methods such as yeast two hybrid (Y2H) and analogous techniques [37, 38] or co-complex associations derived from affinity purification procedures [39]. The co-complex associations represent proteins pairs annotated to the same multiprotein complex.

### *Data Sources*

Three sources of data were used. One is the iRefWeb resource [25] that consolidates PPI data from 14 major public databases each of which curates and archives PPIs from the scientific literature (totaling 60,261 curated publications). The other

sources are two recent single high-throughput proteomics studies, whose data are not yet part of iRefWeb. The first of these is an affinity purification/mass spectrometry analysis of human co-regulator complexes [40]. The raw data from this study (spectral counts) were reprocessed to derive pairwise PPIs by applying the Hypergeometric Spectral Counts score (HGSCore) devised by Guruharsha et al. [17]. A stringent score cutoff (HGSCore > 20) is applied to derive the co-regulator PPI network considered here.

The second study characterized physical associations among human soluble proteins using chromatography-based biochemical fractionation followed by mass spectrometry [18]. We used the published network and scoring scheme but applied a stringent score cutoff (score ≥0.75) to define the co-fractionation PPI network use here.

## *Building the PPI Networks*

The human PPI network consolidated from all three sources comprises 409,223 interactions among 14,727 proteins. Figure 1a, b summarizes the contributions of the individual sources to the consolidated data and illustrates the overlap across these sources as well as the unique contributions of each source. But this consolidated network is likely quite noisy, because much of it is retrieved from public databases, which tend to contain a significant fraction of unreliable PPI [41, 42], and the remainder relies on only two single-study datasets, each suffering from its own biases. We therefore proceeded to build a "high-confidence" human PPI network by pruning the comprehensive network. Using filtering options in iRefWeb (see Legend of Fig. 1) we selected a high-confidence subset of literature curated PPI. This high-confidence subset was supplemented with interactions shared by the unfiltered PPI set from iRefWeb and the two high-throughput studies and those shared by the two studies (Figs 1c, d). The resulting high-confidence consolidated human PPI network was significantly smaller, comprising of 86,429 interactions among 9,107 proteins.

From this high-confidence network we selected all PPIs that involve at least one CM protein (as defined in the previous section) to build the high-confidence CM network used in the present analysis. This network comprises 37,339 interactions (11,647 CM–CM and 25,962 CM–non-CM), encompassing 3,997 proteins. Performing the same operation on the comprehensive PPI network of Fig. 1a, b that has not been pruned, yields a much larger CM network comprising 145,022 interactions among 9,284 human proteins. We verified however that among these unfiltered interactions, only a small fraction (9.0 %) maps into protein complexes defined in the CORUM database [43]. In contrast twice as many PPI (18.6 %) map into CORUM complexes, from the much smaller high-confidence CM network, indicating that this CM network is enriched with stable physical associations. Nevertheless, the fact that more than 80 % of the high-confidence CM interactions fail to map into known protein complexes calls for expanding the landscape of annotated of CM-related complexes, as well as for stepping up their experimental identification.

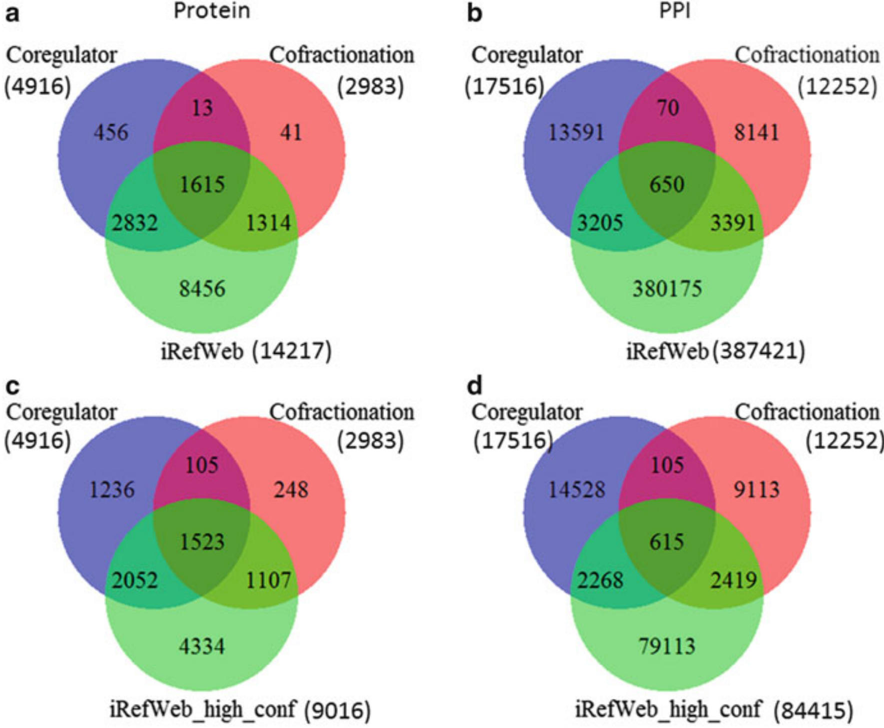



**Fig. 1** Venn diagrams illustrating the overlap between human PPI datasets in terms of proteins (**a** and **c**) and PPIs (**b** and **d**). (**a**) and (**b**): overlap between all pairwise human interactions archived in the iRefWeb with two high-throughput experimental datasets (see [40] for the Coregulator dataset and [18] for the Cofractionation dataset). (**c**) and (**d**): overlap between the high-confidence portion of the iRefWeb human interactions with the Coregulator and Cofractionation datasets. The high-confidence human PPI network from iRefweb was obtained by selecting interactions with a Molecular Interaction (MI) score above 0.43, those reported by at least two publications or those conserved in at least one other organism (see filtering options in iRefWeb: http://www.wodaklab.org/irefweb, for detail)

## *The CM-PPI Network*

A detailed breakdown of the interactions between CM and non-CM proteins, as well as interactions among confirmed and putative CM proteins in the high-confidence CM network are summarized in Fig. 2. Not unexpectedly, confirmed CM proteins are significantly more connected than putative CM proteins (Fig. 2b), given that the ratio of confirmed CM proteins to putative CM proteins is 3:7 (Fig. 2a). This may partly be due to the fact the confirmed CM proteins have been more thoroughly studied. Unless specified otherwise, the analyses of the following sections are all based on the high-confidence CM network.

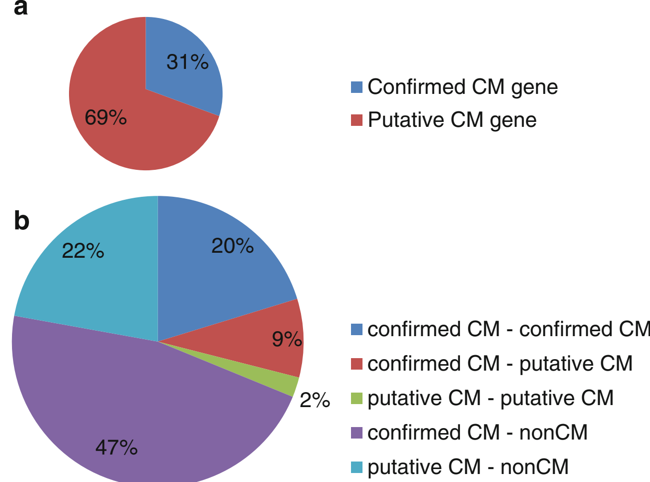



**Fig. 2** The landscape of human CM genes and their protein interaction network. (**a**) Breakdown of human CM genes into “Confirmed” CM genes: those with published evidence on their CM role, based on laboratory experiments; and “Putative” CM gene: those that currently lack experimental evidence, but whose CM role was predicted by bioinformatics methods (see text). (**b**) Breakdown of the high-confidence human CM network. Most interactions occur between a CM protein and a non-CM protein (69 %). Interactions are particular scarce between putative CM proteins (2 %). Confirmed CM proteins consistently participate in more interactions than putative CM proteins in both CM–CM and CM–non CM fractions

## Interactomes of Chromatin-Modifying Enzymes

Histone acetylation and methylation as well as DNA methylation are by and large the most important chromatin modifications [2]. While histone acetylation and H3K4 methylation are linked to transcriptional activation, H3K9 methylation, and H3K27 methylation serve as repressing epigenetic marks. DNA methylation (conversion of cytosine to 5-methylcytosine in the CpG dinucleotide) is responsible for genomic imprinting, X-chromosome inactivation, and tissue-specific gene silencing by promoting heterochromatin formation [1, 2, 44]. We characterize the interactomes of these CMEs by examining interactions among them and by exploring their interaction neighborhoods as defined by the closest neighbors in the network.

Extensive interactions exist among these CMEs, indicating that these CMEs not only modify chromatin but may also modify each other to regulate their enzymatic activities (Fig. 3). For instance, the DNA methyltransferase DNMT1, responsible for maintenance of DNA methylation during replication, can be acetylated on multiple lysines, mainly by KAT2B/PCAF to regulate cell cycle $G_2$/M transition, deacetylated at Lys-1349 and Lys-1415 by SIRT1 to increase methyltransferase activity, and methylated at Lys-142 by SETD7 to promote DNMT1 proteasomal degradation [45]. Alternatively, these interactions may suggest co-complex or recruitment relationships. For example, DNMT1 may form complexes with histone deacetylases HDAC1, HDAC2, and HMTs SUV39H1 and EZH2 [45–49].

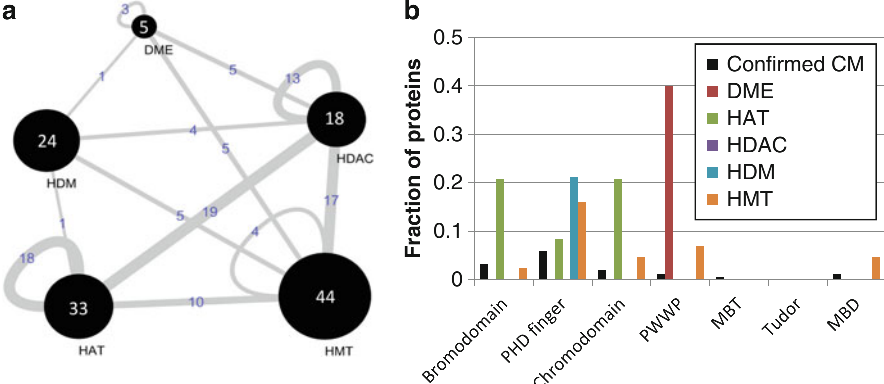



**Fig. 3** (**a**) PPIs that link chromatin-modifying enzymes (CME) in the high-confidence human CM network. The numbers on each edge denote the number of interactions connecting two groups of CME. The size of the nodes is a function of the number of proteins (indicated by the numbers at the center of nodes) belonging to a CME group. *HAT* histone acetyltransferase, *HDAC* histone deacetylase, *HDM* histone demethylase, *HMT* histone methyltransferase, *DME* DNA-modifying enzyme. (**b**) Fractions of CMEs that contain chromatin reader domains. The fractions of confirmed CM proteins that contain these reader domains are included as background references

These co-complex/recruitment relationships are further corroborated by the overlap of their interaction neighborhoods (Fig. 4). Of particular interest are the six proteins that interact with members of all five types of CMEs considered here. Two of these are histones (H1 and H3.3) as expected. The other two are SUMO2 and UBC (poly-ubiquitin chain), indicating that these CMEs are subjected to sumoylation [50] and polyubiquitylation [51]. The fifth one is SIN3A, a member of HDAC complexes and a transcriptional corepressor [52]. And the last protein RB1, is a prototypical tumor suppressor [53] and a regulator of cell cycle and genome stability [54].

This analysis suggests that in addition to SUMO and ubiquitin, SIN3A and RB1 are central regulators or potential substrates of CMEs. Nevertheless, a significant fraction of the proteins in the CME network interact with specific types of CMEs only. This is especially prominent in neighbors of HAT and HDAC (56 % and 45 % HAT and HDAC neighbors interact exclusively with HAT and HDAC, respectively), indicating that these CMEs are regulated/recruited under different conditions.

Another dimension to the functional interplay between different types of CMEs is the enrichments in so-called CM reader domains, including but not limited to bromodomain, PHD finger, chromodomain, PWWP, MBT, Tudor, and MBD [55–57] in CMEs. While bromodomain recognizes histone acetylation marks, chromodomain, PWWP, MBT, and Tudor domains are responsible for recognition of histone methylation marks. The specificity of the small and diverse PHD fingers varies from trimethylated histone to unmodified lysine residues or even acetylated histone. The MBD domain binds methylated DNA. Compared to other CM proteins, HAT are enriched with bromodomain, chromodomain, and PHD finger, which is also abundant among HMT and HDM; whereas PWWP is enriched among DME, MBD is over-represented in HMT (Fig. 3b). By virtue of coexistence of "writer"

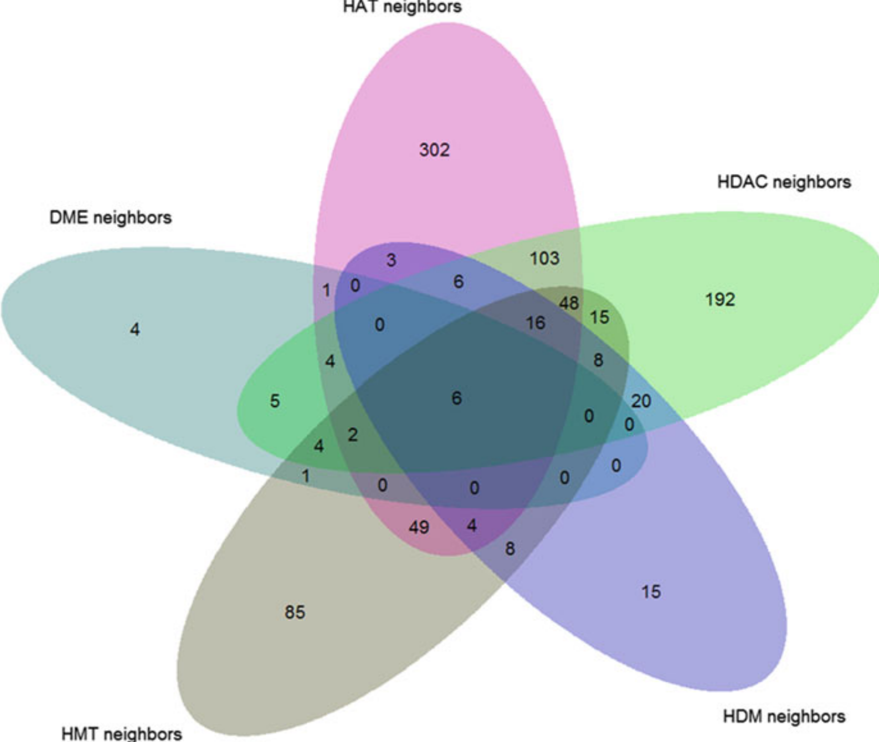



**Fig. 4** Venn diagram illustrating the overlap of the interaction neighbors of different types of CME in the high-confidence human CM network. For each group of CME (HMT, DME, HAT, HDAC, HDM) proteins directly associated (neighbors) with all members of the group were compiled from the high confidence human CM network. The Venn diagram illustrates the extent of overlap between the neighbors of the five CME groups. The different areas delimited by the overlapping ellipsoids are not proportional to the fraction of neighbors

and "reader" domains for the same or different types of chromatin modifications in the same protein, CMEs serve as key links in the transduction of epigenetic regulatory signals. Altogether, the results summarized in Figs. 3 and 4 strongly suggest that CMEs exhibit extensive, multilayer cross talk among themselves in order to produce concerted and complex epigenetic marks.

## CM Networks and Cancer

The Gene Ontology (GO) [32] terms over-represented in the neighborhoods (interaction partners in the high-confidence CM network) of CMEs may shed light on the functional link between CM and cancer. The results of an over-representation (enrichement) analysis using the BiNGO tool [58] are summarized in Fig. 5.

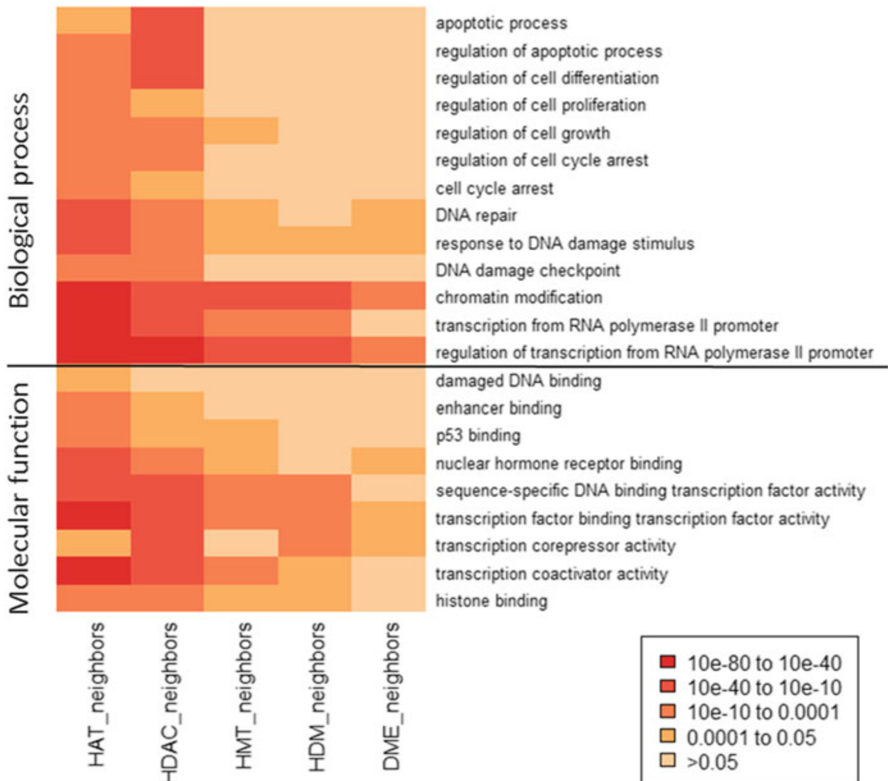



**Fig. 5** A heatmap showing enriched cancer-related GO terms in the interaction neighborhood of CMEs. Colors represent corrected *p*-values (following multi-test correction) as determined using the BiNGO software (see [58]). Significance level is set at $p<0.05$. Not all enriched GO terms are included here; only those related to transcription, chromatin modification, and cancer are listed

In the Biological Process category, neighbors of histone acetylation and methylation enzymes are highly enriched with "chromatin modification," "transcription from RNA polymerase II promoter," and "regulation of transcription from RNA polymerase II promoter." Moreover, neighbors of HAT and HDAC are also enriched with GO terms that are directly related to cancer, such as "apoptotic process," "regulation of apoptotic process," "DNA repair," and so on (see Fig. 5 for details). Accordingly, in the Molecular Function category, neighbors of CMEs are enriched with "sequence-specific DNA binding transcription factor activity" and "transcription factor binding transcription factor activity," suggesting that transcription factors and transcription co-regulators are abundant among the interactors of CMEs. Neighbors of HAT, HDAC, and HMT are also enriched with "p53 binding" activity, consistent with their involvement in "apoptotic process."

Using the approach of Higgins et al. [59], we obtained lists of 1,036 oncogenes and 1,240 tumor suppressors by searching the Entrez gene database [60]. Analysis of the associations of CMEs and neighbors of CMEs with oncogenes and tumor

**Table 1** Enrichment of oncogenes, tumor suppressors, and cancer genes in CM enzymes and their neighbors

| | **Oncogenes** (1036) | | | | **Tumor suppressors** (1240) | | | | **Cancer genes** (2932) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Enzymes** | | **Neighbors** | | **Enzymes** | | **Neighbors** | | **Enzymes** | | **Neighbors** | |
| | Fraction | p-value | Fraction | p-value | Fraction | p-value | Fraction | p-value | Fraction | p-value | Fraction | p-value |
| **DME** (5) (27) | 0.600 | 1.393 E-03 | 0.074 | 0.426 | 0.400 | 3.575 E-02 | 0.222 | 6.192 E-03 | 1.000 | 7.800 E-05 | 0.407 | 1.129 E-03 |
| **HAT** (24) (544) | 0.042 | 0.732 | 0.096 | 3.396 E-05 | 0.083 | 0.459 | 0.142 | 3.236 E-11 | 0.167 | 0.500 | 0.329 | 4.237 E-26 |
| **HDAC** (18) (429) | 0.000 | 1.000 | 0.119 | 6.942 E-08 | 0.278 | 4.467 E-03 | 0.193 | 4.528 E-20 | 0.222 | 0.284 | 0.378 | 2.406 E-31 |
| **HDM** (33) (86) | 0.030 | 0.836 | 0.047 | 0.680 | 0.121 | 0.157 | 0.116 | 4.690 E-02 | 0.152 | 0.570 | 0.267 | 3.613 E-03 |
| **HMT** (44) (246) | 0.068 | 0.419 | 0.077 | 0.068 | 0.227 | 3.640 E-04 | 0.126 | 2.265 E-04 | 0.273 | 2.646 E-02 | 0.313 | 9.042 E-11 |

"Fraction" denotes the fraction of CM enzymes or neighbors that also belong to oncogenes, tumor suppressors, or cancer genes. Fisher's exact test was employed to determine the *p*-value of enrichment. Significance level is set at $p<0.05$. The highlighted values in different colors indicate significantly enriched CMEs or interaction neighbors for different types of CMEs. Figures in the parentheses under column names denote the number of genes/proteins in each group. Figures in the first and second parentheses under row names denote the number of genes/proteins in each CME group and in the neighborhood of each CME group, respectively

suppressor genes, indicates that DME and neighbors of HAT and HDAC are enriched with oncogenes. DME, HDAC, and HMT are enriched with tumor suppressors, which are also enriched in neighbors of all CMEs. This analysis provides further mechanistic link between CMEs and cancer (Table 1).

In the final analysis, we generated a list of 2,932 cancer-related genes using the DAnCER database [33], which archives disease–gene associations derived from Gene2Mesh (http://gene2mesh.ncibi.org/), GAD [61], and PharmKGB [62]. Statistical analysis (Fisher's exact test) indicates that DME and HMT as well as

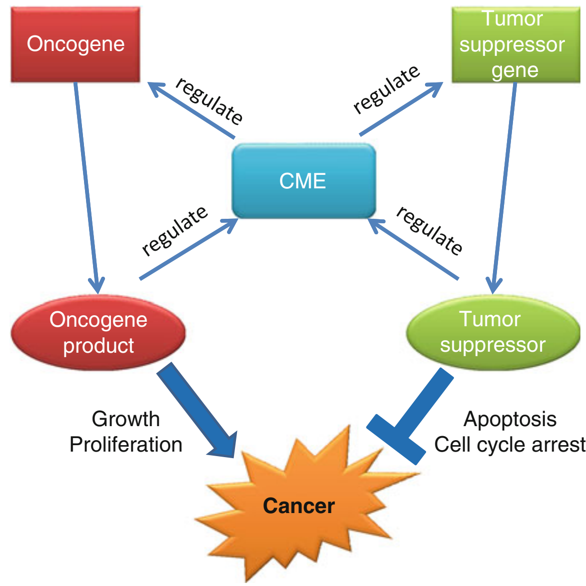



**Fig. 6** A schematic drawing summarizing the relationships among chromatin-modifying enzymes (CME), oncogenes, tumor suppressors, and cancer, as derived from our analysis of the CM-PPI network in human

neighbors of all CMEs are enriched with cancer genes (Table 1). These results suggest that while mutation of some CM enzyme genes per se can cause cancer, as documented by COSMIC [4], perturbation of interactions between CMEs and their neighbors in the PPI network, which we find to be enriched with oncogenes and tumor suppressors, might be a major contributor to tumorigensis as well. It can be envisioned that CME-interacting oncogenes may regulate the expression of tumor suppressors and other genes through interactions with CMEs, and vice versa. As changes in CMEs usually affect a host of genes, the signals carried by CME-interacting oncogenes and tumor suppressors can be profoundly amplified by regulating CMEs. Therefore, by acting as links between oncogenes and tumor suppressors and by acting as signal amplifiers, CMEs occupy strategically critical positions in the cancer-related cellular pathways and thus play a central role in cancer cell biology (Fig. 6).

## CME Interactomes and Epigenetic Therapy

As the role of epigenetics in cancer is increasing appreciated [63], epigenetic therapy has been viewed as a promising new direction for cancer treatment [64–67]. Inhibitors of DNMT (Azacitidine and Decitabine) and HDAC (Vorinostat and Romidepsin) have been approved for clinical use in specific types of cancers

[68–71]. More epigenetic therapies targeting other CMEs (HAT, HMT, HDM, and histone kinases) and chromatin readers (such as the BET family of bromodomain-containing proteins) are being actively developed, with some reaching preclinical and clinical trial stages [30, 65, 72]. As CMEs are involved in regulation of genes with diverse functions, the selectivity of epigenetic inhibitors is critical for drug efficacy and off-target toxicity.

The delineation of CMEs interactomes and the characterization of CMEs-containing protein complexes using high-throughput proteomics approaches (see the chapters by Marcon et al., Olsen et al., Smits and Vermeulen, and Lakshminarasimhan and Washburn combined with identification of downstream genomic targets of CMEs using ChIP-seq techniques, will undoubtedly enhance our understanding of gene regulations that will eventually lead to improvement of selectivity of epigenetic inhibitors. Due to experimental limitations, current PPI networks only represent an ensemble of all interactions detected so far [10]. Availability of tissue-specific and physiological condition-specific CMEs interactomes will be more powerful in defining cancer-specific perturbations, thereby aiding in the discovery of more efficient and more specific epigenetic inhibitors for cancer and other pathological conditions (see the chapter by Richon et al.).

**Acknowledgments** We thank Andrew Emili, Jack Greenblatt, Michael Tyers, John Parkinson, and Zhaolei Zhang as well as members of their teams for many fruitful discussions. We gratefully acknowledge support by the Canadian Institutes of Health Research [MOP#82940], the Ontario Research Fund Global Leadership Program, and the SickKids Foundation. SJW was Canada Research Chair, Tier 1, funded by the Canada Institute of Health Research.

## References

1. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128: 635–8.
2. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128:693–705.
3. Elsasser SJ, Allis CD, Lewis PW. Cancer. New epigenetic drivers of cancers. Science. 2011;331:1145–6.
4. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39:D945–50.
5. Chi P, Allis CD, Wang GG. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer. 2010;10:457–69.
6. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31:27–36.
7. Cross NC. Histone modification defects in developmental disorders and cancer. Oncotarget. 2012;3:3–4.
8. Hake SB, Xiao A, Allis CD. Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer. 2007;96(Suppl):R31–9.
9. Kouzarides T. Wellcome Trust Award Lecture. Chromatin-modifying enzymes in transcription and cancer. Biochem Soc Trans. 2003;31:741–3.
10. Wodak SJ, Pu S, Vlasblom J, Seraphin B. Challenges and rewards of interaction proteomics. Mol Cell Proteomics. 2009;8:3–18.
11. Babu M, Vlasblom J, Pu S, Guo X, Graham C, Bean BD, et al. Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. Nature. 2012;489:585–9.

12. Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, Holstege FC, et al. Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics. 2007;6:439–50.
13. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440:631–6.
14. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440:637–43.
15. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, et al. High-quality binary protein interaction map of the yeast interactome network. Science. 2008;322:104–10.
16. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, et al. A map of the interactome network of the metazoan C. elegans. Science. 2004;303:540–3.
17. Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, et al. A protein complex network of Drosophila melanogaster. Cell. 2011;147:690–703.
18. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, et al. A census of human soluble protein complexes. Cell. 2012;150:1068–81.
19. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;437:1173–8.
20. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 2003;31:248–50.
21. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34:D535–9.
22. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 2002;30:303–5.
23. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct–open source resource for molecular interaction data. Nucleic Acids Res. 2007;35:D561–5.
24. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics. 2008;9:405.
25. Turner B, Razick S, Turinsky AL, Vlasblom J, Crowdy EK, Cho E, et al. iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database (Oxford). 2010;2010:baq023.
26. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, et al. MINT: the Molecular INTeraction database. Nucleic Acids Res. 2007;35:D572–4.
27. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, et al. MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res. 2006;34:D436–41.
28. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43:904–14.
29. Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, et al. New nomenclature for chromatin-modifying enzymes. Cell. 2007;131:633–6.
30. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150:12–27.
31. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21:381–95.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
33. Turinsky AL, Turner B, Borja RC, Gleeson JA, Heath M, Pu S, et al. DAnCER: disease-annotated chromatin epigenetics resource. Nucleic Acids Res. 2011;39:D889–94.
34. On T, Xiong X, Pu S, Turinsky A, Gong Y, Emili A, et al. The evolutionary landscape of the chromatin modification machinery reveals lineage specific gains, expansions, and losses. Proteins. 2010;78:2075–89.
35. Pu S, Turinsky AL, Vlasblom J, On T, Xiong X, Emili A, et al. Expanding the landscape of chromatin modification (CM)-related functional domains and genes in human. PLoS One. 2010;5:e14122.

36. Kouzarides T. SnapShot: Histone-modifying enzymes. Cell. 2007;131:822.
37. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340:245–6.
38. Stagljar I, Korostensky C, Johnsson N, te Heesen S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc Natl Acad Sci USA. 1998;95:5187–92.
39. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods. 2001;24:218–29.
40. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, et al. Analysis of the human endogenous coregulator complexome. Cell. 2011;145:787–99.
41. Turinsky AL, Razick S, Turner B, Donaldson IM, Wodak SJ. Literature curation of protein interactions: measuring agreement across major public databases. Database (Oxford). 2010;2010:baq026.
42. Turinsky AL, Razick S, Turner B, Donaldson IM, Wodak SJ. Interaction databases on the same page. Nat Biotechnol. 2011;29:391–3.
43. Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Stransky M, et al. CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Res. 2008;36:D646–50.
44. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92.
45. Qin W, Leonhardt H, Pichler G. Regulation of DNA methyltransferase 1 by interactions and modifications. Nucleus. 2011;2:392–402.
46. Clements EG, Mohammad HP, Leadem BR, Easwaran H, Cai Y, Van Neste L, et al. DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. Nucleic Acids Res. 2012;40:4334–46.
47. Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 2003;31:2305–12.
48. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet. 2000;25:338–42.
49. Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet. 2000;25:269–77.
50. Lee B, Muller MT. SUMOylation enhances DNA methyltransferase 1 activity. Biochem J. 2009;421:449–61.
51. Oh YM, Kwon YE, Kim JM, Bae SJ, Lee BK, Yoo SJ, et al. Chfr is linked to tumour metastasis through the downregulation of HDAC1. Nat Cell Biol. 2009;11:295–302.
52. Grimes JA, Nielsen SJ, Battaglioli E, Miska EA, Speh JC, Berry DL, et al. The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. J Biol Chem. 2000;275:9461–7.
53. Burkhart DL, Sage J. Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nat Rev Cancer. 2008;8:671–82.
54. Manning AL, Dyson NJ. pRB, a tumor suppressor with a stabilizing presence. Trends Cell Biol. 2011;21:433–41.
55. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci. 2003;28:69–74.
56. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol. 2007;14:1025–40.
57. Wang Y, Fischle W, Cheung W, Jacobs S, Khorasanizadeh S, Allis CD. Beyond the double helix: writing and reading the histone code. Novartis Found Symp. 2004;259:3–17. discussion 17–21, 163–169.

58. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21:3448–9.
59. Higgins ME, Claremont M, Major JE, Sander C, Lash AE. CancerGenes: a gene selection resource for cancer genome projects. Nucleic Acids Res. 2007;35:D721–6.
60. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2011;39:D52–7.
61. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004;36:431–2.
62. Thorn CF, Klein TE, Altman RB. Pharmacogenomics and bioinformatics: PharmGKB. Pharmacogenomics. 2010;11:501–5.
63. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010;19:698–711.
64. Seton-Rogers S. Lymphoma: Epigenetic therapy gains momentum. Nat Rev Cancer. 2012;12:798–9.
65. Rius M, Lyko F. Epigenetic cancer therapy: rationales, targets and drugs. Oncogene. 2012;31:4257–65.
66. Lustberg MB, Ramaswamy B. Epigenetic therapy in breast cancer. Curr Breast Cancer Rep. 2011;3:34–43.
67. Filosa A, Fabiani A. Epigenetic therapy in cancer: perspective and paradoxes. Anal Quant Cytol Histol. 2011;33:303–4.
68. Amato RJ, Stephenson J, Hotte S, Nemunaitis J, Belanger K, Reid G, et al. MG98, a second-generation DNMT1 inhibitor, in the treatment of advanced renal cell carcinoma. Cancer Invest. 2012;30:415–21.
69. Dhawan D., Ramos-Vara JA., Hahn NM, Waddell J, Olbricht GR, Zheng R, Stewart JC, Knapp DW. DNMT1: An emerging target in the treatment of invasive urinary bladder cancer. Urol Oncol 2012. [epub ahead of print]
70. Federico M, Bagella L. Histone deacetylase inhibitors in the treatment of hematological malignancies and solid tumors. J Biomed Biotechnol. 2011;2011:475641.
71. Shankar S, Srivastava RK. Histone deacetylase inhibitors: mechanisms and clinical significance in cancer: HDAC inhibitor-induced apoptosis. Adv Exp Med Biol. 2008;615:261–98.
72. Dawson MA, Kouzarides T, Huntly BJ. Targeting epigenetic readers in cancer. N Engl J Med. 2012;367:647–57.

# Familial and Somatic Mutations of Histone-Modifying Enzymes in Cancer

**Chenyi Liu and Lincoln Stein**

**Abstract** The N-termini of histones are frequent targets for posttranslational modifications. Currently, 16 distinct classes of histone posttranslational modifications have been discovered, but three—acetylation, methylation, and phosphorylation—account for the vast majority of modified sites in the cell. These modifications are reversible and their dynamics are controlled by two classes of enzymes which have the opposing effects of adding and removing chemical groups. Over 150 histone-modifying enzymes (HMEs) have been discovered to date.

Effects of histone modification can be grouped into two categories: (1) effects on the global chromatin environment via the modulation of chromatin conformation and (2) effects on the recruitment of chromatin-associated proteins. Both of these effects are instrumental in regulating chromatin-related processes such as transcription, replication, and DNA damage repair. Dysregulation of histone modification pathways has been associated with an increased risk of cancer. Indeed, mutations in HMEs are now thought to be a common step in the initiation and progression of a variety of cancers.

This chapter reviews histone modifications, the HMEs, the putative functions of histone modification and surveys the evidence for a role of somatic and familial

C. Liu, M.Sc., Ph.D. (✉)
Department of Molecular Genetics, University of Toronto, South Tower,
101 College St., Suite 800, Toronto, Ontraio, Canada, M5G 0A3
e-mail: liuchenyi@hotmail.com

L. Stein, M.D., Ph.D.
Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Informatics and Bio-Computing, Ontario Institute for Cancer Research, Toronto, ON, Canada

Bioinformatics and Genomics, Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY, USA
e-mail: lincoln.stein@gmail.com

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_4,
© Springer Science+Business Media New York 2014

HME mutations in the etiology of cancer. We also review the results of a recent survey of cancer mutation data collected in the Catalogue of Somatic Mutations in Cancer (COSMIC) to reveal some general trends in HME gene mutations.

**Keywords** Histone-modifying enzyme • Histone acetylation • Histone methylation • Histone phosphorylation • Cancer • Somatic mutation • Familial mutation • Sequencing • Epigenetics

## Introduction

Eukaryotic DNA wraps around the core histones H2A, H2B, H3, and H4 (two each) to form nucleosomes, the basic unit of chromatin. Nucleosomes are then further packed into the compact chromosome structure. All the core histones can be post-translationally modified at their N-terminal tails through methylation, acetylation, phosphorylation, and other chemical modifications [1]. These modifications affect both chromatin packing and the ability of other chromatin-associated proteins to bind and are key regulatory steps in the fundamental processes of transcription, DNA replication, and DNA repair [1].

Histone modification is one of the main mechanisms of epigenetics (another being methylation of the DNA), and plays an indispensable role in development, tissue differentiation, and the maintenance of stem cell status [1]. Conversely, abnormal alterations in histone modification patterns have been linked to a wide range of diseases including cancer, neuropsychiatric disorders, inflammation, and metabolic diseases and in some cases have been identified as the causal factor in the etiology of the disease [2]. In recent years, cancer genome-sequencing projects have repeatedly identified frequent mutations in the genes of histone-modifying enzymes (HMEs) which catalyze histone modifications, suggesting that dysregulation of histone modification pathways may play an important role in oncogenesis. This chapter will review histone modifications, HMEs, their functions in normal biological processes, then focus on mutations found in HME genes in the context of cancer and findings related to their causal roles in cancer.

## Histone Modifications and Histone-Modifying Enzymes

Like many other proteins, histones are posttranslationally modified. So far, 16 distinct types of chemical modifications have been documented, including methylation, acetylation, phosphorylation, ubiquitination, deimination, citrullination, SUMOylation, ADP-ribosylation, and isomerization [1]. All the core histone proteins, H2A, H2B, H3, and H4, are modified but the vast majority of modifications are present in H3 and H4 [3]. Among the 16 chemical modifications, acetylation and methylation of lysine residues in histone N-termini are the most common and have been most extensively studied in both human and model species [3, 4].

**Table 1** Histone lysine acetyltransferases [1, 5, 6]

| Gene name | Alternative | Substrate specificity | Function |
|---|---|---|---|
| CDY1 | – | H4 | – |
| CDY1B | – | H4 | – |
| CDY2A | – | H (by similarity) | – |
| CDY2B | – | H | – |
| CDYL | – | H4 | – |
| CLOCK | KAT13D | H3K14, H4 | Transcription activation |
| CREBBP | KAT3A | H2AK5, H2BK12/15, H3K18/27/56, H4K5/8/12/16 | Transcription activation |
| ELP3 | KAT9 | H3K9/18, H4 | – |
| EP300 | KAT3B | H2AK5, H2BK12/15, H3K14/18/27/56, H4K5/8/12/16 | Transcription activation |
| GTF3C4 | KAT12, TFIIIC90 | H3K9/14/18 | Pol III transcription |
| HAT1 | KAT1 | H2AK5, H4K5/12 | Histone deposition, DNA repair |
| KAT2A | hGCN5 | H2B, H3K9/14/18, H4K5/8/12/16/91 | Transcription activation |
| KAT2B | PCAF | H2B, H3K9/14/18, | Transcription activation |
| KAT5 | TIP60, PLIP, HTATIP | H2AK5, H3K14, H4K5/8/12/16 | Transcription activation, DNA repair |
| MGEA5 | – | H3K14, H4K8 | – |
| MYST1 | KAT8, HMOF | H4K16 | Chromatin boundaries, dosage compensation, DNA repair |
| MYST2 | KAT7, HBO1 | H3, H4K5/8/12 | Transcription, DNA replication |
| MYST3 | KAT6A, MOZ | H3K14 | Transcription activation |
| MYST4 | KAT6B, MORF | H3K14 | Transcription activation |
| NAT10 | – | – | – |
| NCOA1 | KAT13A, SRC1 | H3, H4 | Transcription activation |
| NCOA2 | KAT13C, P160 | H3, H4 | Transcription activation |
| NCOA3 | KAT13B, ACTR | H3, H4 | Transcription activation |
| TAF1 | KAT4 | H3, H4 | Transcription activation |

Histone acetylation is reversible and highly dynamic, with the acetyl group transferred from acetyl-CoA to $N^{\varepsilon}$ of lysine residues by acetyltransferases and removed by deacetylases. In human, there are at least 24 annotated (experimentally verified or predicted based on sequence similarity) lysine acetyltransferases (KATs) (Table 1) and 18 deacetylases (HDACs) (Table 2). The KATs can be categorized into two general types. Type A acetyltransferases are capable of modifying chromosomal histones and can be further divided into the three subgroups GNAT (including KAT2A and KAT2B), MYST and EP300/CREBBP families. Type B enzymes predominantly modify free histones in cytoplasm [4].

For their part, histone deacetylases are grouped into four major classes based on sequence homology and substrate specificity: Class I (HDAC1–3 and HDAC8), Class II (HDAC4–7 and HDAC9–10), Class III (SIRT1-7), and Class IV (HDAC11) [7].

**Table 2** Histone deacetylases [1, 5]

| Gene name | Alternative name | Substrate specificity |
|---|---|---|
| SIRT1 | – | H1K25, H2A, H3K9, H4K16 |
| SIRT2 | – | H3K56, K4K16 |
| SIRT3 | – | H3K56 |
| SIRT4 | – | H |
| SIRT5 | – | H |
| SIRT6 | – | H3K9/56 |
| SIRT7 | – | H |
| HDAC1 | – | H3K56 |
| HDAC2 | YAF1 | H |
| HDAC3 | RPD3 | H3K4 |
| HDAC4 | – | H |
| HDAC5 | – | H |
| HDAC6 | – | H |
| HDAC7 | – | H |
| HDAC8 | – | H |
| HDAC9 | – | H |
| HDAC10 | – | H |
| HDAC11 | – | H |

Histone lysine acetylation is a relatively promiscuous process. Multiple acetyltransferases and deacetylases are capable of modifying the same H3 or H4 lysine residue and the same acetyltransferase/deacetylase can typically modify multiple residues. For instance, at least five acetyltransferases (KAT2A, KAT2B, MYST3, MYST4, and GTF3C4) are capable of adding an acetyl group to lysine 14 of histone 3 (H3K14), while three of them (KAT2A, KAT2B, and GTF3C4) are capable of adding an acetyl group to three different lysine residues (H3K9, H3K14, and H3K18).

Histone methylation may occur on either lysine or arginine residues, but lysine methylation has been studied much more intensively than the corresponding arginine modification. So far, methylation on at least seven lysine residues on the core histones has been documented, including H2BK5, H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20. For each lysine, 1–3 methyl groups can be added to become mono, di, and trimethylated lysine (for example, H3K4me1, H3K4me2, and H3K4me3).

Like histone acetylation, histone lysine methylation is reversible and carried out by two opposing classes of enzymes. In human, there are 34 known lysine methyltransferases (KMTs) (Table 3) which can be divided into two classes: SET domain KMTs and non-SET domain KMTs and 24 demethylases (KDMs) (Table 4) which also belong to two groups: LSD1 and JmjC domain KDMs. Both classes of KMTs transfer a methyl group from the donor, S-adenosyl methionine to lysine [8]. However, the two classes of KDMs remove methyl groups by different mechanisms: LSD1 uses FAD as the cofactor to turn the methylated amino group of the lysine into the final products formaldehyde and amine, while JmjC domain KDMs need two cofactors α-ketoglutarate and $Fe^{II}$ to produce the final products formaldehyde and lysine with one methyl group taken off [8].

**Table 3** Histone lysine methyltransferases [1, 5, 6]

| Gene name | Alternative name | Substrate specificity | Function |
|---|---|---|---|
| ASH1L | KMT2H | H3K4, H3K36me1/2 | Transcription activation |
| ASH2L | – | H3K4 | – |
| DOT1L | KMT4 | H3K79me1/2/3 | Transcription activation |
| EHMT1 | KMT1D, GLP | H1K25me1, H1K186me1, H3K9me2, H3K27me1 | Heterochromatin formation/ silencing |
| EHMT2 | KMT1C, G9A | H1K25me1, H1K186me1, H3K9me1/2, H3K27me1 | Heterochromatin formation/ silencing |
| EZH1 | KMT6B | H3K27me1/2/3 | Polycomb silencing, Transcription repression |
| EZH2 | KMT6A | H1K23me1, H3K9, H3K27me1/2/3, | Polycomb silencing, Transcription repression |
| MLL | KMT2A | H3K4, H3K4me3 | Transcription activation |
| MLL2 | KMT2B | H3K4, H3K4me3 | Transcription activation |
| MLL3 | KMT2C | H3K4, H3K4me3 | Transcription activation |
| MLL4 | KMT2D | H3K4, H3K4me3 | Transcription activation |
| MLL5 | KMT2E | H3K4me1/2 | Transcription activation |
| NSD1 | KMT3B | H3K36me2, H4K20me2 | Transcription activation, or repression |
| PRDM2 | KMT8, RIZ1 | H3K9me2 | Transcription repression |
| PRDM8 | – | H3K9me | – |
| PRDM9 | – | H3K4me3 | |
| SETD1A | KMT2F | H3K4me3 | Transcription activation |
| SETD1B | KMT2G | H3K4me3 | Transcription activation |
| SETD2 | KMT3A | H3K36me3 | Transcription activation |
| SETD3 | – | H3K36 | Transcription activation |
| SETD7 | KMT7 | H3K4me1 | Transcription activation |
| SETD8 | KMT5A | H4K20me1/2 | Transcription repression |
| SETDB1 | KMT1E, ESET | H3K9me3 | Transcription repression |
| SETDB2 | KMT1F, CLL8 | H3K9me3 | Transcription repression |
| SETMAR | – | H3K4, H3K36me2 | – |
| SMYD1 | KMT3D | H3K4 | |
| SMYD2 | KMT3C | | Transcription activation |
| SMYD3 | KMT3E | H3K4me2/3 | Transcription activation |
| SUV39H1 | KMT1A | H3K9me3 | Heterochromatin formation/ silencing |
| SUV39H2 | KMT1B | H3K9me3 | Heterochromatin formation/ silencing |
| SUV420H1 | KMT5B | H4K20me2/3 | Transcription repression, DNA damage response |
| SUV420H2 | KMT5C | H4K20me2/3 | Transcription repression |
| WHSC1 | NSD2, MMSET | H3K27me3, H3K36me1/2 | – |
| WHSC1L1 | NSD3 | H3K4me2, H3K27me2/3 | Transcription activation or repression |

**Table 4** Histone lysine demethylases [1, 5, 6]

| Gene name | Alternative name | Substrate specificity | Function |
|---|---|---|---|
| C14orf169 | NO66 | H3K4me3, H3K36me2/3 | – |
| JHDM1D | KDM7 | H3K9me2, H3K27me2, H4K20me1 | – |
| JMJD1C | TRIP8, JHDM2C | H3K9 | – |
| JMJD5 | KDM8 | H3K36me2 | Transcription activation |
| KDM1A | LSD1, BHC110, AOF2 | H3K4me1/2, H3K9me1/2 | Transcription activation and repression, heterochromatin formation |
| KDM1B | AOF1, LSD2 | H3K4me1/2, H3K9me2 | – |
| KDM2A | JHDM1a, FBXL11 | H3K36me1/2 | – |
| KDM2B | JHDM1b, FBXL10 | H3K4me3, H3K36me1/2 | – |
| KDM3A | JHDM2a, JMJD1A | H3K9me1/2 | Androgen receptor gene activation, spermatogenesis |
| KDM3B | JHDM2b, JMJD1B | H3K9me1/2 | – |
| KDM4A | JMJD2A, JHDM3A | H1.4K26me2/3, H3K9me2/3, H3K36me2/3 | Transcription repression, genome integrity |
| KDM4B | JMJD2B | H1.4K26me2/3, H3K9me2/3, H3K36me2/3 | Heterochromatin formation |
| KDM4C | JMJD2C, GASC1 | H1.4K26me2/3, H3K9me2/3, H3K36me2/3 | Putative oncogene |
| KDM4D | JMJD2D | H1K25me1, H1.4K26me2/3, H3K9me2/3, H3K36me2/3 | – |
| KDM5A | JARID1A, RBP2 | H3K4me2/3 | Retinoblastoma-interacting protein |
| KDM5B | JARID1B, PLU-1 | H3K4me1/2/3 | Transcription repression |
| KDM5C | JARID1C, SMCX | H3K4me2/3 | X-linked mental retardation |
| KDM5D | JARID1D/SMCY | H3K4me2/3 | Male-specific antigen |
| KDM6A | UTX | H3K27me2/3 | Transcription activation |
| KDM6B | JMJD3 | H3K27me2/3 | Transcription activation |
| MINA | MDIG | H3K9me3 | Transcription activation of rRNA |
| PHF2 | – | H3K9me2, | Transcription activation |
| PHF8 | – | H3K4me3, H3K9me1/2, H3K27me2, H4K20Me1 | Transcription activation |

Typically more than one methyltransferase and demethylase is capable of modifying the same lysine residue [8]. For example, nine methyltransferases and six demethylases are capable of modifying H3K4. On the other hand, unlike acetyltransferases, most methyltransferases have narrow substrate specificity [8]. For example, all five MLL family members methylate H3K4 only. Some modifiers have even narrower specificity. SETD8 is an example that only catalyzes mono and

**Table 5** Histone arginine methyltransferases and demethylase [1, 5, 6]

| Gene name | Alternative name | Substrate specificity | Function |
|---|---|---|---|
| PRMT1 | ANM1 | H4R3me1/2a | Transcriptional activation |
| PRMT2 | ANM2 | H4 | – |
| CARM1 | PRMT4 | H3R2me1, H3R17me1/2a, H3R26me1 | Transcriptional activation |
| PRMT5 | – | H2AR3, H3R8me2, H4R3me2 | Transcription repression |
| PRMT6 | ANM6 | H2AR3me2, H3R2me2a, H4R3me2 | Transcriptional repression |
| PRMT7 | ANM7 | H2A, H4R3me2s | – |
| PRMT8 | – | – | – |
| JMJD6 | – | H3R2me1/2, H4R3me1/2 | – |

dimethylation of H4K20 [9]. However, others are able to modify two or three different lysine residues, such as EMH1 and EMH2 methylating both H3K9 and H3K27.

There are eight arginine residues on the core histones (H2AR3, H2AR11, H2AR29, H3R2, H3R8, H3R17, H3R26, and H4R3) on which methylation has been reported. Arginines may be mono or dimethylated, and dimethylation may be symmetric with two methyl groups added to two different N atoms, or asymmetric with both methyl groups attached to the same N atom. Compared to the number of lysine methylation modifiers, there are many fewer arginine methylation-modifying enzymes: just seven arginine methyltransferases (PRMTs) (Table 5) and one demethylase JMJD6 have so far been found. Like KMTs, PRMTs also use S-adenosyl methionine as the cofactor for the methyl group transfer. The demethylation by JMJD6 is thought to be similar to JmjC domain KDMs, although the exact molecular mechanism is still unclear [8]. Different PRMTs have distinct methylation patterns which have different biological functions. For example, CARM1, PRMT1, and PRMT6 catalyze asymmetric dimethylation and the resulting arginines serve as transcription-activating marks, while PRMT5 and PRMT7 lead to symmetrically dimethylated arginines as repressive marks [10].

Histone phosphorylation may occur on serine, threonine, tyrosine, and histidine and is controlled by the two opposing classes of enzymes, kinases (phosphate addition), and phosphatases (phosphate removal) [11]. At least 34 kinases (Table 6) and 12 phosphatases (Table 7) have been reported to be involved in histone phosphorylation and most of them modify serine and threonine residues. Many kinases and phosphatases display substrate specificity towards the H2A variant H2AX, which plays an important role in the response and repair of DNA double strand breaks (DSBs) in cells [12].

## Function of Histone Modification

Effects of histone modification may be divided into two general categories: effects on chromatin conformation and compaction and effects on the recruitment of chromatin-associated proteins [1]. The electrostatic attraction between the positive charges on $N^{\varepsilon}$ of lysine residues and the negative charges of DNA keeps DNA tightly bound to the histone core. Histone acetylation on $N^{\varepsilon}$ of lysine residues neutralize a portion of

**Table 6** Histone kinases [1, 5]

| Gene name | Alternative name | Substrate specificity |
|---|---|---|
| ATM | – | H2AS139 |
| ATR | – | H2AS139 |
| AURKB | – | H3S6/10/28 |
| AURKC | – | H3S10/28 |
| BAZ1B | WSTF | H2AY142 |
| BUB1 | – | H2AT120 |
| CDK17 | – | – |
| CDK3 | – | – |
| CDK5 | – | – |
| CHUK | – | H3S10 |
| CSNK2A1 | CKII | H4S1 |
| DAPK3 | – | H3T11 |
| GSG2 | HASPIN | H3T3 |
| GSK3B | – | H1T10 |
| JAK2 | – | H3Y41 |
| LIMK2 | – | – |
| MAP3K8 | – | H3S10 |
| MASTL | – | – |
| NEK6 | – | – |
| NEK9 | – | – |
| PAK2 | – | H3S10 |
| PIM1 | – | H3S10 |
| PKN1 | – | H3T11 |
| PRKCB | – | H3T6 |
| PRKCD | – | H3T45 |
| PRKDC | DNAPK | H2AS139 |
| RPS6KA3 | RSK2 | H2AS16/139 |
| RPS6KA4 | MSK2 | H3S28 |
| RPS6KA5 | MSK1 | H2AS1, H3S10/28 |
| STK10 | – | – |
| STK4 | MST1 | H2AS139, H2BS14 |
| TLK1 | – | H3S10 |
| VRK1 | – | H3S10 |
| ZAK | MLTK | H3S28 |

the positive charge and thus are thought to reduce the electrostatic attraction between histones and DNA, thereby leading to "loosened" or "open" chromatin conformations [1]. These are more accessible to factors involved in DNA-related cellular processes such as transcription, DNA repair, and DNA replication [1]. Phosphorylation of histone adds a negative charge to the histone tail and is thought to alter chromatin structure [11]. Histone methylation, however, does not change either lysine or arginine's charge and has minimal steric effect on chromatin structure due to the small size of the methyl group [8].

Compared to the effects of histone modification on chromatin conformation and compaction, the effects of histone modification on the recruitment of factors involved in various cellular processes have been more extensively studied [1].

**Table 7** Histone phosphatases [1, 5]

| Gene name | Alternative name | Substrate specificity |
|---|---|---|
| DUSP1 | – | H3S10 |
| EYA1 | – | H2AY142 |
| EYA2 | – | H2AY142 |
| EYA3 | – | H2AY142 |
| EYA4 | – | H2AY142 |
| PPM1D | – | H2AS139 |
| PPP1CA | – | H2AFXS140 |
| PPP1CC | – | H3T11 |
| PPP2CA | – | H2AS139 |
| PPP2CB | – | H2AS139 |
| PPP4C | – | H2AS139 |
| PPP5C | – | – |

Although histone methylation has a minimal effect on histone conformation, methylated lysine residues with different number of methyl group can be recognized and differentiated by proteins that contain methylated lysine-binding domains. There are at least five domains that recognize methylated histones: the chromodomain, the plant homeodomain (PHD), TUDOR, the proline–tryptophan–tryptophan–proline domain (PWWP), and the unfortunately-named malicious brain tumor domain (MBT) [1]. The exquisite sensitivity of methylated lysine-binding proteins in recognizing the differences among mono, di and trimethylation provides a high degree of specificity to the binding of chromatin-associated proteins that contain one or more of these domains. In addition to histone methylation, acetylated histones are recognized and bound by bromodomains and PHD and phosphorylated histones by the BRCA-1 C terminus domain (BRCT) [1].

The histone-binding domains can co-occur in proteins with other domains of various functions including DNA-binding (e.g., FYRN and FYRC) and chromatin-related enzymatic activities (e.g., ATP-dependent chromatin-remodeling enzymes and histone modifiers) [13–15], thereby allowing "cross talk" among different histone modifications. For example, repressive mSin3a-HDAC1 deacetylases complex can be recruited to actively transcribing genes through its bound partner ING2, which has an H3K4me3-binding PHD domain, and shut off the transcription [16, 17]. On the other hand, histone modifications may change recognition sites, thus prevent binding by other proteins. For instance, H3K4 methylation occludes the binding of the NuRD complex, which has histone deacetylase activity and is transcription repressive, thereby contributing to transcription activation [18].

## The "Histone Code"

The combination of the effects of histone modifications on chromatin conformation and compaction, plus the ability of histone modifications to regulate the recruitment and binding of non-histone proteins to chromatin, is thought to have profound

**Table 8** Effects of selected histone methylations and acetylations on gene transcription [23–25]

| | Histone modification | | | |
|---|---|---|---|---|
| Amino acid residue | Mono-methylation | Di-methylation | Tri-methylation | Acetylation |
| H2BK5 | Activation | – | Repression | – |
| H3K4 | Activation | – | Activation | – |
| H3K9 | Activation | Repression | Repression | Activation |
| H3K14 | – | – | – | Activation |
| H3K27 | Activation | Repression | Repression | Activation |
| H3K36 | – | – | Activation | – |
| H3K79 | Activation | Activation | Activation | – |
| H4K20 | Activation | – | – | – |

effects on the regulation of transcription, DNA repair, and replication [18]. This has been described as the "histone." However our understanding of the histone code is still rudimentary, and it has not yet been worked out in detail in human or any model organism. This section will summarize what we do know at this point in time.

*Regulation of Gene Expression*. In general, heterochromatin is associated with low levels of acetylation and high levels of methylated H3K9, H3K27, and H4K20. The associations are supported by studies on heterochromatin and transcription suppressive polycomb complex. HP1, an important protein for maintenance of pericentric heterochromatin, binds to methylated H3K9, consistent with the notion that transcription of genes within heterochromatic regions is silenced [1]. PC2, a component of the polycomb complex and recruited by methylated H3K27, is thought to contribute to the inactive state of the X chromosome in human [1]. On the other hand, high levels of acetylation and trimethylated H3K4, H3K36, and H3K79 are typically detected in actively transcribed euchromatin. Consistent with this notion, chromatin immunoprecipitation coupled with sequencing (ChIP-seq) studies show that histone acetylation is present at promoter and enhancer regions of transcriptionally active genes [19, 20].

More recently, results from the comprehensive ENCODE study of functional elements encoded in the human genome provide further evidence for the association between histone modifications and gene transcription. For example, in addition to the fact that H3K4me3 positions precisely around transcription start sites (TSSs), activating marks such as acetylated H3K9 and H3K27, H3K4me3, H3K4me2, H3K79me2, and H3K36me3 correlate positively with transcription and can serve as predictors for transcription, while repressive marks H3K27me3 and H3K9me3 demonstrate negative correlation and are particularly important for accurate prediction of transcription from a subset of promoters [21, 22]. Roles of methylation and acetylation of selected histone sites in transcription regulation are summarized in Table 8.

Findings from stem cells further support the correlations found between histone modifications and transcriptional regulation. In mouse embryonic stem (ES) cells,

both active H3K4me and silencing H3K27me are detected in same domains of development-related genes, while low transcription levels of the genes are detected. It appears that this bivalent histone modification may keep their transcription at a poised state, which may be important for maintenance of pluripotency of these cells. This interpretation is supported by the facts that when ES cells differentiate the poised state is broken where bivalent histone modifications on development-related genes become either H3K4me or H3K27me [26, 27].

Besides the effect of histone acetylation on transcription via modulating global chromatin environment, little is known about how particular acetylated sites contribute to transcription regulation. One exception is the acetylation of H3K56, as the residue faces towards the major groove of the DNA that wraps around the core histones, thus its acetylation has a potential to influence the interaction between the DNA and histones [28].

Phosphorylation of H3S10, controlled by Aurora kinases RPS6KA3/4/5, may play a role in the activation of *c-fos*, *c-jun* during mitosis, and genes regulated by NFKB [29]. Phosphorylated H3S10 may be recognized and bound by the 14-3-3 protein, a conserved protein involved in cell signaling, through phosphoserine-binding modules [30]. In addition, phosphorylation of H3S10 cross talks with other histone modifications to affect their state of modification. For example, H3S10 phosphorylation can increase acetylation of H3K14 [31, 32], eliminate acetylation of H3K9 [33], and inhibit methylation of H3K9 [34], leading to distinct transcription outcomes.

*Regulation of DNA Repair*. Repair of DNA damage is essential for the maintenance of genomic stability. To repair damaged DNA, the compacted chromatin needs first to be relaxed to allow repair proteins to access the DNA damage site, then the repair proteins are recruited to the damage site. Histone modification plays important roles in both steps [3]. Upon DSBs, one type of deleterious DNA lesions, occur, phosphorylation of an H2A variant, H2AX at S139 is triggered and catalyzed by histone kinases *ATM*, *ATR*, or *PRKDC* throughout megabases around the break site. The phosphorylated H2AX then functions as a docking site for recruiting proteins including HATs, resulting in the acetylation of histones and relaxed chromatin which favors access of repair proteins to the DNA lesion [35].

The importance of histone acetylation to repair of DNA damage has been shown by different studies. For instance, KAT5/TIP60, an HAT, is required for recruitment and access of repair proteins to DSBs [36, 37]. Conversely, cells defective in H3 acetylation are sensitive to DNA-damaging agents [38]. Contribution of histone methylation to DNA repair has also been implicated. KAT5/TIP60 can bind to heterochromatic mark H3K9me3 leading to histone acetylation and activation of DNA DSB repair. Although KAT5 may be recruited to DNA damage sites through an alternative route, the interaction with H3K9me3 is required for KAT5 activation and the initiation of subsequent DNA repair [39].

*Regulation of DNA Replication.* During DNA replication, the replication machinery needs to access the DNA, nucleosome structure has to be broken on the parental DNA and re-assembled on the daughter DNA, and the epigenetic state will be rebuilt. Although how histone modifications exactly contribute to DNA replication is still unclear, involvement of histone modification in the regulation of DNA access and nucleosome reorganization have been proposed and links have been made between histone modifications and the regulation of efficiency and timing of the replication origin activity [3]. Indeed, the KAT HBO1 has been shown by several studies to be associated with replication factor MCM2 and with the origin recognition complex 1 subunit of human initiator protein [40, 41], consistent with the proposal that the enhancement of histone acetylation at the replication origin provides a favorable chromatin environment for DNA replication. Histone acetylation has also been reported to correlate with replication timing [42].

## Somatic Mutations of Enzymes and Cancer

The earliest indication that alteration in histone modification pathways might be related to cancer was the finding that *CREBBP,* a KAT enzyme, is able to bind to the adenovirus oncoprotein E1A [43]; this notion was supported by the subsequent detection of recurrent chromosomal translocations in a variety of solid and hematological malignancies [7], resulting in fusion proteins involving KATs and oncoproteins such as MLL-CBP [44] and MOZ-TIF2 [45]. Later, an association of MOZ-TIF2 with leukemia was discovered in murine models [45, 46]. Many KATs have since been implicated in various cancers due to altered expression patterns and somatic mutations, including *CREBBP/KAT3A*, *EP300/KAT3B*, *KAT6A/MOZ*, and *KAT6B/MORF*. Another line of evidence of involvement of HMEs in cancer came from aberrant HME expression in various solid tumors [47].

In general, recurrent mutations are very rare in HDACs [7]. However, chimeric proteins found in leukemia such as PML-RARα, PLZF-RARα, and AML1-ETO are able to recruit HDACs which deacetylate histone leading to abnormal gene silencing and thus contributing to leukemogenesis [48]. Chemical inhibitors of class I HDACs have been tested in patients carrying these chimeric proteins in their malignant cells and in some patients the aberrantly silenced genes were reactivated, leading to cell growth arrest, apoptosis, and differentiation of their tumors [48, 49]. Two such pan-HDAC inhibitors, Vorinostat and Romidepsin, have been approved by FDA for use in cutaneous T cell lymphoma patients [50, 51].

HMEs involved in methylation have also been linked to cancer. PRDM2 is an H3K9-specific KMT and was the first PRDM shown to be tumor suppressor [52, 53] Mutations inactivating *PRDM2* have been found in various types of solid tumors [54]. Conversely, experimental overexpression of *PRDM2* in cancer cell lines results in apoptosis and/or cell cycle arrest, supporting the notion of PRDM2 as a tumor

suppressor [55]. However, the exact molecular mechanism by which PRDM2 functions as a tumor suppressor is still unclear [54].

One of extensively researched HMEs implicated in cancer is EZH2 [7]. EZH2 is the enzymatic component of suppressive PRC2 complex, which has methyltransferase activity specific for mono, di, and trimethylation of H3K27. Elevated *EZH2* expression is correlated with a poor prognosis in prostate and breast cancer [56], suggesting an oncogene role of *EZH2* in the cancers. Heterozygous missense mutations at tyrosine 641 are found in 22 % of diffuse large B-cell lymphoma samples and the mutations increase its catalytic activity for converting H3K27me1 to H3K27me2/3 [57]. However, in myeloid malignancies [58, 59] and T-ALL [60, 61], loss-of-function mutations in the *EZH2* gene are associated with a poor prognosis, suggesting that it can act as a tumor suppressor in other contexts. Inhibitors targeting EZH2 have been designed and tested in lymphoma [62] and leukemia [63] cells.

The *MLL* family of genes has 5 members in human (*MLL1* through *MLL5*), each of which forms the catalytic component of regulatory complexes with H3K4-specific methyltransferase activity associated with actively transcribed genes [64]. The *MLL* genes are frequently mutated in a variety of cancers. For instance, genetic alterations are present in *MLL2* in 89 % of follicular lymphoma and 32 % of diffuse large B-cell lymphoma [65]. Mutations in *MLL2* and *MLL3* are found in 16 % of medulloblastoma cases [66]. *MLL3* is also frequently mutated in pancreatic cancers [67].

The mutations in *MLL2* and *MLL3* are distributed throughout the genes in a pattern typical of loss-of-function alterations in tumor suppressors. Knockout of the *Mll3* catalytic SET domain in *p53*$^{+/-}$ mice leads to the development of spontaneous urothelial tumors, providing evidence for a *p53*-dependent *Mll3* tumor suppressor role [68]. The prototype family member *MLL*, however, differs in mutation patterns and functions, as chromosome translocations are the dominant mutations resulting in MLL fusion proteins that are apparently causal in certain acute lymphoid leukemias and acute myeloid leukemias [69]. Recurrent mutations have also been reported in KDM genes, *KDM5A, KDM5C,* and *KDM6A* [7].

Because of the increasing evidence for involvement of HMEs in cancer, we recently undertook a survey of HME mutations reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) [70] administered by the Wellcome Trust Sanger Institute and collected from a large number of cancer studies published in the primary literature. To get an overview of somatic mutations in HMEs in cancer, we analyzed the mutation data in the most recent version (v61) of the database, which was released on September 26 of 2012. This version includes somatic mutations from 24 large-scale screen papers and 64 whole genome shotgun sequencing projects, in addition to those found in many small-scale studies [71].

We have compiled a list of the 154 HME genes involved in acetylation (42 genes), methylation (66 genes), and phosphorylation (46 genes) described in the previous sections and used this list to analyze the frequency of mutant genes in all the gene groups. We found that among the 19,923 genes collected in the COSMIC, 18,320 genes have at least one non-fusion mutation and the overall mean mutant frequency of all the genes is 9.8 %. However, histone phosphorylation genes (both

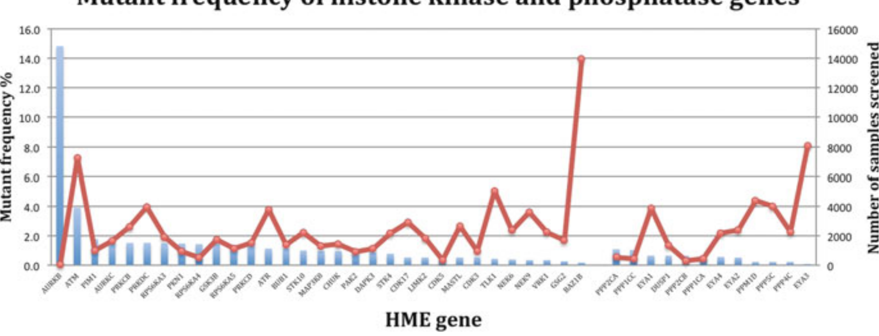



**Fig. 1** Mutant frequencies of histone acetylation genes. HME mutant frequency is calculated as follows: dividing the number of samples that have at least one mutation in the coding region of an HME gene by the number of the samples that are screened/sequenced for the HME gene and then the rate is turned into percentage. The gene of a sample may carry more than one mutation, but in terms of number of mutant HME gene, it is counted as one. The chart shows two groups of HMEs, KATs at *left* and HDACs at *right*, separated by the space

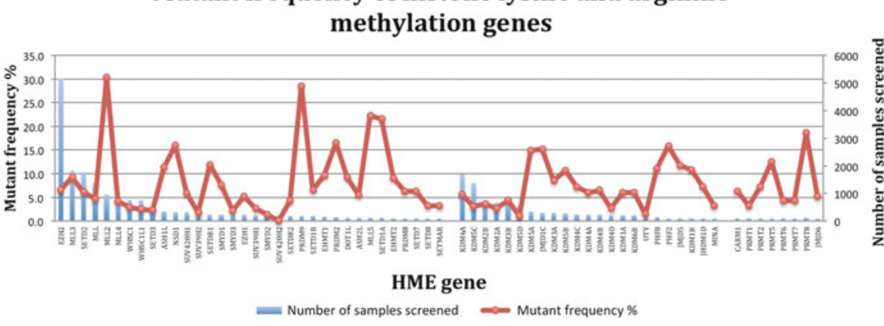



**Fig. 2** Mutant frequencies of histone methylation genes. The calculation of HME mutant frequency is the same as that described in Fig. 1 legend. The chart shows three groups of HMEs, KMTs at *left*, KDMs in the *middle*, and seven RMTs and one RDM (JMJD6) at *right*

kinases and phosphatases) show very low mutant frequencies (mean = 2.5 %) and the majority are under 5 % with four exceptions (*BAZ1B* at 14 %, *EYA3* 8.1 %, *ATM* 7.3 %, and *TLK1* 5.1 %) (Fig. 1). In contrast, many methylation genes (over 25 %) (both KMTs and KDMs) display mutant frequencies above 10 % (mean = 7.9 %) with the highest at 30.4 % (*MLL2*) (Fig. 2). The acetylation gene group demonstrates a somewhat different pattern: mutant frequencies of the majority of KATs are above 5 %, while the majority of HDACs below 5 % with a prominent exception of *HDAC9* at 21 % (mean = 5.1 %) (Fig. 3). The differences among the three groups in mutant frequency are statistically significant, shown by an ANOVA test ($p < 0.0001$) (Fig. 4).

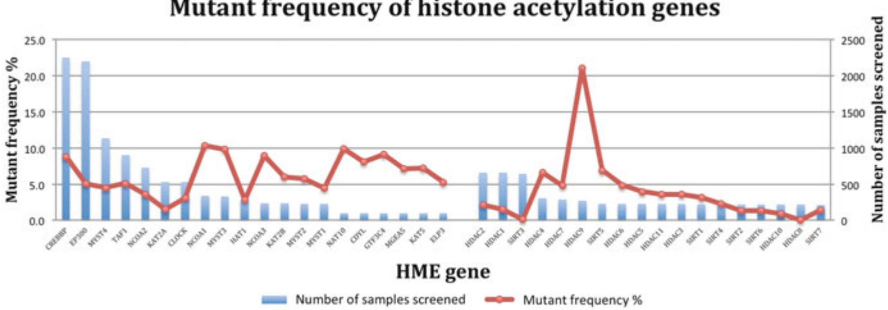
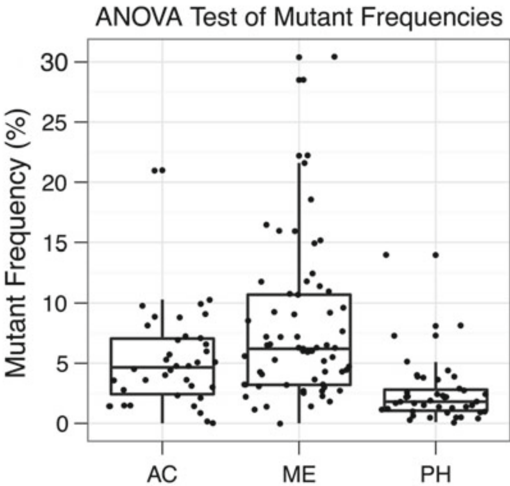



**Fig. 3** Mutant frequencies of histone phosphorylation genes. The calculation of HME mutant frequency is the same as that described in Fig. 1 legend. The chart shows two groups of HMEs, kinases at *left* and phosphatases at *right*. *JAK2* gene is not included in the calculation because the number of mutant *JAK2* in the COSMIC is over 32,000 (with over 99 % of mutations being missense amino acid mutation V617F and the overall mutant frequency is 37.1 %), which, if present in the chart, would make the numbers of all other HME genes look much smaller and their details hard to be seen

**Fig. 4** ANOVA test of mutant frequencies of HME genes. *Notes*: AC stands for HMEs of acetylation, ME for HMEs of methylation, and PH for HMEs of phosphorylation




The trends seen in our analyses are consistent with previous observations [7], for example, low mutant frequencies found in HDACs, but higher in methyltransferases. It is tempting to relate these trends to their substrate specificity (low for HDACs and high for methyltransferases). We speculate that loss-of-function mutations in HDACs, which are highly overlapping in their substrate specificity, have little effect on the biology of the cancer cell due to the extensive redundancy of the system. In contrast, a mutation in a highly specific HME such as a methyltransferase, will have a large impact on the gene(s) that it regulates, and more likely to be subject to positive selection in the early cancer cell.

The HME mutation rate appears to be dependent on specific tumor types. For example, *MLL2* has an overall mutation frequency of 30 %, but the mutation frequency ranges from extremely high in hematopoietic and lymphoid tissues (100 % in 90 samples) and prostate (100 % in 8 samples) to extremely low in liver (0 % from 18 samples) and kidney (2.99 % in 501 samples) [72]. Therefore, our analyses are useful in finding general trends in mutations and in providing guidance for further more focused analyses, especially in the context of accumulation of cancer sequencing data with ever increasing speed.

## Familial Mutations of Enzymes and Cancer

HMEs have been implicated in a few diseases as an inherited factor, including MLL2 in Kabuki syndrome, a pediatric congenital disorder with multiple congenital anomalies and intellectual [73], and HDAC6 in Parkinson's disease, which has both familial and sporadic forms [74], and ATM in ataxia telangiectasia, a rare neurodegenerative, inherited disease causing severe disability [75]. In general, familial cancers are rare. Currently, there is very limited knowledge about roles played by HMEs in familial cancers. Recently, ATM, a kinase responsible for the phosphorylation of H2AS139 has been implicated in familial pancreatic cancer [76], which consists of 5–10 % of all pancreatic cancer cases in the USA. In this study, whole-genome and whole-exome sequencing was used to screen for genetic alterations in 36 subjects from 16 families. Two different heterozygous nonsense mutations in *ATM* were identified in three subjects in each of two families, respectively. Subsequent sequencing of *ATM* in an additional 166 familial pancreatic cancer patients found four more mutations, three of which were considered to be protein function-damaging.

Germ-line *EZH2* mutations identical to some loss-of-function mutations found in myeloid malignancies and T-ALL malignancies have been identified in Weaver syndrome patients, a congenital genetic syndrome associated with rapid growth beginning in the prenatal period and a characteristic facial appearance [77]. However, only two of 19 Weaver patients with *EZH2* mutations develop a malignant disorder.

Germ-line mutations in *CREBBP*, a KAT, appear to be the cause of Rubinstein–Taybi syndrome, a condition characterized by learning difficulties and distinctive facial features, and an association with tumors in some cases [78, 79].

Several other HMEs including MLL2, EP300, and ASXL1 (responsible for deubiquitination of monoubiquitinated H2AK119) have also been implicated in inherited tumors since somatic mutations of these genes have been observed in tumors and similar germline mutations in developmental disorders [80]. In addition, germ-line mutations in protein components of HME complexes have also been linked to familial tumors. MEN1 is an essential component of MLL complex, which catalyzes methylation of H3K4. MEN1 functions as a tumor suppressor and certain

germ-line mutations in *MEN1* resulted in multiple endocrine neoplasia type 1, an inherited tumor syndrome [81]. Because research on HMEs in familial tumors is still in early stage, we may see more to be revealed in future.

## Conclusions and Perspectives

Recurrent and non-recurrent somatic mutations have been identified in most of 150 some HME genes in a wide range of cancers. In the case of 20 genes, the high frequency of mutation suggests a causal or contributory association with the cancer. Some of the mutations have been confirmed to affect functions of the HMEs and to be causative to the cancers. Our analysis of the somatic mutations in HME genes collected in the COSMIC database shows some interesting trends towards low mutation frequencies in histone acetylation genes, and higher frequencies in genes involved in histone methylation.

While the somatic and germline mutation of HMEs is one mechanism for HMEs to contribute to cancer, there are other types of alterations that can also be contributory, for example, alterations in HME gene expression, mutations and changes in expression in genes of components of HME complexes that are essential for the execution of functions of HMEs and in genes coding for proteins that recognize and bind to histone modification sites. There is also complication resulted from overlaps of substrates of histone modification by HMEs. In fact, the situation is further complicated given the fact that many, if not all, of HMEs also modify other proteins in addition to histones, many of which are known oncoproteins and tumor suppressors. Therefore, future efforts on revealing roles played by histone modifications will preferably be designed to examine all related factors involved at different levels including genetics, expression, and function. However, the accelerating production and accumulation of mutation data from large-scale and high-throughput sequencing of cancers will continue to provide highly valuable guidance to more focused research to reveal roles of histone modification in cancer.

## References

1. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
2. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012;11(5):384–400.
3. Sawan C, Herceg Z. Histone modifications and cancer. Adv Genet. 2010;70:57–85.
4. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381–95.
5. Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. HIstome—a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 2012;40(Database issue):D337–42.
6. Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, et al. New nomenclature for chromatin-modifying enzymes. Cell. 2007;131(4):633–6.

7. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012; 150(1):12–27.
8. Spannhoff A, Hauser AT, Heinke R, Sippl W, Jung M. The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors. ChemMedChem. 2009;4(10):1568–82.
9. Julien E, Herr W. A switch in mitotic histone H4 lysine 20 methylation status is linked to M phase defects upon loss of HCF-1. Mol Cell. 2004;14(6):713–25.
10. Di Lorenzo A, Bedford MT. Histone arginine methylation. FEBS Lett. 2011;585(13): 2024–31.
11. Banerjee T, Chakravarti D. A peek into the complex realm of histone phosphorylation. Mol Cell Biol. 2011;31(24):4858–73.
12. Firsanov DV, Solovjeva LV, Svetlova MP. H2AX phosphorylation at the sites of DNA double-strand breaks in cultivated mammalian cells and tissues. Clin Epigenetics. 2011;2(2):283–97.
13. Pray-Grant MG, Daniel JA, Schieltz D, Yates 3rd JR, Grant PA. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature. 2005; 433(7024):434–8.
14. Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006;312(5774):748–51.
15. Pu S, Turinsky AL, Vlasblom J, On T, Xiong X, Emili A, et al. Expanding the landscape of chromatin modification (CM)-related functional domains and genes in human. PLoS One. 2010;5(11):e14122.
16. Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature. 2006; 442(7098):100–3.
17. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature. 2006;442(7098):96–9.
18. Füllgrabe J, Hajji N, Joseph B. Cracking the death code: apoptosis-related histone modifications. Cell Death Differ. 2010;17(8):1238–43.
19. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311–8.
20. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008;40(7):897–903.
21. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75–82.
22. ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74.
23. Benevolenskaya EV. Histone H3K4 demethylases are essential in development and differentiation. Biochem Cell Biol. 2007;85(4):435–43.
24. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37.
25. Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol. 2008;28(8):2825–39.
26. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell. 2005; 120(2):169–81.
27. Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, et al. Chromatin signatures of pluripotent cell lines. Nat Cell Biol. 2006;8(5):532–8.
28. Xu F, Zhang K, Grunstein M. Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell. 2005;121:375–85.
29. Mahadevan LC, Willis AC, Barratt MJ. Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. Cell. 1991; 65(5):775–83.

30. Macdonald N, Welburn JP, Noble ME, Nguyen A, Yaffe MB, Clynes D, et al. Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol Cell. 2005;20(2):199–211.
31. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell. 2000;5:905–15.
32. Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, et al. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell. 2000;5:917–26.
33. Edmondson DG, Davie JK, Zhou J, Mirnikjoo B, Tatchell K, Dent SY. Site-specific loss of acetylation upon phosphorylation of histone H3. J Biol Chem. 2002;277:29496–502.
34. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000;406:593–9.
35. Downs JA, Allard S, Jobin-Robitaille O, Javaheri A, Auger A, Bouchard N, et al. Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. Mol Cell. 2004;16:979–90.
36. Loizou JI, Murr R, Finkbeiner MG, Sawan C, Wang ZQ, Herceg Z. Epigenetic information in chromatin: The code of entry for DNA repair. Cell Cycle. 2006;5:696–701.
37. Murr R, Loizou JI, Yang YG, Cuenin C, Li H, Wang ZQ, et al. Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. Nat Cell Biol. 2006;8:91–9.
38. Masumoto H, Hawke D, Kobayashi R, Verreault A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature. 2005;436:294–8.
39. Sun Y, Jiang X, Xu Y, Ayrapetov MK, Moreau LA, Whetstine JR, et al. Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. Nat Cell Biol. 2009;11:1376–82.
40. Burke TW, Cook JG, Asano M, Nevins JR. Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. J Biol Chem. 2001;276:15397–408.
41. Iizuka M, Stillman B. Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J Biol Chem. 1999;274:23027–34.
42. Cimbora DM, Schubeler D, Reik A, Hamilton J, Francastel C, Epner EM, et al. Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region. Mol Cell Biol. 2000;20:5581–91.
43. Bannister AJ, Kouzarides T. The CBP co-activator is a histone acetyltransferase. Nature. 1996;384:641–3.
44. Wang J, Iwasaki H, Krivtsov A, Febbo PG, Thorner AR, Ernst P, et al. Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease. EMBO J. 2005; 24:368–81.
45. Huntly BJ, Shigematsu H, Deguchi K, Lee BH, Mizuno S, Duclos N, et al. MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. Cancer Cell. 2004;6:587–96.
46. Deguchi K, Ayton PM, Carapeti M, Kutok JL, Snyder CS, Williams IR, et al. MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. Cancer Cell. 2003;3:259–71.
47. Ozdag H, Teschendorff AE, Ahmed AA, Hyland SJ, Blenkiron C, Bobrow L, et al. Differential expression of selected histone modifier genes in human solid cancers. BMC Genomics. 2006;7:90.
48. Johnstone RW, Licht JD. Histone deacetylase inhibitors in cancer therapy: is transcription the primary target? Cancer Cell. 2003;4:13–8.
49. Federico M, Bagella L. Histone deacetylase inhibitors in the treatment of hematological malignancies and solid tumours. J Biomed Biotechnol. 2011;2011:475641.
50. Olsen EA, Kim YH, Kuzel TM, Pacheco TR, Foss FM, Parker S, et al. Phase IIb multicenter trial of vorinostat in patients with persistent, progressive, or treatment refractory cutaneous T-cell lymphoma. J Clin Oncol. 2007;25:3109–15.

51. Piekarz RL, Frye R, Turner M, Wright JJ, Allen SL, Kirschbaum MH, et al. Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. J Clin Oncol. 2009;27:5410–7.
52. Steele-Perkins G, Fang W, Yang XH, Van Gele M, et al. Tumour formation and inactivation of RIZ1, an Rb-binding member of a nuclear protein-methyltransferase superfamily. Genes Dev. 2001;15:2250–62.
53. Kim KC, Geng L, Huang S. Inactivationofahistonemethyltrans- ferase by mutations in human cancers. Cancer Res. 2003;63:7619–23.
54. Fog CK, Galli GG, Lund AH. PRDM proteins: important players in differentiation and disease. Bioessays. 2012;34(1):50–60.
55. Pastural E, Takahashi N, Dong WF, Bainbridge M, et al. RIZ1 repression is associated with insulin-like growth factor-1 signaling activation in chronic myeloid leukemia cell lines. Oncogene. 2007;26:1586–94.
56. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011;469:343–9.
57. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010;42:181–5.
58. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet. 2010; 42:722–6.
59. Nikoloski G, Langemeijer SM, Kuiper RP, Knops R, Massop M, Tonnissen ER, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat Genet. 2010;42:665–7.
60. Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med. 2012;18:298–301.
61. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481:157–63.
62. Knutson SK, Wigle TJ, Warholic NM, Sneeringer CJ, Allain CJ, Klaus CR, et al. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nat Chem Biol. 2012;8(11):890–6.
63. Valente S, Lepore I, Dell'aversana C, Tardugno M, Castellano S, Sbardella G, et al. Identification of PR-SET7 and EZH2 selective inhibitors inducing cell death in human leukemia U937 cells. Biochimie. 2012;94(11):2308–13.
64. Mandal SS. Mixed lineage leukemia: versatile player in epigenetics and human disease. FEBS J. 2010;277:1789.
65. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476(7360):298–303.
66. Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, et al. The genetic landscape of the childhood cancer medulloblastoma. Science. 2011;331(6016):435–9.
67. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science. 2008; 321(5897):1801–6.
68. Lee J, Kim DH, Lee S, Yang QH, Lee DK, Lee SK, et al. A tumour suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proc Natl Acad Sci USA. 2009;106(21):8513–8.
69. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer. 2007;7:823–33.
70. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39(Database issue):D945–50.

71. The Catalogue of Somatic Mutations in Cancer (COSMIC) database administered by the Wellcome Trust Sanger Institute. Data were downloaded from ftp://ftp.sanger.ac.uk/pub/CGP/cosmic/data_export/ which is version 61 updated and released on September 26, 2012.
72. The Catalogue of Somatic Mutations in Cancer (COSMIC) database administered by the Wellcome Trust Sanger Institute. Tissue distribution of MLL2 mutations seen on November 6, 2012 on the COSMIC website: http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=MLL2#histo which is version 61 updated and released on September 26, 2012.
73. Kokitsu-Nakata NM, Petrin AL, Heard JP, Vendramini-Pittoli S, Henkle LE, dos Santos DV, et al. Analysis of MLL2 gene in the first Brazilian family with Kabuki syndrome. Am J Med Genet A. 2012;158A(8):2003–8.
74. Coppede F. Genetics and epigenetics of Parkinson's disease. Scientific World J. 2012; 2012:489830.
75. Lynch HT, Deters CA, Lynch JF, Brand RE. Familial pancreatic carcinoma in Jews. Fam Cancer. 2004;3(3–4):233–40.
76. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, et al. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2012;2(1):41–6.
77. Tatton-Brown K, Hanks S, Ruark E, Zachariou A, Duarte Sdel V, Ramsay E, et al. Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. Oncotarget. 2011;2(12):1127–33.
78. Kurosawa K, Fukutani K, Masuno M, Kawame H, Ochiai Y. Gonadal sex cord stromal tumour in a patient with Rubinstein-Taybi syndrome. Pediatr Int. 2002;44(3):330–2.
79. Verstegen MJ, van den Munckhof P, Troost D, Bouma GJ. Multiple meningiomas in a patient with Rubinstein-Taybi syndrome. Case report. J Neurosurg. 2005;102(1):167–8.
80. Cross NC. Histone modification defects in developmental disorders and cancer. Oncotarget. 2012;3(1):3–4.
81. La P, Yang Y, Karnik SK, Silva AC, Schnepp RW, Kim SK, et al. Menin-mediated caspase 8 expression in suppressing multiple endocrine neoplasia type 1. J Biol Chem. 2007; 282(43):31332–40.

# ncRNA–Protein Interactions in Development and Disease from the Perspective of High-Throughput Studies

**Dorothy Yanling Zhao, Yue Li, Jack Greenblatt, and Zhaolei Zhang**

**Abstract** Genome-scale studies have provided strong support to the prevalent transcription of nonprotein-coding RNAs (ncRNAs) in various organisms. The immense functional potential embedded in long intergenic ncRNAs (>200 nucleotides) have especially aroused excitement in the scientific community. ncRNA–protein complexes are now known to participate in an astonishing array of processes in transcriptional and posttranscriptional gene regulations; the mechanisms reviewed here may just be a minor fraction of the vast unknowns. These ncRNA-mediated cellular activities contribute to an expansive spectrum of developmental processes such as embryogenesis and organogenesis, and the misregulation of these leads to tumor growth and diseases. The characterization of ncRNAs is progressing at a very fast pace, so much aided by innovations in genome-wide high-throughput

D.Y. Zhao, B.Sc., M.Sc., Ph.D. Candidate
Department of Molecular Genetics, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada
e-mail: dorothy0728@gmail.com

Y. Li, B.Sc., M.Sc., Ph.D. Candidate
Department of Computer Science, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada
e-mail: yueli@cs.toronto.edu

J. Greenblatt
Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, ON, Canada, M5S 3E1
e-mail: jack.greenblatt@utoronto.ca

Z. Zhang, B.Sc., Ph.D. (✉)
Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, 6th Floor, Toronto, ON, Canada, M5S 3E1
e-mail: zhaolei.zhang@utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_5,
© Springer Science+Business Media New York 2014

technologies (e.g., mass spectrometry, deep sequencing, large scale RNAi screen), and the development of computational tools for data processing. We here present an up-to-date view of the expanding field of ncRNAs in epigenetic regulation, and the development of genome-wide high-throughput methodologies that facilitates our understanding of ncRNAs in a systematic manner.

**Keywords** Epigenetics • Chromatin • Noncoding RNA • Mass spectrometry • Sequencing • RNAi

## Abbreviations

| | |
|---|---|
| CD | Chromodomain |
| CHART | Capture hybridization analysis of RNA targets |
| ChIP-seq | Chromatin immunoprecipitation followed by sequencing |
| ChIRP | Chromatin isolation by RNA Purification |
| CLIP | Cross-linking and immunoprecipitation |
| CRP | Chromatin-remodeling proteins |
| CUTS | Cryptic unstable transcripts |
| eRNAs | Enhancer derived RNAs |
| ESCs | Embryonic stem cell |
| HOTAIR | Hox antisense intergenic RNA |
| lncRNAs | Long intergenic noncoding RNAs |
| LSD/co-rest | Lysine-specific demethylase corepressor for element1 silencing transcription factor |
| mRNA | Messenger RNA |
| Nats | Natural antisense transcripts |
| ncRNAs | Noncoding RNAs |
| NoDS | ncRNA with nuclear detention sequence |
| PAR-CLIP | Photo-activable ribonucleoside-enhanced CLIP |
| PASRs | Promoter-associated short RNAs |
| PRC2 | Polycome repressive complex 2 |
| PROMPTS | Promoter upstream transcripts |
| RBP | RNA-binding proteins |
| RIP | Ribonucleoprotein immunoprecipitation for RNA |
| RIP-seq | RIP sequencing |
| | RNAcompete |
| RNAPII | RNA polymerase II |
| SELEX | Systematic evolution of ligands by exponential enrichment |
| TASRs | Terminator-associated short RNAs |
| TF | Transcription factor |
| tiRNAs | Transcription initiation-associated RNAs |
| TSSa-RNAs | Transcription start site-associated RNAs |

| | |
|---|---|
| uaRNAs | Upstream antisense RNAs |
| UNTS | Upstream noncoding transcripts |
| Xuts | Xrn1 sensitive unstable transcripts |

# Introduction

Recently, our understanding of the biological roles of noncoding RNAs (ncRNA) had a paradigm shift due to the advent of next-generation sequencing technology. Despite the tremendous efforts made, we have just started to appreciate the various classes of ncRNAs, which are implicated in almost all of the biological processes including embryonic developments and tumorigenesis. In this chapter, Sect. 2 focuses on our current knowledge of a major class of ncRNA—the long ncRNAs; Sect. 3 reviews the experimental and computational techniques employed to characterize genome-wide interactions between ncRNAs and proteins; Sect. 4 proposes our perspectives on the future of ncRNA research.

# Biological Roles of ncRNAs: Known and Unknowns

## *The Expanding Universe of Noncoding RNAs (ncRNAs)*

The human genome is pervasively transcribed, but only 1–2 % of the transcripts generate protein-coding messenger RNA (mRNA) [1–3]. The better-characterized groups of ncRNAs include tRNA, rRNA, snRNA, snoRNA, microRNA (miRNAs), and piwi-interacting piRNAs. Each group is either processed by the same set of enzymes or is involved in common metabolic or cellular pathways, thus sharing common sequence features common cellular compartments or common ribonucleoprotein modules. High-density tiling array and deep sequencing have identified ncRNAs of varying lengths [1–9]. Because of the lack of a common feature, we arbitrarily classify these ncRNAs into two groups based on size and stability (Table 1). The first group, cryptic unstable ncRNAs, are short (20–200 bp) RNAs that represent ~10 % of the transcriptome in human; they are not protected by ribonucleoproteins and tend to be stabilized upon the deletion of exonucleases [49, 50]. Short ncRNAs identified from several studies using different experimental designs and organisms may represent the same category of ncRNAs participating-related cellular activities [49]. These include the promoter-associated small RNAs (PASRs) [51–53], terminator-associated small RNAs (TASRs) [53], transcription start site associated (TSSa-RNAs), upstream antisense RNAs (uaRNAs), transcription initiation associated (tiRNAs) [54], promoter upstream transcripts (PROMTS) [55], cryptic unstable transcripts (CUTS) [56], upstream noncoding transcripts (UNTS),

**Table 1** A list of ncRNA–protein complexes that are known to be involved in epigenetic gene regulation

| ncRNA name | Protein names | Regulatory function | References |
|---|---|---|---|
| Complexes that function in *cis* | | | |
| pRNA | DNMT3b, PARP1, TIP5 | rRNA gene silencing | [10, 11] |
| Kcnq1ot1 | G9a, PRC2 | Imprinting | [12, 13] |
| Air | G9a | Imprinting | [14] |
| Xist | PRC1, PRC2, YY1 | Nucleation of X chromosome | [15–18] |
| Anril | PRC1, PRC2 | Gene silencing mediated by NATs | [19, 20] |
| Gtl2 | PRC2 | Imprinting | [21] |
| CCND1 ncRNA | TLS | Gene silencing | [22] |
| Hottip | MLL-WDR5 | Hox activate Hox cluster gene | [23] |
| evf-2 | DLX-2 | Gene activation | [24] |
| eRNA | p300, CBP | Gene activation | [25–27] |
| BDNF-AS | PRC2 | Gene silencing mediated by NATs | [28] |
| roX1/2 | MOF | X chromosome compensation | [29, 30] |
| mistral | MLL | Hox gene activation | [31] |
| DEB-T | Ash1l | Gene mis-activation in genetic disease | [32] |
| Complexes that function in *trans* | | | |
| SRA | CTCF, cohesin | Higher order looping of chromatin | [33] |
| Gas5 | Glucocorticoid receptor | Repress stress responsive gene activation by sequestering TF | [34] |
| Linc-p21 | hnRNP-K, HuR | p53 pathway gene repression | [35, 36] |
| Hotair | PRC2, LSD1/co-REST | Hox gene repression | [37–39] |
| NF-YA | PANDA | Repress stress responsive gene activation by sequestering TF | [40] |
| Malat1 | SR splicing factors, Pc2 | Splicing speckle formation/regulate splicing, recruit to actively transcribed regions | [41, 42] |
| NRON | NFAT | Transport transcription factor | [35] |
| 1/2 SBS RNAs | Staufen | Mediate nonsense-mediated decay | [43] |
| DHFR | TFIIB | Gene repression by blocking promoter region | [44] |
| TERC | Telomerase complex | Telomere extension | |
| 7sk | P-Tefb | Transcription elongation | [45, 46] |
| HSR1 | HSF1 | Heat shock gene regulation | [47] |
| Tug1 | Pc2, PRC2 | Recruit to repressed gene regions, repress cell cycle gene expression | [41, 48] |

and Xrn1 sensitive unstable transcripts (Xuts) [49, 50, 57–61]. The second group, stable and long ncRNAs (lincRNA, >200 bp ~ 100 kb) are often part of ribonucleoprotein complex [8, 48, 62–65]. For this review, we will focus our attention on the stable ncRNAs that regulate gene expression transcriptionally at the chromatin level; their biological importance is often manifested in disease phenotypes.

## *The Debate: ncRNAs as Functional Modules or Transcription Byproducts*

It might be true that the majority of ncRNAs in cells are transcriptional noise as 90 % of POL II is known to be engaged in inefficient transcription and that these short ncRNAs are degraded shortly after synthesis [66, 67]. Various studies have shown that some ncRNAs are byproducts of RNA polymerase II (RNAPII) initiation, stalled RNAPII due to pausing factors (NELF, DSIF), ripples from neighboring transcription, splicing intermediates, or random transcription from nucleosome-free regions at 5′ or 3′ ends of genes [28, 50, 56, 60, 61, 68–70]. One study showed the activity of transcription rather than the transcript product itself carry a function in cell as it opens the promoter region to make it more accessible to transcription factors and the RNAPII [71]. Some ncRNAs are rare with low expression in cell; they may interfere with transcription but it is doubtful that such observed cellular activity would elicit meaningful biological outcome in an organism [56, 72]. It is also possible that some ncRNA loci are undergoing positive selections to evolve into protein-coding genes, and we are just capturing the evolutionary intermediates.

In a manner similar to protein-coding transcripts, many ncRNAs are transcribed by RNAPII and contain splice sites, 5′ caps, and 3′ polyA modifications [6]. Bioinformatics analyses show that these ncRNAs are evolutionarily more conserved than intergenic regions and introns but are less conserved than protein-coding genes [2, 48, 73, 74]. It is hypothesized that certain secondary structures or short stretch of sequences within the ncRNAs may be under higher level of selective constraint, thus the sequence conservation over the entire length may not be necessarily very high [25, 62, 75]. To predict whether a specific ncRNA is functional, the conservation of its sequence across species, potential secondary structures, expression pattern, and motif features may all need to be considered [62, 64].

Experimentally, ncRNAs' tissue specificity can provide support to their function; studies have shown that they tend to be more tissue-specific than protein-coding genes [64]. Many have been shown to express in different tissues and brain parts, as well as in various developmental time points [15, 64, 76, 77]. More directly, large-scale RNAi screens in cell lines provide phenotypic outcome [62]. These studies have shown that (1) ncRNAs are differentially expressed in differentiation stages and have parallel expression patterns to some regulatory genes [7, 15], (2) they show distinct subnuclear or cytoplasmic localization patterns [2, 78], and (3) knocking-down some ncRNAs can lead to pluripotency changes [7, 63, 79]. It is also important to perform gain-of-function screen to investigate the ncRNAs' ectopic role but such study has not been done at large scale. Overall, it seems that computational prediction for evolutionary conservation combined with phenotypic screening using knockdown or overexpression can best support that an ncRNA carries a functional role.

## *ncRNAs' Functional Diversities in Transcriptional Regulations*

Long ncRNAs show prevalent activities in transcriptional gene regulations [7, 21, 48, 72, 80–84]. An ncRNA acts in *cis* if it regulates regions at or near the locus of its synthesis or acts in *trans* at other genomic loci. We classify their activities into four nonexclusive groups, all of which depend on ncRNAs' interaction with transcription factors (TF), chromatin-remodeling proteins (CRP), and sequence-specific pairing with the DNA (Table 1).

### *cis* Gene Upregulation Through ncRNAs Transcribed from Enhancer Regions

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) of enhancer markers (H3K4me1, H3K27ac) and p300/CBP histone acetyltransferase-bound sites revealed that these regions are transcribed into ncRNAs that subsequently recruit factors to potentiate nearby gene expression [19, 40, 69, 74, 85]. Two genome-wide studies have been performed to look at RNAs transcribed from enhancers (eRNA). One study examined the GENCODE annotated 1,000 ncRNAs in human and showed that the depletion of ncRNAs affect neighboring gene activation [74]. The second study showed that 2,000 genomic sites bound by enhancer p300/CBP overlapped with RNAPII-binding sites for active transcription, and these sites were changed upon neuronal membrane depolarization [19]. One mechanistic study showed that the ncRNA Evf2 transcribed from a distal enhancer recruited the transcription factor DLX2 to activate genes involved in neuronal differentiation [12, 74, 77]. Thus ncRNAs from intergenic or enhancer regions can alter neighboring gene expressions [25, 74].

### *cis* Gene Repression

The ncRNAs of this category function to reduce the expression of their neighboring genes rather than to promote them. One example is the cyclin D1 ncRNA (CCND1), which is transcribed from the cyclin D1 gene promoter upon ionizing radiation [37]. CCND1 binds to the transcription modulator TLS, which then blocks the activity of CREB and p300, leading to the repression of the target genes [37]. Another example is the ncRNA DHFR, which binds to the promoter of the gene dihydrofolatereductase, forming a triple helix to block the binding of TFIID for gene transcription [26]. Also playing a prevalent role in *cis* gene repression is the natural antisense transcripts (NATs) that originate from the opposite strand of many protein-coding genes [22, 24, 27]. NAT scan regulate the sense transcript through Dicer-dependent RNA cleavage or through CRP recruitment mechanisms [22, 24, 27]. For instance, the ncRNA BDNF-AS represses the sense transcript that encodes for the gene BDNF (brain-derived neurotrophic factor) through the recruitment of the PRC2 complex [67].

### *trans* Gene Regulation Through Binding to and Regulating the Activity of Transcription Factors

The ncRNAs of this category can form a complex with transcription factors (TF) to promote or repress their transcription activities. One example is the 7sk ncRNA, which binds to the P-TEFb complex to block its kinase activity for RNAPII Serine 2, therefore represses RNAPII elongation [44, 86]. Another ncRNA, PANDA (P21 associated ncRNA DNA damage activated), blocks the activity of the TF NF-Y downstream of p53 signaling, thus regulating a large subset of stress response genes [2, 82, 87]. A third ncRNA, Gas5, enhances apoptosis upon nutrient starvation through blocking the glucocorticoid receptor from binding to gene promoters with glucocorticoid response elements [88]. A fourth example is the heat-shock TF HSF1 that complexes with the ncRNA HSR1 under stress stimuli; this complex then acquires DNA-binding specificity [89]. Other studies have also shown that ncRNAs complex with TF such as YY1, MECP2 to alter their binding behaviors and activities in gene regulation [7, 34, 45–47, 74, 90].

### *trans* Gene Regulation Through Binding to Chromatin-Remodeling Proteins to Reprogram the Epigenetic Code and Alter Gene Expression Indirectly

Many ncRNAs are in a complex with histone modifiers (methyltransferases, demethylases, acetyltransferases, deacetylases) and their binders (e.g., PRC2, PRC1, MLL, Ash1l, CBP, P300, SetD8, P400, Eset/SETDB1, Suv39h1, G9a, LSD, Hdac1, SMCX/Jarid1c/Kdm5c, Jmjd6, Cbx1, Cbx3, Cbx4/Pc2, Tip60/Kat5, and Jarid1b/Kdm5b) [2, 16, 23, 40, 41, 48, 63, 72, 80, 84, 90–92]. The activity of this category of ncRNAs overlaps with the previously mentioned natural antisense transcripts (NATs) but can act in *trans*. The ability of the ncRNAs to bind multiple chromatin complexes enables them to act as flexible molecular adaptor to process combinatorial chromatin code [42]. A few examples include the ncRNA TERC that acts as a template for the telomerase complex, the ncRNA HOTAIR that recruits PRC2 and LSD/co-Rest for gene repression, and the ncRNA Xist that interacts with PRC2 and YY1 for X chromosome inactivation [13, 14, 20, 31, 38, 47, 92–95]. One may envision that ncRNA adaptors can recruit chromatin remodelers, transcription factors, and enhancers to generate distinct ribonucleoprotein clusters on the chromatins that they can also pair with specificity. These functions enable the ncRNAs to regulate the expression of groups of genes synergistically with specificity and selectivity (Table 2).

### Other Regulatory Roles of ncRNAs

ncRNAs' posttranscriptional regulatory activities in the cytoplasm are also diverse; we will briefly summarize them here [95]. (1) trafficking: ncRNA (NRON) regulates

**Table 2** ncRNA as templates/adaptors for the assembly of ribonucleoprotein modules [13, 14, 20, 31, 38, 47, 92–95]

| ncRNAs | Protein complexes | Functions |
|---|---|---|
| TERC | Telomerase complex | Telomere regulation |
| HOTAIR | PRC2, LSD/co-Rest | Gene repression at the HoxD cluster |
| Xist | PRC2, YY1 | X chromosome nucleation |
| ANRIL | PRC1, PRC2 | Tumor suppressor locus regulation |
| MALAT1, NEAT1 | SR splicing factors | Splicing regulation and splicing speckle formation |
| pRNA | DMNT3b, PARP1, and TIP5 NoRC complex | Establish and maintain silenced rDNA chromatin during cell division |
| SRA | Insulator CTCF, cohesin | Control higher order looping and imprinted gene expression |
| ncRNAs with NoDS | VHL, Hsp70, and MDM2/PML | Capture and immobilize proteins to regulate their dynamics |

nuclear trafficking of TF NFAT [10, 33]; (2) translation regulation: HuR recruits Ago2 for ncRNA linc-p21 degradation, otherwise, linc-p21 accumulation leads to enhanced JUNB and b-catenin translation [96]; (3) mRNA decay: Staufen mediates mRNA decay through base pairing mRNA 3′ UTR with Alu repeats [97]; (4) ceRNA/miRNA sponges derived from pseudogene transcripts (e.g., PTENP1) can dramatically upregulate the expression of their functional genetic counterparts (e.g., PTEN transcript) through competing for and diluting away their regulatory miRNAs [35, 98, 99]. Another example for miRNA sponge is ncRNAlinc-MD1 that sponges the miR133/135 to upregulate the expression of TFs (MAML1, MEF2C) in muscle differentiation [98].

## *Chromatin Remodeling Through ncRNAs*

ncRNAs's role in chromatin remodeling has been widely studied using imprinting, X chromosome inactivation, Hox gene cluster, and ESC differentiation or reprogramming systems; please refer to references [43, 91] for more details.

### Imprinting and Dosage Compensation

Imprinting is an essential process that initiates paternal- or maternal-chromosome-specific expression patterns in many loci. Most mammalian imprinted regions contain antisense ncRNAs whose expression represses the loci in *cis*; examples include the paternal-allelic expression of ncRNA Air and Kcnq1ot that silence Igf2r and Kcnq1genes through recruitment of histone lysine methyltransferases G9a and PRC2 [16, 40, 84, 91, 100]. Similarly, the stochastic X chromosome inactivation in epiblast involves the interplay of several ncRNAs (Xist, Jpx, Tsix, Xite,

Linx) through the recruitment of repressive chromatin complexes: DNMT3a, PRC2, PRC1, and macroH2A [41, 43, 75, 91, 101–104]. In particular, ncRNA Xist and its antisense ncRNA Tsix are transcribed on both X chromosomes; Tsix activated by ncRNA Xite silences Xist in *cis* with the recruitment of DNMT3a and blocks the interaction between Xist and PRC2; and the ncRNA Jpx activates Xist in *trans* thus antagonizes Tsix [101]. Eventually, Xist is only expressed by the chromosome that will be silenced, and the RepA region on Xist interacts with Suz12 of PRC2 through a 28 nucleotide motif, which then deposits H3K27me3 marks to nucleate the X chromosome [41, 91]. The RepF region within Xist interacts with the sequence-specific TFYY1 as its zinc finger domain binds DNA and RNA simultaneously [47]. Xist is thus an ncRNA adaptor that bridges PRC2 and YY1 to achieve synergistic repression of the DNA loci that it hybridizes to. Similar dosage compensation upregulates gene expression from the single X chromosome in Drosophila acts through the ncRNAs roX1 and roX2 that recruit the MOF histone acetyltransferase for histone H4K16 acetylation [39, 91, 105].

## Hox Gene Cluster Regulation

The Hox gene clusters in bilateria contain homeobox TF genes that are organized in tandem; these genes determine the anterior–posterior axis patterning process that is conserved from fly to man. Diverse ncRNAs regulate the clusters with repressive and activating activities in *cis* and *trans* [91, 100]. In flies, ncRNAs transcribed from the Bxd and Ubx Hox regions can induce and repress Ubx gene expressions [17, 106]. In mammals, HOTAIR transcribed from HoxC cluster *trans* silences the HoxD cluster through recruiting PRC2 and LSD/co-Rest [42, 100, 107]. HOTAIR binds PRC2 (Suz12, Ezh2) through its 5′ end, and LSD1 through its 3′ end [42, 45]. Another Hox cluster ncRNA HOTTIP is transcribed 5′ to the HoxA cluster; it then recruits MLL/WD5 to drive H3K4 trimethylation and upregulates genes in *cis* [92]. A third ncRNA Mistral activates Hoxa6/7 and cell differentiation via MLL1 recruitment [42, 92, 100, 107].

## The Roles of ncRNAs in Stem Cell Pluripotency, Differentiation, and Reprogramming

### ncRNA Expression Profiling

In vitro stem cell pluripotency, cell lineage specification, and induced pluripotency, stem cell reprogramming (iPSC) assays have been used to investigate ncRNAs' functional roles. ncRNA's expression was examined in ESCs through profiling the expression of 954 ncRNAs to show that 174 of them were differentially expressed through a 16-day differentiation time course [1]. In this study, some ncRNAs were found to share expression profile with developmental genes, and two ncRNAs (Evxas and Hoxb5/6as) were found to associate with MLL1 to generate active

histone mark H3K4me3 [1]. One ingenious study examined intergenic ncRNAs using ChIP-seq of H3K4me3 (which marks the gene promoter), H3K36me3 (which marks the gene body), and RNAPII (which marks active transcribed regions) to profile active transcribed intergenic regions in ESCs [6]. These signatures reveal that ~1,600 and ~2,500 ncRNA regions in mouse and human are active, respectively; and 24 % of them are physically associated with PRC2 [48]. However, a limitation of this approach is that it only considers regions that do not intersect with protein-coding genes, thus excluding ncRNAs that are intronic or antisense with respect to genes.

### ChIP-seq-Binding Pattern and ncRNA Knockdown Studies

Approaches using ChIP-seq against the protein-ncRNA complexes and RNAi against ncRNAs have directly shown that ncRNAs are regulated by major stem cell TFs and play active roles in cell differentiation. One study showed that ncRNAs were bound by p53, NF-kappaB, Sox2, Oct4, and Nanog in ESCs, and the knockdown of these ncRNAs altered cell pluripotency, suggesting that the ncRNA and the pluripotency TFs are in a regulatory feedback loop [6, 29]. A second study showed that hundreds of ncRNAs such as linc-ROR, linc-SFMBT2, and linc-VLDLR had upregulated expression in iPSC reprogramming and were directly bound by the TF Oct4, SOX2, or Nanog at their promoters [30]. The knockdown of linc-ROR reduced reprogramming efficiency and led to expression of genes in p53 signaling and apoptosis [30]. It is still not clear how linc-ROR affects the p53 pathway, i.e., whether through transcriptional or posttranscriptional means. A third study using RNAi screen against 237 ncRNAs in ESCs showed that these ncRNAs mediated gene expression in *trans* through associating with CRPs, and these nRNAs maintained ESC pluripotency [7]. The knockdown of these ncRNAs induces specific differentiation pathways for endoderm, ectoderm, and mesoderm formations [30, 40].

### ncRNA–Chromatin Protein Complexes in Stem Cells

Studies have shown that ncRNAs regulate pluripotency through binding to CRPs, TFs, and enhancers [2, 7, 63]. Some ncRNAs bridge multiple CRP complexes (e.g., linking PRC2 and Eset and Jarid1c, PRC1, and Jarid1b) and the knockdown of these components lead to similar gene expression changes [7]. As these CRP complexes may act redundantly in gene regulation, it remains unclear if single ncRNA is necessary and sufficient to bridge them specifically, and if some ncRNAs function redundantly. One way to elucidate this would be through performing co-immunoprecipitation of the protein complexes with and without RNase. As well, in cells that lack a particular ncRNA expression, one does not expect the protein clusters to form if they require the particular ncRNA bridge.

Studies often reveal the same set of protein complexes (e.g., PRC2, G9a. MLL, Ash1l, DNMT3a, CTCF) to remodel chromatin loci via various ncRNAs; PRC2

RIP-seq experiments indeed show that PRC2 is associated with a large number of ncRNAs [48, 80]. One study suggested that transcription start site-associated ncRNAs are GC rich and forms stem loop; such loops are preferentially recognized by PRC2 to nucleate the loci in *cis* [72]. It remains unclear whether gene silencing specified by the ncRNA–dsDNA formation requires triple helix formation in general, and what are the structure or sequence features in ncRNA that are necessary for this activity.

## *ncRNA–Protein Interactions Are Regulated by Posttranslational Modifications*

The RNA-binding activity of the RNA-binding protein (RBP) is regulated by the protein's posttranslational modifications. For example, chromodomain (CD) is now shown to have RNA/DNA-binding capacity in addition to histone lysine methylation-binding ability (for H3K9me, H3K27me) [90]. The Drosophila CD in MOF and MSL3 that are involved in dosage compensation can bind RNA, and the mammalian Cbx7 CD also possesses RNA-binding capacity for ncRNA Anril [39, 83]. One NMR study of fission yeast (*Schizosaccharomyces pombe*) Chp1 CD shows that it has RNA-binding capacity that is enhanced when it is simultaneously bound to the histone H3K9me2 mark [108]. A study of the Pc2/Cbx4's binding to ncRNAs shows that it is regulated by methylation of Pc2/Cbx4 at Lysine191 by the methyltransferases SUV39H1 and the demethylase KDM4C [90]. Methylated Pc2/Cbx4 is associated with repressive Pc group bodies, whereas unmethylated Pc2/Cbx4 localizes to the active interchromatin granules. ncRNAs effect the recruitment through Pc2 CD—methylated Pc2 is recruited by ncRNA TUG1, whereas unmethylated Pc2 is recruited by the ncRNAs NEAT2/MALAT1 to these compartments [90]. Binding of these ncRNAs also alters Pc2's preference for repressive (H3K9me3, H4R3me2s, H3K27me2) or active (H2AK5ac, H2AK13ac) histone marks [90]. Regarding the role of phosphorylation in regulating ncRNA–protein interactions, one example is Ezh2, the methyltransferase component of PRC2 [45]. Ezh2 is phosphorylated by cyclin-dependent kinase 1 at Threonine 345/487 in a cell cycle-dependent manner, and a phospho-mimic at 345 enhances its binding of ncRNA HOTAIR.

## *The Roles of ncRNAs in Tumorigenesis and Diseases*

ncRNAs are now viewed as programmers for various developmental processes that include maternal–zygotic transition, epithelial–mesenchymal transition, organ morphogenesis, and tissue differentiation; thus their role in tumor genesis is evident [6, 7, 62, 63, 74, 91, 109–112]. ncRNAs' roles in cancer are diverse, ranging from mis-recruitment of chromatin-remodeling proteins (CRPs) and TFs, to mis-repression of sense transcript, to mis-splicing-regulations, to misregulation at the translational

level. One epigenetic example comes from a study showing that breast carcinoma contain differential ncRNA expression from the HoxC cluster, including HOTAIR [40]. Forced expression of HOTAIR recruits PRC2 for enhanced H3K27me3, changing the expression of genes that are inhibitory to metastasis [40], while inhibition of HOTAIR leads to reduced metastasis. A comprehensive ncRNA analysis has been performed in prostate cancer, uncovering 121 ncRNAs (PCATs) that have differential expressions across cancer stages. ncRNA PCAT-1 has been found to be a prostate-specific cell proliferation regulator that it is normally repressed by PRC2 [110]. The mechanisms of these PCATs still await elucidation.

Many tumor suppressor genes have antisense ncRNAs [113]. One example is the INK4b/ARF/INK4a locus that encodes 3 tumor suppressors, and the antisense ncRNA ANRIL that represses the locus in *cis* through interacting with Cbx7 of PRC1 and Suz12 of PRC2 [83, 93, 113]. Upregulation of ANRIL and Cbx7 lead to cancer and cardiac disease [83, 113, 114]. There are a few examples of PRC1 binding to ncRNAs, suggesting that unlike PRC2 that regulates a large number of loci through ncRNAs, the PRC1–ncRNA complex is probably more specific to a few loci.

Three ncRNAs have been shown to participate in p53 pathway while many have been found to have altered expression upon p53 induction [6]. One is the aforementioned lincRNA-21 that is transcribed with the nearby CDKN1A gene upon DNA damage [115]. LincRNA-21 binds to hnRNP-K through its 5′ end to upregulate a large subset of p53 target genes in apoptosis [115]. Another study identified more than 200 long ncRNAs in regions next to cell cycle-controlling genes (cyclins, cdks, cdk inhibitors, etc.) and showed that these mRNAs and the adjacent ncRNAs had similar cell cycle-dependent expression profiles [82]. The ncRNA PANDA (P21-associated ncRNA DNA damage activated) blocks p53-dependent apoptosis through blocking the TF NF-Y from binding to and activating apoptotic genes downstream of p53 [82]. A third ncRNA, TUG1, is bound by PRC2 and is induced by p53; its depletion leads to upregulation of genes involved in cell cycle regulation [48].

Recently, one ncRNA was characterized to be an epigenetic activator in rare genetic disease facioscapulohumeral muscular dystrophy (FSHMD or FSHD) that is caused by copy number variation-dependent de-repression of genes in the 4q35 locus [116]. In somatic cells, this locus is normally repressed by PRC. But in the diseases state, it is transcribed into an ncRNA (DBE-T), which then directly interacts with Ash1l, recruiting it to the locus to dimethylate H3K36 for active gene transcription [116].

## Methods for Characterizing RBP–ncRNA Interactions: Ribonucleoprotein IP Coupled with Sequencing (RIP-seq)

High-throughput sequencing technologies have revolutionized genome and transcriptome studies; datasets generated from different platforms can often be integrated for the purpose of cross-validation and comparison [117]. For example, mass spectrometry and deep sequencing performed on multiple protein baits within

**Table 3** Experimental techniques that are amenable for genome-wide ncRNA study

| Experimental type | Experimental technique | Description |
|---|---|---|
| ncRNA–protein purification | Ribonucleoprotein IP (RIP) | Antibody-based purification of the protein |
| | HITS-CLIP | RIP coupled with RNA–protein crosslinking (UV) and RNAase treatment to reduce background |
| | PAR-CLIP | CLIP with the addition of 4-SU for enhanced signal detection |
| | MS2-tagged RNA | Purify the RNA and bound protein, DNA tag facilitates RNA visualization |
| ncRNA-DNA purification | ChIRP | Tiling oligos complementary to ncRNAs to purify the bound DNA |
| | CHART | Oligos complementary to ncRNAs to purify the bound DNA and protein |
| Strand specific sequencing | Chemical based | Bisulfite treatment for C to U conversion detected in sequencing |
| | | d-UTP incorporation during cDNA synthesis followed by destruction |
| | Template switch | Generates first-strand cDNA with triple G overhang for second adaptor that contains triple C to recognize |
| | Direct sequencing strategy | |
| ncRNA–protein interaction in vitro | SELEX | Protein domain incubated with synthetic RNA library to detect favored sequences |
| | RNAcompete | Protein domain incubated with synthetic RNA library to detect favored sequences |

the same complex allows protein–ncRNA interactions to be identified, cross-validated with high confidence [117]. RIP-seq data can be compared with ChIP-seq of the same protein bait to examine if the protein plays a *cis* or *trans* role with respect to the ncRNAs and the genes it regulates. In this section we will review these high-throughput approaches; these techniques are also summarized in Table 3.

## *Large-Scale Protein-Tag or Antibody-Based Systems for Protein–Protein and Protein–RNA Interactions*

There is a nontarget-specific approach to enrich ncRNA–protein fraction from cellular lysates through glycerol gradient to facilitate ncRNA–protein interaction studies by sequencing or mass spectrometry [118]. More target-specific ncRNA–protein interaction studies rely on precipitating the complex for RNA extraction and sequencing [7, 80]. Variations to the method include high-throughput sequencing with UV-cross-linking and immunoprecipitation (HITS-CLIP) that can be used with the addition of 4-thiouridine (4SU) prior to cross-linking for better signal detection,

which is termed as Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) [119]. With respect to ncRNAs' role in chromatin regulation, Guttman et al. had performed native RIP-seq to identify 11 chromatin proteins to interact with ncRNAs [7].

Large-scale RIP-seq screen is ideal for the identification and validation of ncRNAs that function as linker–adaptors for multiple chromatin complexes. Protein tagging and antibody-based approaches can both be used. The advantages of the protein-tagging system are that the levels of tagged protein can be controlled and standardized across the experiments, and the tag allows easy manipulation for protein–protein, protein–ncRNA interaction studies, as well as localization study with staining [36]. The disadvantage is that nonphysiological interactions may be observed due to bait protein overexpression. The advantages of antibody-based studies are that they allow for endogenous protein–protein, protein–RNA interactions to be captured, and for tissue-specific or developmental-restricted interactions to be readily examined [32]. If the antibody recognizes the same protein in multiple species, the interactions could then be compared across species. Phage display for synthetic antibody production has been a popular high-throughput approach for generating antibodies with high in vitro binding affinity and specificity; but the in vivo specificities often requires laborious testing [32]. A disadvantage of the antibody approach is that the low abundance interactions may not be readily detected. A recent study shows that synthetic antibodies can be generated to identify RBP-bound RNAs or to block RBP–RNA interactions [32].

## *RNA Deep Sequencing with Multiplexing Barcodes*

Two early studies using high density tiling arrays provided the first evidence that ncRNAs are prevalently transcribed in the genome [120, 121]. The advantages of deep sequencing are many folds and it is gradually replacing the hybridization-based microarray platform. Unlike the array-based techniques, which require prior knowledge of the sequence, deep sequencing can capture transcript isoforms and splicing variations de novo [4]. Additionally, deep sequencing allows for a higher dynamic range of detection, for a more quantitative measurement, and for single base resolution [4]. The reads from deep sequencing range in length from 30 to 400 bp, depending on the platform chosen (Illumina/Solexa, Roche/454, Life/APG SOLiD, Helicos Biosciences). Illumina platform is more commonly used, generating 100–200 million pair-end reads per flow cell lane. The cDNA library preparation involves heat or enzyme fragmentation of RNAs, ligation of adaptors that contain universal priming sites onto the RNAs, reverse transcription into cDNA, and PCR amplification. The samples are then loaded onto a flow cell covered with forward and reverse primers, and amplification steps take place to generate clones of DNA for detection. The sequences are detected using the 4 color cyclic reversible

termination strategy—each cycle starts with the incorporation of 4 nucleotides each labeled with a different dye, followed by washing and imaging, and cleavage that removes the dyes and regenerates the free 3′end for the next cycle [5]. At present, the sequencing cost is becoming increasingly lower; furthermore multiple samples can be barcoded (multiplexed) during adaptor ligation step and sequenced in one single lane. The sequencing reads can then be separated by these barcodes into individual bins to reflect reads from each sample during data processing.

## *Strand-Specific RIP-seq*

DNA strand information from RIP-seq can be retained to determine the origin of the transcript that is being sequenced: i.e., sense or antisense. There exist several strand-specific cDNA preparation strategies; some rely on chemical modifications or orientation-specific priming adaptors in cDNA preparation, others use innovative direct sequencing technologies. The application of strand-specific sequencing to RIP is limited due to small amount of starting material; the complexity of the sample also tends to be reduced through extensive preparation steps. One method is based on changing all C residues to U through bisulfite treatment of RNA [122]. A second method incorporates deoxy-UTP during second strand cDNA synthesis and subsequent destruction of the U-containing strand [123]. A third method, template-switching strand-specific cDNA synthesis, involves first-strand cDNA synthesis from random hexamers that creates a triple C overhang at the 3′ end, which then base pairs with triple G containing adaptor for second-strand synthesis [80, 124]. The lack of uniformity of transcript coverage and the introduction of transcript length bias and spurious reads are common problems for this method [125]. This method was used in a genome-wide study of PRC2–ncRNA interactome, identifying >9,000 PRC2-interacting RNAs [80]. The PRC2 interactome is enriched with transcripts that correspond to over 40 % of the oncogenes or tumor suppressors, including c-Myc, Brac1, Klf4, and Dnmt1. It also encompassed many previously confirmed ncRNAs from imprinting regions [80].

Developments of direct sequencing technologies allow the first-strand cDNA to be determined directly, thus retaining its strand information [126–129]. The first direct RNA sequencing was developed using Helicos platform that relies on hybridization of 3′ polyA RNA to channels of polyT-coated surface; RNA species lacking polyA tail can be modified in vitro prior to sequencing [130]. Only femto-mole or less amount of sample is needed for this approach [125]. A second strategy is to use flow cell reverse transcription sequencing (FRT-seq) [129]. The reverse transcription step is performed on the flow cell using the adapter-ligated mRNA as template, thus preserving the strand information. PCR amplification is skipped to reduce bias and allows more accurate quantification [129].

## CLIP-seq and PAR-CLIP

Comparing to native RIP-seq, CLIP-seq (cross-linking and immunoprecipitation coupled with high-throughput sequencing) and PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) are more advantageous for elucidating RNA components of known protein–RNA complexes as stringent experimental condition is used for enhanced specificity. UV and formaldehyde cross-linking have both been used successfully with CLIP, with the addition of RNase (T1, A, I, or a combination of these) to reduce background. RNase digestion often requires optimization and might introduce bias [131]. In HITS-CLIP (high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation), the RBP-binding site is often skipped by reverse transcriptase in cDNA preparation, resulting in signatures such as T substitution and nucleotide deletion that can be used to elucidate the exact protein–RNA-binding sites in the analysis [131, 132]. HITS-CLIP can detect at a resolution of 30–60 nucleotides bracketing the RBP-binding sites; it has been used for RNA binding analysis for Nova, SRSF1, Fox2, PTB, and TDP43-splicing factors [132].

PAR-CLIP incorporates 4-thiouridine (4-SU) or 6-thioguanosine (6-SG) in culture media prior to UV 365 nm irradiation and can determine the exact residue involved in binding at high resolution [119]. The protein–RNA interaction site is determined by T to C transition. Methods like CLIP and PAR-CLIP can provide high resolution identification of miRNA-binding sites on the transcripts as well [131]. A quantitative analysis using CLIP and PAR-CLIP has shown that both methods work well for low (HuR) and high (Ago2) complexity sequence binders; CLIP giving better correlation between RNA enrichment and affinity and PAR-CLIP enabling better recovery of miRNA seed-complementary sites [131]. The major disadvantage of PAR-CLIP is that it cannot be performed in tissues in general, though one PAR-CLIP study has been done using *Caenorhabditis elegans* fed with 4-SU [133].

Recently, two studies used PAR-CLIP and quantitative MS for oligodT precipitation of total mRNA to examine the general mRNA–protein interactome [134, 135]. These approaches are nontarget-specific but reveal the power of integrative high-throughput studies. The “interactome capture” unveils ~800 proteins significantly enriched in RNA binding and pinned down that mRNA 3′ UTR regions important for the interactions [134, 135].

## RIP-seq Computational Analysis

Following sequencing, the reads can either be aligned to a reference genome or transcriptome or be assembled de novo into longer transcripts. Paired-end sequencing strategy is often used to increase read coverage, allowing more confident alignment or assembly. The experimental challenges with respect to RIP-seq are associated with library quality (such that it should be relatively free of adaptor

contaminants and is not be overamplified in PCR to produce highly redundant artifacts) and sequence quality. The computational challenges include RIP-seq read alignment and assembly of ncRNA from the alignments [136].

### Aligning RIP-seq Reads

Deep sequencing reads include those that span exon–intron junctions that cannot be directly mapped to a reference genome. TopHat is a gapped aligner that is specially designed to align RNA-seq reads [137]. TopHat first maps all the reads to the reference genome using a fast alignment tool called Bowtie [138]; it then attempts to align unmapped reads to splice junctions using a seed-and-extend strategy. To report a splice junction, TopHat adopts a heuristic filter to discard hits with low coverage across the junction relative to the highest coverage on either side of the junction. The latest TopHat version (release 1.4) is able to take advantage of the mate-pair information in paired-end sequencing, enabling it to search for splice junctions within the genomic interval between the read pair without splitting the reads. Some reads, known as multihits, may be mapped to multiple genomic loci due to gene duplication or repetitive elements. A common practice is to discard these multihits; an alternative approach is to distribute the multihits to the aligned loci with probabilities proportional to the number of uniquely mapped reads that correspond to adjacent regions [139].

### ncRNA Assembly

#### Annotation-Based Strategy

The first strategy is a "rule-based" approach that relies on existing reference transcriptome from databases such as Ensembl or UCSC. The abundance of each transcript is computed as reads per kilobase of exon per million mapped reads (RPKM). A transcript is found to be enriched in the sample above control if its RPKM and the fold-change ratio are above certain thresholds [80, 139]. In practice, however, these thresholds are sometimes chosen arbitrarily and lack statistical rigor. Another disadvantage of the approach is that it cannot identify novel transcripts. With replicated samples and controls, a more rigorous statistical test can be conducted to detect enrichment from background using either the read counts or the RPKM as the metrics for relative transcripts abundance. For a parametric statistical test, Gaussian distribution might provide a good approximation to RPKM but it is inadequate to model the read counts due to their discrete and skewed nature. Poisson and negative binomial (NB) distributions are both popular for modeling read count distribution. Alternatively, nonparametric methods such as permutation test may be used on datasets with sufficient number of replicates [140].

Reference-based assembly methods such as Scripture [141] and Cufflinks [142] operate on read alignments (rather than reads) produced by a spliced aligner such as

TopHat. Scripture first transforms the genome sequence into a graph with nodes as single bases and edges representing all possible connections of bases implicated in the (spliced) alignments. It then identifies statistically significant transcript paths across the graph based on their read coverage. Cufflinks first divides read alignments into nonoverlapping components, each considered as a directed graph. Within each component, two overlapping fragments are defined to be compatible and belong to the same path in the graph if they imply the same introns or splicing event. Following this definition, Cufflinks finds the minimum path cover(s) for all the read alignments by implementing a constructive proof of the Dilworth's theorem. The theorem indicates that finding such path cover(s) is equivalent to finding the largest set of mutually incompatible fragments in a bipartite graph. For a more detailed description of transcript assembly methods, please refer to previous reviews [136, 143]. These methods do not need to model the two types of background noise in RIP-seq data, which represent the nonspecific RNA interactions with a protein of interest (technical background) and the nonspecific RNA input from the control dataset (biological background).

#### De Novo Transcript-Assembly Strategy

Genome-independent transcript reconstruction algorithms often model the assembly problem with a de Bruijn graph with nodes representing subsequences (k-mers) and edges representing sequence overlap between the k-mers. The problem of transcript assembly is then reduced to finding the minimum number of Eulerian paths that go through every edge only once [144]. The disadvantages of this approach are that it is very sensitive to sequencing error and that the choice of the k-mer length can lead to very different assembly outcomes.

#### Peak Calling Strategy

The third strategy is based on ChIP-seq peak calling algorithm as the peaks refer to read enriched regions relative to background noise. These algorithms can distinguish bona fide transcription factor binding sites (TFBS) or histone modification patterns from background [145]. Among many programs, MACS [146] and QuEST [147] represent the parametric and nonparametric framework, respectively. MACS applies a dynamic Poisson distribution to model local background bias within the sample and control; the $p$-values of the candidate peaks are computed using Poisson parameters. To correct for multiple testing, MACS computes an empirical false discover rate (FDR) using a "sample swap" trick, i.e., treating the data from control as sample. At each $p$-value, MACS uses the same parameters to find sample peaks over control and control peaks over sample. QuEST applies a Gaussian kernel density estimation function to model the read enrichments across the genome as a continuous spectrum [147]. Similar to MACS, QuEST estimates an FDR for each candidate peak using the control library. QuEST randomly separates the control data into two subsets:

one set is used as pseudo-positive and the other as background. At the same *p*-value cut-off, an FDR is defined as the number of peaks detected in the random comparison divided by the number of peaks in the actual comparison.

Although the underlying statistical frameworks of peak callers may be generalized to RIP-seq, most of them are specific for ChIP-seq. Unlike RIP-seq reads (25–100 bp), the ChIP-seq DNA fragments (200–400 bp) are longer; so only ends are sequenced with one end coming from the plus strand and the other from the minus strand. Many algorithms exploit such strand-dependent bimodality [148] by searching for the symmetry and shifting the twin peaks toward the middle before modeling the peak distribution [146, 147, 149]. Most peak callers extend the aligned reads uni or bidirectionally (before/after the shift) by the number of bases proportional to the length of default or estimated DNA fragment. The read extension is intended to improve the power of peak detection. Notably, both assumptions are invalid for RIP-seq analysis due to the single-stranded nature of the RNA and the existence of splicing variations. Most peak callers were developed to predict TF-binding sites (TFBS), which are usually <10 bp long, much shorter than the length of transcripts that RIP-seq aims to identify. Some recently developed algorithms [149–152] allow for broader regions associated with histone marks to be predicted [149, 153]. However, the performances of these algorithms tend to heavily depend on the nature of the ChIP-seq dataset [148, 154]. The aforementioned strategies are not ideal for RIP-seq analysis; the authors of this chapter have recently developed a software tool (RIPSeeker) that is specialized for RIP-seq analysis (http://www.bioconductor.org/packages/2.12/bioc/html/RIPSeeker.html).

## *Other Genome-Wide Studies to Define ncRNAs' Functions*

### Large-Scale Cell-Based RNAi Screen and RNA Localization Studies

Several large-scale RNAi screens have been performed to characterize ncRNAs' regulatory roles in pluripotency and differentiation [6, 7, 78]. One study used Nanog-promoter-driven luciferase reporter to show that 26 ncRNAs promote pluripotency and used marker genes for lineage specification to show that various ncRNAs are associated with particular differentiation paths [7]. This information was combined with antibody-based RIP-seq to reveal ncRNAs that are in complex with CRPs to regulate gene expression circuit [7]. Another study combined RNAi screen with localization study (c-KLAN), in which in vitro transcribed dsRNAs were used to make an esiRNA library against 594 ncRNAs to examine their knock-down effect in ESC identity [78]. The same set of RNA was converted into probes for in situ hybridization (FISH) to detect ncRNA localization [78]. Localization studies could crudely classify ncRNAs into transcriptional/chromatin related or posttranscriptional gene regulation pathways. RNA/DNA FISH used with protein immunostaining have been done in both cells and tissues to provide more refined pattern thus is a more informative strategy [155].

### MS2-Tagged RNA

This method uses an ectopically expressed construct that contains the ncRNA linked to a MS2 hairpin loop, a 19-nt viral sequence [156]. Co-currently, the cells also ectopically express the MS2-binding protein (RBP MS2) fused to an affinity tag [156]. The MS2-ncRNA affinity purification has several versatile applications: (1) proteins associated with the ncRNAs can be identified by mass spectrometry; (2) associated RNAs (e.g., microRNAs) can be identified through deep sequencing that complex with the ncRNA bait; (3) DNA cross-linked to the tagged ncRNA can be identified in a manner similar to ChIP-seq [157]. Once the putative protein and DNA interactors are identified, the ncRNA–MS2 system can be used to test for truncated ncRNAs to identify the functional modules within the sequence that are essential for the interactions. GFP binding to the M2 loop on the tagged ncRNA enables straightforward in vivo visualization of the complex; a variant of the method that uses cell permeable fluorophore rather than endogenously expressed GFP is an exciting recent advancement in the field [158].

### ChIRP and CHART

One proposed regulatory role of ncRNAs is that they can function as RNA decoys or guides that can form ncRNA–RNA duplex, or ncRNA–dsDNA triplex, to enhance or repress gene expression [159, 160]. An example of ncRNA–dsDNA formation is ncRNA DHFR, which binds to the promoter to block its own transcription; an example of ncRNA–RNA formation is the ncRNA pRNA, which binds to the rDNA loci for gene repression [18, 26]. Two groups have developed two similar strategies to examine genome-wide ncRNA–dsDNA interaction, as well as the bound chromatin proteins [161, 162]. ChIRP (Chromatin Isolation by RNA Purification) relies on glutaraldehyde cross-linking of ncRNAs to DNA and subsequent precipitation with tiling oligos targeting the full length of the ncRNAs [162]. ChIRP-seq has been applied to examine 3 lncRNAs (roX2, HOTAIR, and TERC), which revealed that these interactions are sequence specific and are present in multiple foci genome wide. Sequence motif search showed that HOTAIR preferentially occupies a GA-rich DNA motif to nucleate broad domains, which are then associated with the PRC and the repressive H3K27me3 marks. Importantly, HOTAIR interacts with DNA in the absence of Ezh2, indicating that ncRNA–DNA binding is at the top of the epigenetic signaling cascade [162].

CHART (Capture Hybridization Analysis of RNA Targets) is a similar method that uses tagged oligo complementary to the ncRNA target for the pull-down of associated proteins and DNAs [161]. The method shows that the ncRNA roX2 is part of the MSL complex and localizes to active gene regions for dosage compensation [161]. The major caveat of both methods is that the oligos used to retrieve ncRNAs may directly interact with the DNA targets, thus generating a high background and false positive rate [161]. Both studies also require rigorous negative controls, which would be samples in which the ncRNA target has been efficiently depleted through in vivo or in vitro means.

### ncRNA Stability

As RNA is incorporated into ribonucleoprotein complexes co-transcriptionally, ncRNAs that complex with proteins are therefore more stable than those that are just transcriptional byproducts [163]. Large-scale ncRNA stability measurement is thus an informative determinant of ncRNAs' functionality. One study employed actinomycin D for transcription blockage in cells and probed RNA stability with custom arrays to determine the half-lives of ~800 ncRNAs and 12,000 mRNAs [11]. The measured half-lives of ncRNAs and mRNAs vary over a wide range with the median/mean of ncRNAs (3.5 h/4.8 h) to be less than that of mRNAs (5.1 h/7.7 h) [11]. Intergenic and *cis*-antisense ncRNAs that have gone through splicing are more stable than those that are nuclear. A second study used 4-thiouridine labeling of RNA and measured transcript decay at different time points in cells using microarrays [164]. Such study could be adapted to measure *cis*-regulatory elements that underlie ncRNA stability, employing deep RNA-seq for high sequence resolution.

### ncRNA Structure Predictions

High-throughput genome-wide RNA structure probing methods have utilized nucleases to cleave various structures within RNAs, followed by RNA sequencing of the fragments and computational analysis [165, 166]. One study used the combination of V1 RNase for ssRNA and S1 RNase for dsRNA cleavage, and another used nuclease P1 for ssRNA cleavage [165, 166]. A third method (hydroxyl radical foot printing) used known secondary structure to model tertiary structure to generate nicks in RNA backbone using hydroxyl radical, which report the solvent accessibility of the RNA [167]. Reverse transcription at the nick generates a cDNA profile; the lower the cDNA signature for a sequence, the more buried it is within the 3D structure [167]. Regardless of the accuracy of these predictions, the in vitro studies do not take into account the protein–ncRNA complex formation that affects ncRNA structure and accessibility in vivo. All these approaches require RNA extraction and renaturation in vitro [168]. Combining the structure probing strategies with the aforementioned purification techniques, one may better predict the structure of the RBP–ncRNA interaction complex, e.g. which region of the ncRNA is buried in protein, which is exposed in solution and form secondary structure, and which may associate with dsDNA.

### ncRNA–Protein Interaction Motif Predictions

It has been proposed that the interactions between RBP and dsRNA are not sequence specific, i.e. not dependent on specific RNA sequence motifs, but the interactions between RBP and dsRNA are often sequence specific. Such interactions have been studied through both in vivo and in vitro means [169]. The well-characterized RBP domains include RNA recognition motif, K homology, SAM, DEAD box helicase,

dsRNA binding domain, Piwi, Paz, Pumilio, TRAP, S1, and Zinc finger; there are also reports on noncanonical RNA-binding domains (Chromo, SAO, WD40 motif, LSD, RGG box, etc.) [135, 170]. Many of these domains exist in tandem or in combination, forming diverse modules that enhance RNA-binding affinity and specificity in various regulatory pathways [135, 170]. In RNA interactome capture studies, proteins containing unstructured repeated regions that are enriched of Glycine, Arginine, Lysine, and Tyrosine are overrepresented (e.g., RGG Box) and may include additional novel RNA-binding domains [135]. It is not clear how the canonical RNA-binding domains would bind ncRNAs, or if many novel RNA-binding domains are yet to be identified for ncRNA binding.

In vivo, HITS-CLIP or PAR-CLIP allow identification of protein binding sites on ncRNAs, and the motif enriched in binding can be searched computationally using motif search algorithms such as MEME [171]. The structure of the bound region on ncRNAs could also be computationally screened for secondary structure predictions. In vitro, methods such as SELEX (Systematic Evolution of Ligands by EXponential enrichment) and RNACompete allow RNA motif or structure to be predicted through incubating a library of synthetic RNAs with purified protein domains [172–174]. SELEX begins with a pool of synthetic random RNA library; the sequences purified by the bound proteins are then amplified by PCR and sequenced for consensus motif [172–174]. RNACompete workflow relies on the generation of an RNA pool through in vitro transcription that comprises 213,130 unique RNAs with all possible 10-bp sequences and all possible 7- or 8-bp loops. A pull-down of RNAs bound to tagged RBP of interest, and microarrays are used to examine enrichment of RNAs in the bound fraction [172].

## Future Perspectives

While numerous ncRNAs are identified at rapid pace, we are overwhelmed by how little we know in detail about their function. The building of a genome-wide ncRNA–protein interaction network will be the first step towards examining ncRNA function systematically. ncRNA expression profile overlapped with ChIP-seq data should also be performed at system scale to reveal in detail, how ncRNAs are regulated, and what acts as their stimuli or elicits a feedback loop at their transcription level. We also know little about how the RBP and ncRNA interact and the exact domains and sequence motifs involved, as well as the overall 3D structures. Does the binding of ncRNA alter the protein structure or does ncRNA act as a linker–adaptor, or whether the ncRNA enhances or represses the protein's catalytic activity? Modifications such as phosphorylation or methylations associated with ncRNA–RBP are also of great interest. The information embedded within the sequence or structure of the ncRNAs needs to be clearly elucidated as it is important to know how minor perturbations in ncRNAs can be translated into disease phenotypes. Such study could be facilitated by Genome-wide Association Studies (GWAS) that uncover correlations between mutations in the ncRNA loci to disease phenotypes.

With better understanding of the versatility in ncRNA–protein interactions in development and disease, we can design ncRNA as guides, inhibitors, decoys, or linker–adaptors to reroute the epigenetic program. Such genetic reprogramming through ncRNA may be of great therapeutic value for cancer or disease treatment in the future.

**Acknowledgment** This work was supported by Ontario Research Fund—Global Leader (Round 2) to JG and ZZ. YL is co-funded by NSERC Canada Graduate Scholarship and Ontario Graduate Scholarship.

## References

1. Dinger ME et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 2008;18(9):1433–45.
2. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136(4):629–41.
3. Birney E et al. Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799–816.
4. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
5. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010;11(1):31–46.
6. Guttman M et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–7.
7. Guttman M et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295–300.
8. Okazaki Y et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002;420(6915):563–73.
9. Carninci P et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559–63.
10. Willingham AT et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science. 2005;309(5740):1570–3.
11. Clark MB et al. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012;22(5):885–98.
12. Feng J et al. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev. 2006;20(11):1470–84.
13. Zappulla DC, Cech TR. RNA as a flexible scaffold for proteins: yeast telomerase and beyond. Cold Spring Harb Symp Quant Biol. 2006;71:217–24.
14. Audas TE, Jacob MD, Lee S. Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. Mol Cell. 2012;45(2):147–57.
15. Mercer TR et al. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci USA. 2008;105(2):716–21.
16. Nagano T et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 2008;322(5908):1717–20.
17. Petruk S et al. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell. 2006;127(6):1209–21.
18. Schmitz KM et al. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 2010;24(20):2264–9.
19. Kim TK et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465(7295):182–7.

20. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell. 2009;35(4):467–78.
21. Plath K et al. Role of histone H3 lysine 27 methylation in X inactivation. Science. 2003;300(5616):131–5.
22. Katayama S et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309(5740):1564–6.
23. Bertani S et al. The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. Mol Cell. 2011;43(6):1040–6.
24. Watanabe T et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. 2008;453(7194):539–43.
25. Kutter C et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet. 2012;8(7):e1002841.
26. Martianov I et al. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007;445(7128):666–70.
27. Magistri M et al. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. Trends Genet. 2012;28(8):389–96.
28. Xu Z et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009;457(7232):1033–7.
29. Sheik Mohamed J et al. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA. 2010;16(2): 324–37.
30. Loewer S et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010;42(12):1113–7.
31. Yao H et al. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev. 2010;24(22):2543–55.
32. Laver JD et al. Synthetic antibodies as tools to probe RNA-binding protein function. Mol Biosyst. 2012;8(6):1650–7.
33. Mattick JS. The functional genomics of noncoding RNA. Science. 2005;309(5740):1527–8.
34. Bond AM et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci. 2009;12(8):1020–7.
35. Poliseno L et al. Deletion of PTENP1 pseudogene in human melanoma. J Invest Dermatol. 2011;131(12):2497–500.
36. Mak AB et al. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. Mol Cell Proteomics. 2010;9(5): 811–23.
37. Wang X et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 2008;454(7200):126–30.
38. Sopher BL et al. CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. Neuron. 2011;70(6):1071–84.
39. Akhtar A, Becker PB. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell. 2000;5(2): 367–75.
40. Gupta RA et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6.
41. Zhao J et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322(5902):750–6.
42. Tsai MC et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689–93.
43. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev. 2009;23(16):1831–42.
44. Yang Z et al. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature. 2001;414(6861):317–22.

45. Kaneko S et al. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev. 2010;24(23):2615–20.
46. Shin E et al. Gene expression profiling of human hepatoblastoma using archived formalin-fixed and paraffin-embedded tissues. Virchows Archiv. 2011;458(4):453–65.
47. Jeon Y, Lee JT. YY1 tethers Xist RNA to the inactive X nucleation center. Cell. 2011;146(1):119–33.
48. Khalil AM et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA. 2009;106(28): 11667–72.
49. Jacquier A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet. 2009;10(12):833–44.
50. Flynn RA et al. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci USA. 2011;108(26):10460–5.
51. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature 2009;457(7232):1028–32.
52. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet. 2007;8(6):413–23.
53. Kapranov P et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316(5830):1484–8.
54. Taft RJ et al. Tiny RNAs associated with transcription start sites in animals. Nat Genet. 2009;41(5):572–8.
55. Preker P et al. RNA exosome depletion reveals transcription upstream of active human promoters. Science. 2008;322(5909):1851–4.
56. Neil H et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature. 2009;457(7232):1038–42.
57. Chekanova JA et al. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell. 2007;131(7):1340–53.
58. van Dijk EL et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature. 2011;475(7354):114–7.
59. Wyers F et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell. 2005;121(5):725–37.
60. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322(5909):1845–8.
61. Seila AC et al. Divergent transcription from active promoters. Science. 2008;322(5909): 1849–51.
62. Mattick JS. The genetic signatures of noncoding RNAs. PLoS Genet. 2009;5(4):e1000459.
63. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10(3):155–9.
64. Cabili MN et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18):1915–27.
65. Jia H et al. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. RNA. 2010;16(8):1478–87.
66. Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol. 2007;14(2):103–5.
67. Modarresi F et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotechnol. 2012;30(5):453–9.
68. Muse GW et al. RNA polymerase is poised for activation across the genome. Nat Genet. 2007;39(12):1507–11.
69. Ebisuya M et al. Ripples from neighbouring transcription. Nat Cell Biol. 2008;10(9): 1106–13.
70. Valen E et al. Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. Nat Struct Mol Biol. 2011;18(9):1075–82.
71. Hirota K et al. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature. 2008;456(7218):130–4.

72. Kanhere A et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell. 2010;38(5):675–88.
73. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. 2007;17(5):556–65.
74. Orom UA et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010;143(1):46–58.
75. Maenner S et al. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS Biol. 2010;8(1):e1000276.
76. Ginger MR et al. A noncoding RNA is a potential marker of cell fate during mammary gland development. Proc Natl Acad Sci USA. 2006;103(15):5781–6.
77. Pollard KS et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;443(7108):167–72.
78. Chakraborty D et al. Combined RNAi and localization for functionally dissecting long noncoding RNAs. Nat Methods. 2012;9(4):360–2.
79. Ulitsky I et al. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell. 2011;147(7):1537–50.
80. Zhao J et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010;40(6):939–53.
81. Yap DB et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood. 2011;117(8):2451–9.
82. Hung T et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011;43(7):621–9.
83. Yap KL et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell. 2010;38(5):662–74.
84. Pandey RR et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 2008;32(2):232–46.
85. Wang F et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. Nat Genet. 2011;43(4):345–9.
86. Nguyen VT et al. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature. 2001;414(6861):322–5.
87. Jeronimo C et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell. 2007;27(2): 262–74.
88. Kino T et al. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal. 2010;3(107):ra8.
89. Shamovsky I et al. RNA-mediated response to heat shock in mammalian cells. Nature. 2006;440(7083):556–60.
90. Yang L et al. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell. 2011;147(4):773–88.
91. Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet. 2011;12(2):136–49.
92. Tripathi V et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010;39(6):925–38.
93. Kotake Y et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011;30(16):1956–62.
94. Guetg C et al. Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. Mol Cell. 2012;45(6):790–800.
95. Shevtsov SP, Dundr M. Nucleation of nuclear bodies by RNA. Nat Cell Biol. 2011;13(2): 167–73.
96. Yoon JH et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012;47(4): 648–55.

97. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature. 2011;470(7333):284–8.
98. Cesana M et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell. 2011;147(2):358–69.
99. Kalyana-Sundaram S et al. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell. 2012;149(7):1622–34.
100. Rinn JL et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129(7):1311–23.
101. Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell. 2010;143(3):390–403.
102. Wutz A, Gribnau J. X inactivation Xplained. Curr Opin Genet Dev. 2007;17(5):387–93.
103. Jeon Y, Sarma K, Lee JT. New and Xisting regulatory mechanisms of X chromosome inactivation. Curr Opin Genet Dev. 2012;22(2):62–71.
104. Nora EP et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381–5.
105. Kind J et al. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell. 2008;133(5):813–28.
106. Sanchez-Elsner T et al. Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science. 2006;311(5764):1118–23.
107. Croce CM. LINCing chromatin remodeling to metastasis. Nat Biotechnol. 2010;28(9): 931–2.
108. Ishida M et al. Intrinsic nucleic acid-binding activity of chp1 chromodomain is required for heterochromatic gene silencing. Mol Cell. 2012;47(2):228–41.
109. Hu W et al. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev. 2011;25(24):2573–8.
110. Prensner JR et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011;29(8):742–9.
111. Meola N et al. The long noncoding RNA Vax2os1 controls the cell cycle progression of photoreceptor progenitors in the mouse retina. RNA. 2012;18(1):111–23.
112. Kretz M et al. Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes Dev. 2012;26(4):338–43.
113. Yu W et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 2008;451(7175):202–6.
114. Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol. 2011;21(6):354–61.
115. Huarte M et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409–19.
116. Cabianca DS et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell. 2012;149(4):819–31.
117. Scheibe M et al. Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. Nucleic Acids Res. 2012;40(19):9897–902.
118. Rederstorff M, Huttenhofer A. cDNA library generation from ribonucleoprotein particles. Nat Protoc. 2011;6(2):166–74.
119. Hafner M et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141(1):129–41.
120. Kapranov P et al. Large-scale transcriptional activity in chromosomes 21 and 22. Science. 2002;296(5569):916–9.
121. Rinn JL et al. The transcriptional activity of human chromosome 22. Genes Dev. 2003;17(4):529–40.
122. He Y et al. The antisense transcriptomes of human cells. Science. 2008;322(5909):1855–7.
123. Parkhomchuk D et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009;37(18):e123.
124. Cloonan N et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods. 2008;5(7):613–9.

125. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87–98.
126. Lipson D et al. Quantification of the yeast transcriptome by single-molecule sequencing. Nat Biotechnol. 2009;27(7):652–8.
127. Ozsolak F et al. Amplification-free digital gene expression profiling from minute cell quantities. Nat Methods. 2010;7(8):619–21.
128. Ozsolak F et al. Digital transcriptome profiling from attomole-level RNA samples. Genome Res. 2010;20(4):519–25.
129. Mamanova L, Turner DJ. Low-bias, strand-specific transcriptome illumina sequencing by on-flowcell reverse transcription (FRT-seq). Nat Protoc. 2011;6(11):1736–47.
130. Ozsolak F et al. Direct RNA sequencing. Nature. 2009;461(7265):814–8.
131. Kishore S et al. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods. 2011;8(7):559–64.
132. Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol. 2011;29(7):607–14.
133. Jungkamp AC et al. In vivo and transcriptome-wide identification of RNA binding protein target sites. Mol Cell. 2011;44(5):828–40.
134. Baltz AG et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell. 2012;46(5):674–90.
135. Castello A et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell. 2012;149(6):1393–406.
136. Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011;12(10):671–82.
137. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
138. Langmead B et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
139. Mortazavi A et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621–8.
140. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
141. Guttman M et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010;28(5): 503–10.
142. Trapnell C et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5): 511–5.
143. Garber M et al. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods. 2011;8(6):469–77.
144. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.
145. Mikkelsen TS et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553–60.
146. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
147. Valouev A et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods. 2008;5(9):829–34.
148. Wilbanks EG, Facciotti MT. Evaluation of algorithm performance in ChIP-seq peak detection. PLoS One. 2010;5(7):e11471.
149. Feng X, Grossman R, Stein L. PeakRanger: a cloud-enabled peak caller for ChIP-seq data. BMC Bioinformatics. 2011;12:139.
150. Song Q, Smith AD. Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinformatics. 2011;27(6):870–1.
151. Xu H et al. A signal-noise model for significance analysis of ChIP-seq with negative control. Bioinformatics. 2010;26(9):1199–204.

152. Zang C et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009;25(15):1952–8.
153. Rashid NU et al. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. 2011;12(7):R67.
154. Laajala TD et al. A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments. BMC Genomics. 2009;10:618.
155. Namekawa SH, Lee JT. Detection of nascent RNA, single-copy DNA and protein localization by immunoFISH in mouse germ cells and preimplantation embryos. Nat Protoc. 2011;6(3): 270–84.
156. Yoon JH, Srikantan S, Gorospe M. MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. Methods. 2012;58(2):81–7.
157. Ng K et al. A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells. Mol Biol Cell. 2011;22(14):2634–45.
158. Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. Science. 2011;333(6042):642–6.
159. Rinn JL, Huarte M. To repress or not to repress: this is the guardian's question. Trends Cell Biol. 2011;21(6):344–53.
160. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145–66.
161. Simon MD et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci USA. 2011;108(51):20497–502.
162. Chu C et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44(4):667–78.
163. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol. 2002;3(3):195–205.
164. Goodarzi H et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature. 2012;485(7397):264–8.
165. Underwood JG et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods. 2010;7(12):995–1001.
166. Kertesz M et al. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010;467(7311):103–7.
167. Ding F et al. Three-dimensional RNA structure refinement by hydroxyl radical probing. Nat Methods. 2012;9(6):603–8.
168. Westhof E, Romby P. The RNA structurome: high-throughput probing. Nat Methods. 2010;7(12):965–7.
169. Anko ML, Neugebauer KM. RNA-protein interactions in vivo: global gets specific. Trends Biochem Sci. 2012;37(7):255–62.
170. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol. 2007;8(6):479–90.
171. Bailey TL et al. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006;34(Web Server issue):W369–73.
172. Ray D et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat Biotechnol. 2009;27(7):667–70.
173. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. 1990;249(4968):505–10.
174. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990;346(6287):818–22.

# Identification of Chromatin-Binding Protein Complexes

**Alain Doucet and Daniel Figeys**

**Abstract** Cellular localization and activity of most proteins are highly dependent on the formation of complexes with other proteins and/or other biomolecules. Moreover, protein–protein interactions and the nature of the complexes formed are modulated in cell differentiation, de-differentiation, and carcinogenesis. Therefore, identification of protein interactors is essential to the understanding of protein biological activity in given cell states. Formation of DNA-binding complexes implicating proteins such as DNA repair enzymes, DNA- and histone-modifying enzymes, transcription factors, and gene activators, repressors, and silencers greatly affect DNA integrity and gene expression associated with multiple processes including cellular division and cancer development and progression. Protein–protein interaction mapping proved to be particularly difficult for chromatin-bound protein complexes due to the insolubility of these high molecular weight biomolecule assemblies. Recently, DNA-based methods were developed to isolate DNA-bound protein complexes formed at a specific locus. Techniques were also developed to purify the breath of proteins associated with a particular protein target using modified chromatin immunoprecipitation. The isolated proteins are then identified by tandem mass spectrometry and protein complexes are represented using dedicated softwares. In this chapter, we present the recently developed methods allowing for the isolation and identification of chromatin-bound protein complexes and to assign biological functions to these biomolecule. We also discuss the future challenges in the field of chromatin-bound protein interactor isolation and identification.

**Keywords** Chromatin-associated proteins • Modified chromatin immunoprecipitation • mChIP • Protein–protein interactions • Mass spectrometry

---

A. Doucet, Ph.D. • D. Figeys, Ph.D. (✉)
Department of Biochemistry Microbiology and Immunology,
Ottawa Institute of Systems Biology, University of Ottawa,
451 Smyth Road, Ottawa, ON, Canada, K1H 8M5
e-mail: douceta@medicago.com; dfigeys@uottawa.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_6,
© Springer Science+Business Media New York 2014

## Abbreviations

| | |
|---|---|
| CENPA | Centromer protein A |
| ChAP-MS | Chromatin affinity purification with mass spectrometry |
| CTCF | CCCTC-binding factor |
| E-MAP | Epistasis mini-array profile |
| GENECAPP | Global exonuclease-based enrichment of chromatin-associated proteins for proteomics |
| H3K4me3 | Histone H3 lysine 4 (K4) tri-methylation (me3) |
| HJURP | Holliday junction recognition protein |
| HSP90 | Heat shock protein 90 |
| Kog1p | Kontrol of growth protein |
| Lge1p | Depletion of large cell 1 |
| MChIP | Modified chromatin immunoprecipitation |
| NPM1 | Nucleophosmin1 |
| PICh | Proteomics of isolated chromatin segment |
| SMYD2 | SET and MYND-containing protein 2. |

Protein–protein interactions and the formation of protein complexes dictate the biological activity of most proteins. Very successful mass spectrometry-based methods developed to identify the set of proteins that can interact with a given protein (its interactome) [1–3] were applied to large-scale interactome mapping in various organisms [4–9]. These studies solved technically challenging aspects associated with protein–protein interaction studies and spur the development of bioinformatics tools to study protein networks and to query the interactome. Intrinsically, these large-scale studies rely on a cookie cutter approach in which one protocol is applied to all the bait proteins. Retrospective analysis of these datasets revealed gaps in the interactome for subgroup of proteins including membrane proteins and protein bound to macromolecules. In particular, we noticed that the interactome of DNA-associated proteins was often missing or only represented the in-solution interactions of the proteins rather than its interactions while associated with DNA. Moreover, the notion of protein complexes is evolving and includes complexes formed using a macromolecular backbone on which direct and indirect interconnections between protein complexes occurs. This is driven by the emergence of techniques that allows the study of macromolecular based connections including protein, DNA, and other biomolecules.

In eukaryotes, DNA is compacted in chromatin structures composed of repeated nucleosome units where 146 DNA base pairs are wrapped around an octamer of histone (two copies of histone H2A, H2B, H3, and H4 and/or their isoforms) [10]. We defined the interaction of DNA-associated proteins into three subgroups (Fig. 1). The first subgroup represents the interactions of a bait protein while it is free in solution. These can be interactions required to regulate association of the bait to DNA (or other proteins on DNA) or interactions corresponding to other functions not directly associated with DNA (Fig. 1). For example, although the SET and

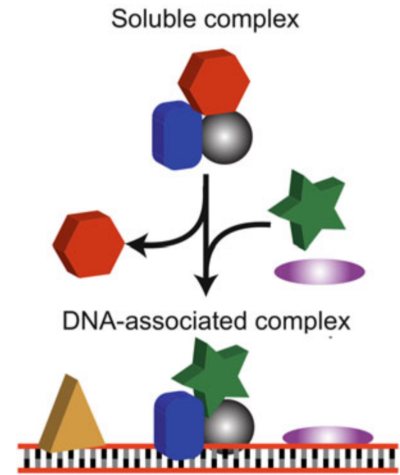



**Fig. 1** Type of protein complexes. Chromatin-binding proteins can be found within protein complexes (1) free in solution, (2) directly interacting with DNA-bound proteins, or (3) recruited to DNA and indirectly interacting with DNA-bound proteins (*purple ellipse*). Other noninteracting, generic DNA-binding proteins can be found in the vicinity of the complex (*yellow triangle*)

MYND-containing protein 2 (SMYD2) associates with DNA to methylate histone H3K4 [11], it also methylates heat shock protein 90 (HSP90) [12] independently of the presence of DNA. The second subgroup of interactions is composed of the direct interactors of the bait proteins while it is associated with DNA (Fig. 1). Finally, the third subgroup of interactors is composed of proteins not physically attached to the bait proteins but rather proteins that are indirectly associated through the common DNA strand. These "neighboring" proteins either generically appears on DNA, but others correlate with the presence of the bait protein (Fig. 1).

One important challenge for the isolation and identification of bait proteins in complex with DNA is maintaining the complex while using techniques to solubilize chromatin. Because of its large size, chromatin and its associated proteins are eliminated with other insoluble material during classical immunopurification involving clarification of cell extracts. Moreover, classical approaches to isolate DNA-bound proteins such as acid and high-salt extraction perform well to enrich proteins such as histone [13, 14]. However, these approaches do not maintain the protein/chromatin structure (denaturing conditions) and therefore are not suitable to identify histone interactors and the interactions of other DNA associated proteins. Here we present methods developed specifically for the identification of protein complexes associated with chromatin.

## Nucleic Acid Probes

Immobilized DNA probes were first used in affinity chromatography to purify DNA-binding proteins [15]. This general principle was further refined to identify protein complexes bound to chromatin. The goal of those methods is to define the set of proteins interacting with a defined, exogenous, short DNA sequence. The overall concept is presented in Fig. 2 and examples of such methods are presented in this section.

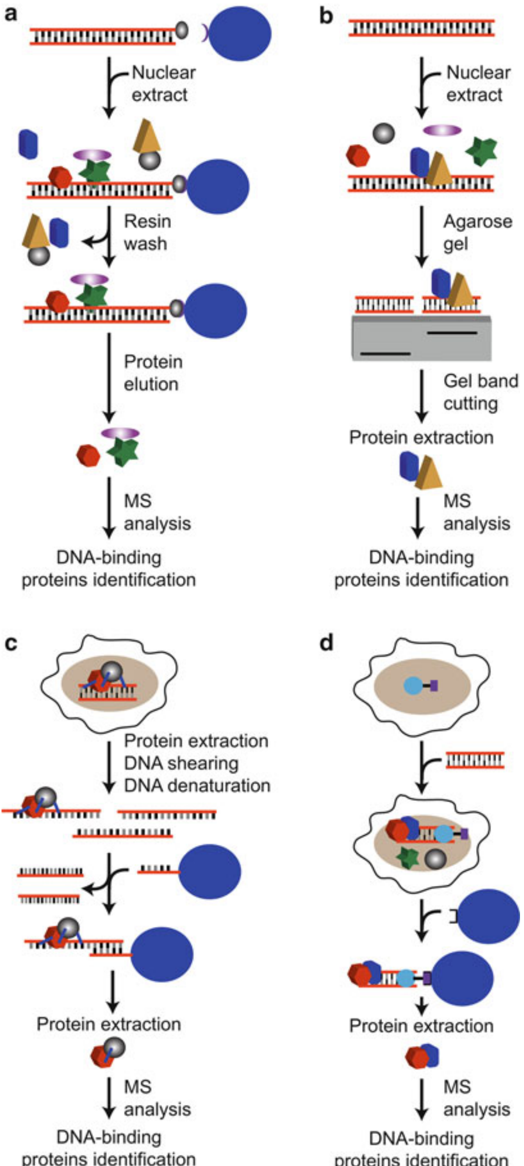



**Fig. 2** DNA probes to identify DNA-binding protein complexes. (**a**) A double-stranded, biotin conjugated (*gray sphere*) DNA probe is immobilized on straptavidin-conjugated beads (*blue circles*). Modified histones (acetylated for example) can be incorporated to the DNA probe to generate chromatin and screen for proteins binding to these modifications if desired. The immobilized DNA probe is incubated with nuclear-extract proteins (represented by the different forms), noninteracting proteins are washed, and DNA-binding proteins are eluted and identified by mass spectrometry. (**b**) DNA probe is incubated with nuclear extract and the mixture is separated on agarose gel (*gray slab*). Electrophoretic shifts to higher molecular weights are observed when protein complexes are bound to DNA. Agarose gel bands are cut, proteins are extracted, and identified by mass spectrometry.

## *Reconstitution of DNA-Bound Complexes In Vitro*

In the nucleus, distant chromatin segments are sometimes brought together and reside in the same nuclear area. Insulators and boundary elements are involved in the clustering of DNA in "nuclear territories" characterized by specific expression profile. DNA segregation is mediated by long-range chromatin interactions involving DNA-bound protein complexes. CTCF (CCCTC-binding factor, CCCTC being DNA bases) is the only known protein with insulator functions in vertebrates, but its mode of action was mostly unknown. In an effort to identify CTCF interactors, Rubio et al. designed an experiment to enrich CTCF and its interacting partners using DNA probes [16]. Biotinylated DNA templates of 163 bp containing the CTCF-binding sequence of the c-myc insulator element sequence were generated by PCR and coupled to streptavidin-linked magnetic beads (Fig. 2a). Negative control templates with mutations in the CTCF binding site were also produced. These immobilized DNA probes were incubated with nuclear extracts to capture CTCF along with its binding partners. Proteins from the specific DNA probe and the control probe were eluted and analyzed by quantitative mass spectrometry. The cohesin subunit Scc3/SA1 was found to interact with CTCF and was greatly enriched in the CTCF-binding probe sample compared to the negative control. Supporting this result, cohesin subunit Scc3/SA1 also colocalized with a subset of genomic CTCF binding sites, as demonstrated by ChIP-Chip large-scale experiment. Cohesin is essential for chromatid cohesion in mitotic metaphase and CTCF was found in centromeric region of metaphase chromosomes using immunofluorescence. These results suggest a role for CTCF sister chromatide cohesion [16]. Thus, immobilized DNA probes proved very efficient and useful as a screening tool for the enrichment and identification of chromatin-bound protein complex.

In another study, small double-stranded oligonucleotide probes with an estrogen response element sequence were used to identify large protein complexes recruited to DNA by the estrogen receptor α [17]. The DNA probe was incubated in vitro with the recombinant estrogen receptor α and nuclear extracts from HeLa cells (Fig. 2b). DNA–protein complexes were separated from unbound probes and free proteins on an agarose gel. Shifts in electrophoretic migration pattern were observed between free and protein-bound probes, indicating that protein–DNA complexes were

**Fig. 2** (continued) (**c**) Cells are fixed with formaldehyde, which crosslink protein complexes to DNA. Cells are lysed, DNA is fragmented and denatured, but proteins remains attached to a DNA strand due to crosslinking. A single-stranded DNA probe complementary for a sequence of interest is coupled to beads and used to purify the DNA of interest with its coupled proteins. DNA-binding proteins are then identified by mass spectrometry. (**d**) Cells expressing a flag-tagged Lac1 protein (*light blue circle with purple box attached*) are transfected with a minichromosome with a binding sequence for the protein of interest and another sequence for Lac1. Then minichromosome is replicated by the cells and binds protein complexes at one locus and the flag-tagged Lac1 at another. The cells are lysed and the minichromosome is affinity-isolated using the tagged Lac1 protein. Proteins binding to the minichromosome are identified by mass spectrometry

maintained intact during electrophoresis. Probe bands shifted to higher apparent molecular masses were cut, and the proteins were extracted and analyzed by mass spectrometry. Large interconnected enzyme networks involved in DNA repair and members of the INHAT complex interacting with the estrogen receptor α were identified in this study [17].

Instead of being developed around specific DNA sequence, some protein complex formations are triggered by posttranslational modifications of histones N-terminal tail. Tri-methylation of histone H3 lysine 4 (H3K4me3) and 9 (H3K9me3) is associated with activation and repression of transcription, respectively. The number and sites of these modifications are modulated upon carcinogenesis. To identify proteins specifically interacting with differently marked histone H3, biotinylated, uniformly modified oligonucleosmes composed of a double-stranded DNA segment wrapped around the core histones H2A, H2B, H4, and H3 with H3K4me3, H3K9me3, or unmodified H3 were separately generated [18]. These oligonucleosmes were immobilized on streptavidin beads, isotope-labeled Hela cell extracts were added to the oligonucleosomes, and proteins interacting with the oligonuclosomes were identified by affinity purification and quantitative mass spectrometry. This method identified proteins directly interacting with modified histone H3 as well as indirect interactors recruited to the site by the direct interactors. Proteins interacting with the unmodified histone H3, but excluded by histone H3 trimethylation were also identified. Interestingly, the set of interactors identified using modified histone H3 embedded in the oligonucleosome was significantly different from interactors of modified histone H3 tail peptides and mono-nucleosomes, indicating that the histone H3 environment greatly influences its interacting partners [18].

Although very successful in the context of the three studies describe here, these methods are based on the assumption that the nuclear or cellular extracts contain a reasonable amount of the soluble form of the targeted protein and its interactors, which might not be the case for many nuclear proteins binding strongly to DNA. The two first methods also require previous knowledge of the DNA binding site for the protein of interest and thus, cannot be applied to proteins with unknown DNA binding sequence. More studies are needed to establish whether this approach can be used for the systematic study of protein complexes associated with DNA.

### *Isolation of DNA–Protein Complexes Generated In Vivo*

Déjardin et al. also used DNA probes to isolate chromatin but developed a very different workflow for their method named Proteomics of Isolated Chromatin Segment (PICh) [19]. In PICh, protein–protein and protein–DNA interactions are stabilized by crosslinking with formaldehyde during cell fixation (Fig. 2c). Cells are then lysed, the chromatin is fragmented by sonication, denatured, and hybridized with sequence specific, single-stranded, biotin-labeled DNA probes containing modified nucleotides. Protein–chromatin–DNA probe complexes are captured on streptavidin-coupled magnetic beads and proteins are identified by mass spectrometry. Compared

to the two previous methods described above, PICh identifies protein complexes formed inside the cells instead of test tubes and the DNA probes are used to retrieve chromatin fragments with bound proteins instead of being the bait to isolate DNA-binding proteins. Using PICh, the authors identified multiple telomere binding proteins including orphan nuclear receptors [19]. One limitation of this method is the isolation of sufficient material for identification of proteins by mass spectrometry. This can be challenging for the study of rare loci, organisms with complex genome, or when limited amount of starting material is available. In the study described here, $3 \times 10^9$ cells (equivalent to 3 L of Hela S3 cell culture) are required for each purification, despite the high frequency of the telomere loci (2 per chromosome). Therefore, PICh is best suited to study abundant genomic loci or for organisms with low genetic complexity and easy to grow in large quantity such as yeast and bacteria.

Another similar technique to PICh is the Global Exonuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics (GENECAPP). In this approach, formaldehyde is used to crosslink DNA and its associated proteins, and the DNA is then fragmented. Then exonuclease III is used to generate single-stranded DNA–protein complexes [20]. In GENECAPP, complexes are captured on a nucleotide array that allows screening of multiple DNA sequences on a modified microscope slide. Captured protein complexes are trypsinized on-chip and peptides are analyzed by mass spectrometry to identify interactors. The potential of this approach was demonstrated for the study of the FoxO1-IGFBP1 promoter region complex. *In vitro* reconstitution of the complex followed by capture on tilling arrays was used to demonstrate the feasibility of this approach.

A different strategy was developed to specifically isolate protein complexes involved in DNA segregation. Chromosome segregation is a crucial aspect of cell division and allows even distribution of the chromosomes to the dividing cells. Chromosome segregation involves formation of the kinetochore: a protein macromolecular complex located at centromeric chromatin loci that attach to microtubules. Although systematic studies of the kinetochore have been performed [21, 22], the full repertoire of proteins participating to the kinetochore is still unknown. Moreover posttranslational modifications, in particular protein phosphorylations, are key regulators of the dynamic of the kinetochore. As well, many articles have studies the role of protein phosphorylations in kinetochore dynamic [23], but the repertoire of phosphorylations and their temporal modulation is still incomplete. This is in part due to difficulties in isolating sufficient amount of kinetochore protein complexes with sufficient purity. In order to tackle this problem, Akiyoshi et al. used yeast as model organism and devised a 2 kb circular minichromosome containing the yeast centromere from chromosome 3 and the lactose operon purification system [24] (Fig. 2d). This system allows for the affinity purification of the minichromosome from a yeast strain expressing a Flag-tagged version of Lac1. Minichromosomes replicate and segregate in cells and can be used for the isolation of kinetochore protein complexes bound to the centromere element of the minichromosome. The authors identified 35 out of the 38 previously known kinetochore associated proteins using this method and 10 phosphorylation sites on 7 proteins by coupling this

minichromosome purification system to quantitative tandem mass spectrometry. More interestingly, the protein Fin1 was identified as a new element of the kineto-chore. Fin1 interacts with protein phosphorylase 1 (PP1) and regulates its association to the kineotochore. Fin1–PP1 complex formation is inhibited upon phosphorylation of Fin1 by Cdk1 [25] and association of phosphorylated Fin1 with the 14-3-3 proteins [26]. Proper regulation of Fin1 through phosphorylation is essential for proper formation of kinetochore and spindle function [24].

## Immunoaffinity Purification of Chromatin-Bound Complexes

While methods based on DNA probes are very good at identifying proteins interacting at a precise DNA loci, to date their design does not permit the broad identification of interactors of a given nuclear protein throughout the genome. Other research groups have focused on the purification of protein complexes while maintaining their association to DNA (Fig. 3).

One approach to maintain the protein–DNA complex and solubilize the DNA is to introduce micrococcal nuclease digestion during the purification of the complex. This was demonstrated by Sullivan et al. [27, 28] who, using this approach,

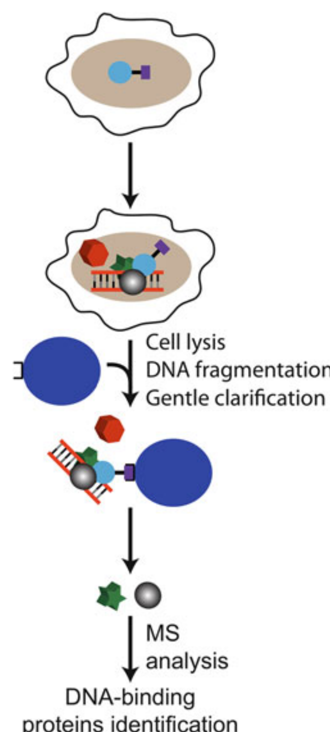



**Fig. 3** Immuno-affinity purification of DNA-bound protein complexes. Cells are engineered to express a protein of interest (*yellow circle*) fused to a tag (*purple hexagon*). The protein of interest interacts with DNA-bound protein complexes. Cells are lysed, the DNA is fragmented either by sonication or digestion with nucleases, and the cell lysate is clarified by low speed centrifugation to maintain DNA-bound complexes in solution. The clarified extract is incubated with beads coupled to an antibody specific for the tag on the protein of interest. Interactors of the protein of interest are immunopurified and identified by mass spectrometry

identified the mammalian protein complex attached to centromer protein A (CENP-A), which replaces the histone H3 protein in nucleosomes of active centromers. Briefly, in that study Hela cells stably expressing a TAP-tagged CENP-A (or histone H3.1 as control) were generated to allow for immunopurification of CENP-A protein complexes [29]. The stable cell lines were lysed and their genomic DNA was digested with micrococcal nuclease. This digestion step is very important as it permits the solubilisation of DNA–protein complexes assembly by releasing mostly single, soluble nucleosomes with their attached intact protein complexes. Affinity purification of CENP-A using the TAP-tag method allowed for the isolation of nucleosomes from active centromers. Proteins extracted from the CENP-A interactors were identified by mass spectrometry. Known (CENP-B, -C, -H, -U) and new (CENP-M, -N, -T) CENP-A interactors were identified. CENP-A containing nucleosomes were also enriched in histone H2A.Z and macroH2A [29]. The importance of the DNA digestion by micrococcal nuclease was later demonstrated by the same author [29]. Omitting the nuclease digestion step and pelleting insoluble genomic DNA resulted in isolation of soluble TAP-Tagged CENP-A complexed to a completely different set of interactors, namely HJURP (Holliday JUnction Recognition Protein) and NPM1 (Nucleophosmin1) [30]. These two proteins associate with CENP-A and histone H4 as pre-nucleasome soluble complex and HJURP participates to the recruitment of CENP-A to nucleosomes of active centromers. This clearly demonstrated that the interactors of a bait protein can be drastically different when it is attached to DNA compared to when it is in solution. As well, it demonstrated the importance of solubilizing DNA–protein complexes while maintaining the interaction of the proteins with DNA in order to identify chromatin-bound complexes.

In 2009, we developed a method named modified Chromatin ImmunoPrecipitation (mChIP) to isolate DNA fragments with chromatin-bound protein complexes. The mChIP development was based on the chromatin immunoprecipitation (ChIP) genomic method where affinity purification of DNA–protein complexes leads to the identification of DNA binding sites. In mChIP, Tap-tagged DNA-binding proteins are overexpressed in yeast [31]. The cells are then collected, lysed, and the DNA is fragmented by sonication, generating fragments <2,000 bp carrying protein complexes. The sample is then clarified at low centrifugal force to maintain DNA–protein complexes in solution. The TAP-tagged protein of interest is isolated by affinity purification and the members of the protein complex are identified by tandem mass spectrometry. mChIP identified 98 interactors of the yeast histone protein Hta2p during the validation of the method. To put this number in perspective, only 42 Hta2p interactors were listed in the Biogrid database at the time of the experiment [32]. This demonstrates the sensitivity of mChIP to isolate chromatin-bound proteins compared to the regular immunoprecipitation methods. It was also possible to amplify the *GAL1* promoter region by PCR from affinity purified Htz1p with mChIP. The HSP12 promoter region was successfully amplified by PCR using mChIP technology in another study [33], demonstrating that whole chromatin sections are isolated with this technology. Based on this success, the authors tackled a more difficult problem: the identification of protein complexes associated to Lge1, Mcm5p, and

Yta7p, three chromatin-bound proteins with interactors that proved difficult to identify by conventional affinity purification. Depletion of Large Cell 1 (Lge1p) leads to large yeast cells, but the mechanism of action of this protein was unknown and only two proteins were known to interact with Lge1p. We identified 40 specific interactors of Lge1p including Kontrol of Growth (Kog1p), which is involved in the control of yeast cell growth. Immunoprecipitation of Mcm5p and Yta7p largely increased our understanding of the biological roles of these proteins through the identification of 13 and 23 interactors, respectively [31]. The mChIP protocol was later used in the first large-scale study of chromatin-bound protein complex mapping where 102 C-terminally TAP-Tagged proteins were used to identified 2,966 high confidence interactions involving 724 proteins [34]. mChIP identified more interactors for 75 % of the baits when compared to conventional large scale interactome study and 18 % of the baits that previously failed to identify any interactors in previous studies returned results using mChIP [34]. This demonstrates the power of mChIP to identify new chromatin-associated networks of protein interactions. Moreover, Lambert et al. also demonstrated that using different enzymes (MNase and DNase) the isolated stretch of DNA fragment could be reduced to the core complex to help differentiate direct interactors versus neighbors through DNA binding.

The current disadvantage of the Immunoprecipitation of chromatin-bound complexes is that it results in the global identification of interactors of a given protein and does not discriminate between loci. Some proteins participate to the formation of different protein complexes with opposite physiological activity (activator vs. repressor for example). Those proteins are recruited to various loci depending on the presence of other members in the complex. As well, global identification of interactors does not indicate the function of the protein under study at different genomic locations. Standard chromatin immunoprecipitation and mChIP could be used to identify both the breath of proteins interacting with a bait as well as its genomic localization. Based on this information, DNA probes or minochromosome technologies could be used to identify the subset of interactors at precise loci to better understand the role of the protein under study. However, this workflow would be difficult to apply to high throughput studies and would require significant amount of starting material.

## Immunopurification of Protein Complexes at Specific Chromatin Loci

Techniques have also been developed to study protein complexes that occur at specific loci within the genome. An example of such an approach was provided by Butala et al. who used a plasmid containing *E. coli* promoter region of the colicin K gene (*cka*) coupled the LacI repressor DNA-binding sequence, engineered to identify protein complexes binding to the *cka* gene reculatory region [35]. The plasmid is devised to release a linear DNA fragment composed of the *cka* promoter and LacI binding sequence in the *e coli.* cells. These particular *E. coli* cells also expressed a

Flag-tagged LacI protein that binds to the linear DNA LacI-binding sequence. This linear DNA-bound Flag-tagged LacI is used as a handle to allow for immunoaffinity purification of the linear DNA fragment and identification of the proteins bound to the *cka* promoter on the same linear DNA. This method can be used to study other *E. coli* DNA sequences and is facilitated by the use of this prokaryotic model that does not require targeting of the DNA fragment under study to the nucleus.

Byrum et al. developed a method called Chromatin Affinity Purification with Mass Spectrometry (ChAP-MS) whereby protein complexes at a specific DNA locus in eucaryote are purified and identified by mass spectrometry [36]. In ChAP-MS, a LexA DNA binding site is added upstream of the gene of interest and this construct is inserted in the yeast genome by homologous recombination and replaces the endogenous copy of the gene. A LexA-Protein A fusion protein expressed by the engineered yeast binds to the LexA DNA-binding site and, following cross-linking of DNA–Protein complexes and sonication of the chromatin in ~1,000 bp fragments, the gene of interest is purified using Protein A. Protein complexes at this locus are identified by tryptic digestion and quantitative mass spectrometry analysis. Protein complexes at the *GAL1* locus in yeast grown with galactose vs. lactose media were identified with ChAPS-MS. ChAP-MS also isolates chromatin segments containing core histone proteins and thus allows for the identification of histone tails modifications at the targeted locus using the same mass spectrometry data. Multiple acetylation and methylation marks were identified in histones N-terminal tails in this study [36]. However, N-terminal tails of histones have tightly spaced lysine and arginine residues and tryptic digestion (trypsin cleaves C-terminal of lysine and arginine) of histone tails generate very small peptides not amenable to mass spectrometry. Moreover, lysines and arginines are subjected to posttranslational modifications (phosphorylation, acetylation, methylation, ubiquitinylation, etc.) and these affect trypsin cleavage specificity and turn-over, thus complicating the tryptic digestion. Identification of histone tails modifications in ChAPS-MS could be greatly improved by using proteases other than trypsin to generate peptides for mass spectrometry analysis. Genomic insertion of the LexA binding site at the locus of interest was also used by Fukita and Fujii to identify components of the chicken insulator HS4 (cHS4) [37]. A DNA segment containing 24 copies of cHS4 core sequence with LexA DNA-binding sequences was inserted in the genome of a mouse cell line (Ba/F3) engineered to express a modified LexA protein that will be used for immunoprecipitation of the cHS4 DNA. The RNA helicase p68/DDX5, the matrix protein Matrin-3 and the RNA species SAR1 were identified at the cHS4 locus using this insertional chromatin immunoprecipitation (iChIP) method. This demonstrates that genomic insertion of LexA DNA-binding sequence to isolate specific DNA loci is not only applicable to yeast but also to mammalian cell in culture. However, considering the high number of gene copies inserted at a nonbiological site in the cell genome the biological significance of the findings using this system have to be validated. The methods presented in this section identify protein complexes at a given DNA locus but do not identify specific interactors of a protein of interest at this locus. A combination of DNA locus isolation followed by immunoprecipitation of the protein of interest would be suitable for this purpose, but would require large amount of starting material.

## Assigning Biological Functions to Protein–Protein Interactions

An interesting approach to associate protein–protein interactions to biological function named epistasis mini-array profile (E-MAP) was developed by Collins SR et al. in 2007. E-MAP is based on growth rates measurements of *Saccharomyces cerevisiae* cells carrying pairs of mutations in a gene subset associated to a biological process [38]. Comparing experimental growth rate of *S. cerevisiae* cells carrying mutations in two genes (double mutants) to the theoretical growth rate based on the additive effect of the two single mutations reveals genetic interactions. This method allows for the identification of (a) negative interactions where the deletion of two genes has a larger impact than the individual deletions, indicating a synergistic effect of the two genes and (b) positive interactions where the double mutant fare better than the two single mutants (rescue effect) or is comparable to the single mutants (the two proteins act in a common complex or pathway). Hierarchical clustering of pairwise interactions of 743 genes (754 alleles) participating to chromosome biology using E-MAP segregates genes based on their biological functions. For example genes involved in general DNA replication are distinguished from genes participating in DNA replication checkpoint complexes, and from genes associated with sensing and repair of damaged DNA. By comparing protein physical interaction maps derived from high-throughput affinity purification experiments to genetic interaction clustering generated with E-MAP it is now possible to identify sub-groups or modules composed of proteins involved in a common biological function. Protein from this complex with distinct or opposing functions can also be identified. Analysis of the 25 protein complex named Mediator identified modules mediating a common function as well as one module with opposing actions on the other module functions [38]. Therefore, by combining physical and genetic interactions it is now possible to gain a better understanding of the biological roles of protein complexes. This approach is very interesting when using the budding yeast as model organism, but is difficult (or impossible) to apply to higher organisms such as mice and humans due to the limitation to generate large number of double mutants.

## Future Challenges

Many challenges still lie ahead for the study of DNA-associated protein complexes. Especially, the studies described here all required large amount of starting material and used either model organisms or cell cultures. Low sensitivity associated to identification of chromatin-bound protein complexes still prevents the use of limited amount of starting material such as stem cells and tissues. In cancer cells, chromatin-bound protein networks are expected to be highly modified and it would be interesting to identify the DNA-bound protein complexes affected in human biopsies from cancer patients. Stem cells and primary cells are also very interesting to study, but

are often hard to cultivate in large amount and thus, are not yet amenable to chromatin-bound protein complexes studies using current approaches. Another limitation associated with low sensitivity in most studies described here is the use of overexpressed tagged proteins. The unusually high concentration of a given protein due to overexpression could modify the equilibrium of protein complex formation and force protein associations not seen with endogenous level of the same protein. This would results in identification of physiologically irrelevant protein complexes. Possible cellular miss-localization of overexpressed proteins is another problem of this system and can also result in non-natural protein complex identification. An apparently simple solution would be the use of endogenous proteins, but this is not a trouble-free approach either. Apart from sensitivity issues, high affinity, high specificity antibodies are not available for many proteins of interest, greatly limiting this approach. In cell culture, the challenge ahead is to be able to identify chromatin-associated complex at specific loci and study the dynamic changes in these complexes and loci.

Our ability to identify chromatin-associated protein networks has improved drastically in recent years due to novel methods and to technological developments. In some instances it is feasible to even perform large-scale identification of complex associated with DNA and better understand their roles and functions. It is foreseeable that future laboratory protocol improvements and engineering advances will overcome current limitations and allow for a deeper understanding of chromatin-bound complex formation, their dynamics of assembly in defined cell states, and their biological roles in homeostasis and disease states.

**Acknowledgments** D.F. Acknowledges a Canada Research Chair in Proteomics and Systems Biology.

## References


1. Gavin A-C, Maeda K, Kühner S. Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. Curr Opin Biotechnol. 2011;22(1):42–9.
2. Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nat Rev Genet. 2009;10(9):617–27.
3. Gingras A-C, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol. 2007;8(8):645–54.
4. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440(7084):637–43.
5. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084):631–6.
6. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams S, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 2002;415:2–5.
7. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 2007;3(89):89.
8. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W, et al. Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. PLoS Biol. 2009;7(4):e96.
9. Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, et al. Proteome organization in a genome-reduced bacterium. Science. 2009;326(5957):1235–40.

10. Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. Cell. 2004;116:259–72.
11. Abu-Farha M, Lambert J-P, Al-Madhoun AS, Elisma F, Skerjanc IS, Figeys D. The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. Mol Cell Proteomics. 2008;7(3):560–72.
12. Abu-Farha M, Lanouette S, Elisma F, Tremblay V, Butson J, Figeys D, et al. Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. J Mol Cell Biol. 2011;3(5):301–8.
13. Galasinski S. Protein mass analysis of histones. Methods. 2003;31(1):3–11.
14. Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. Nat Protoc. 2007;2(6):1445–57.
15. Kadonaga JT, Tjian R. Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci USA. 1986;83(16):5889–93.
16. Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, et al. CTCF physically links cohesin to chromatin. Proc Natl Acad Sci USA. 2008;105(24):8309–14.
17. Schultz-norton JR, Ziegler YS, Likhite VS, Yates JR, Nardulli AM. Isolation of novel coregulatory protein networks associated with DNA-bound estrogen receptor alpha. BMC Mol Biol. 2008;12:1–12.
18. Nikolov M, Stützer A, Mosch K, Krasauskas A, Soeroes S, Stark H, et al. Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. Mol Cell Proteomics. 2011;10(11):M110.005371.
19. Déjardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. Cell. 2009;136(1):175–86.
20. Wu C, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Brian L, et al. Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. PLoS One. 2011;6(10):e26217.
21. Takeuchi K, Fukagawa T. Molecular architecture of vertebrate kinetochores. Exp Cell Res. 2012;318(12):1367–74.
22. Gascoigne KE, Cheeseman IM. Kinetochore assembly: If you build it, they will come. Curr Opin Cell Biol. 2012;23(1):102–8.
23. Ma HT, Poon RYC. How protein kinases co-ordinate mitosis in animal cells. Biochem J. 2011;435(1):17–31.
24. Akiyoshi B, Nelson CR, Ranish JA, Biggins S. Quantitative proteomic analysis of purified yeast kinetochores identifies a PP1 regulatory subunit. Genes Dev. 2009;23(24):2887–99.
25. Woodbury EL, Morgan DO. Cdk and APC activities limit the spindle-stabilizing function of Fin1 to anaphase. Nat Cell Biol. 2007;9(1):106–12.
26. van Hemert MJ, Deelder AM, Molenaar C, Steensma HY, van Heusden GPH. Self-association of the spindle pole body-related intermediate filament protein Fin1p and its phosphorylation-dependent interaction with 14-3-3 proteins in yeast. J Biol Chem. 2003;278(17):15049–55.
27. Sullivan BA, Blower MD, Karpen GH. Determining centromere identity: cyclical stories and forking paths. Nat Rev Genet. 2001;2(8):584–96.
28. Cleveland DW, Mao Y, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell. 2003;112(4):407–21.
29. Foltz DR, Jansen LET, Black BE, Bailey AO, Yates JR, Cleveland DW. The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol. 2006;8(5):458–69.
30. Foltz DR, Jansen LET, Bailey AO, Iii JRY, Emily A, Wood S, et al. Centromere specific assembly of CENP-A nucleosomes is mediated by HJURP. Cell. 2010;137(3):472–84.
31. Lambert J-P, Mitchell L, Rudner A, Baetz K, Figeys D. A novel proteomics approach for the discovery of chromatin-associated protein networks. Mol Cell Proteomics. 2009;8(4):870–82.
32. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, et al. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res. 2011;39((Database issue)):D698–704.

33. Mitchell L, Lambert J-P, Gerdes M, Al-Madhoun AS, Skerjanc IS, Figeys D, et al. Functional dissection of the NuA4 histone acetyltransferase reveals its role as a genetic hub and that Eaf1 is essential for complex integrity. Mol Cell Biol. 2008;28(7):2244–56.
34. Lambert J-P, Fillingham J, Siahbazi M, Greenblatt J, Baetz K, Figeys D. Defining the budding yeast chromatin-associated interactome. Mol Syst Biol. 2010;6(448):448.
35. Butala M, Busby SJW, Lee DJ. DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. Nucleic Acids Res. 2009;37(5):e37.
36. Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep. 2012;2(1):198–205.
37. Fujita T, Fujii H. Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS One. 2011;6(10):e26109.
38. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature. 2007;446(7137):806–10.

# Exploring Chromatin Readers Using High-Accuracy Quantitative Mass Spectrometry-Based Proteomics

**Arne H. Smits and Michiel Vermeulen**

**Abstract** In recent years, quantitative mass spectrometry-based proteomics has proven itself as a powerful technology to study chromatin structure and function in eukaryotic cells. Multiple methodologies have been developed which enable a comprehensive identification of chromatin readers and the characterization of the dynamic protein complexes these readers assemble in. These advancements in technology have made a big impact in the field of chromatin biology and have led to new fundamental insights. In this chapter, we will discuss the quantitative mass spectrometry-based methodologies used for identification and characterization of chromatin readers and the new biological insights that these approaches have generated.

**Keywords** Epigenetics • Chromatin readers • Quantitative mass spectrometry • Proteomics

## Introduction

During the last two decades, our molecular understanding of chromatin structure and function has increased tremendously. Nucleosomes, which form the basic repeating unit of chromatin, were previously thought to merely serve for compaction and storage of DNA inside eukaryotic cells. This view has radically changed and

A.H. Smits, M.Sc. • M. Vermeulen, Ph.D. (✉)
Department of Molecular Cancer Research, University Medical Center Utrecht,
Startenum 3.126, Universiteitsweg 100, Utrecht, 3584 CG, The Netherlands

Department of Proteomics and Chromatin Biology,
University Medical Center Utrecht, Utrecht, The Netherlands
e-mail: A.Smits-4@umcutrecht.nl; M.Vermeulen-3@umcutrecht.n

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_7,
© Springer Science+Business Media New York 2014

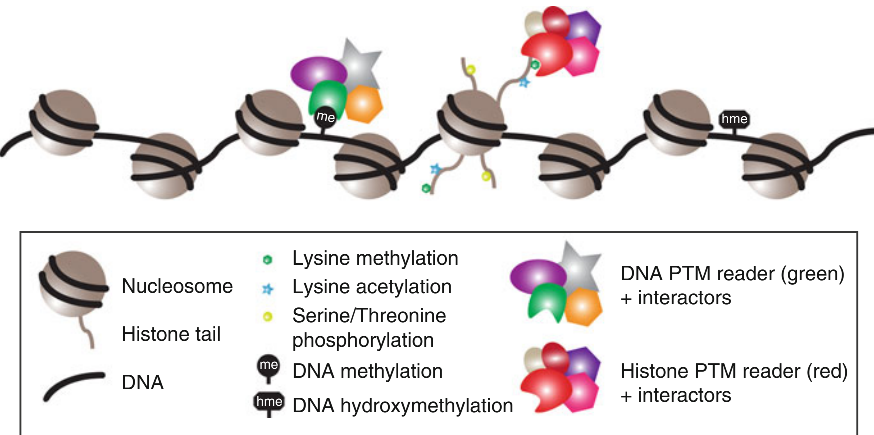



**Fig. 1** Graphical representation of chromatin structure and chromatin-associated complexes. Nucleosomes form the basic repeating unit of chromatin and both the histone tails and the DNA can be modified. The chromatin can be specifically bound by “readers” that have a strong affinity for either the modified or the non-modified molecule. These readers assemble in protein complexes

nucleosomes are now appreciated to actively regulate nuclear processes such as replication, transcription, and DNA repair. Chromatin-associated proteins, which interact with nucleosomes and some of which can modify them, form a major downstream target in signal transduction pathways and chromatin therefore plays a central role in cell cycle control and mitosis, growth factor signaling, and in stress response pathways. Fundamental in these processes are the nucleosomes and the posttranslational modifications (PTMs) on the N-terminal histone tails that protrude from the nucleosome. These tails are subjected to a large number of PTMs, such as acetylation, phosphorylation, methylation, and ubiquitination [1] (Fig. 1). In addition, the DNA itself can be modified through (hydroxy)methylation of cytosine residues (Fig. 1). These modifications can affect gene expression and cellular phenotype, and modification patterns can be inherited from mother to daughter cells. Histone PTMs and DNA (hydroxy)methylation therefore provide epigenetic information which, together with genetic information embedded in DNA, determines the phenotype of a eukaryotic cell or organism.

One of the major downstream functions of histone PTMs is the recruitment or stabilization of effector proteins which are also called “readers” [2] (Fig. 1). The biological function of these readers often correlates to the biology of the epigenetic mark they bind to; indicating that the “reading” function of epigenetic modifications is rather important [3]. Identifying and characterizing chromatin readers and the (dynamic) protein complexes that these readers assemble in is therefore crucial to further our understanding of epigenetic modifications and their role in determining gene expression and cell fate. We have recently developed and applied a number of methods based on quantitative mass spectrometry-based proteomics technology

that can be used to identify and characterize chromatin readers. In this chapter the workflow and methodology behind these methods as well as the biological insights that these approaches have generated will be discussed in detail.

## Quantitative Mass Spectrometry-Based Proteomics to Decipher the Chromatin Interactome

As mentioned above, core histones are subjected to a large number of PTMs, such as lysine acetylation and lysine and arginine methylation [1]. Most of these site-specific modifications are associated with particular functional chromatin states. For example, trimethylation of histone H3 on lysine four (H3K4me3) is associated with promoters of genes that are being actively transcribed. In contrast, H3K27me3 is associated with transcriptional repression. Similarly, ubiquitination of H2B at lysine 120 is linked to activation of transcription, whereas H2A ubiquitination at lysine 119 is linked to gene repression [4]. Other modifications, such as phosphorylation of H3S10 and H3S28 are important for mitosis, whereas tyrosine phosphorylation of the histone variant H2A.X plays an important role in the DNA damage response [5]. In order to understand the molecular mechanisms underlying these associations, it is essential to characterize the proteins and protein complexes that specifically interact with the epigenetic modifications. Several domains capable of binding selectively to a particular histone modification have recently been described. Examples include the chromodomain of heterochromatin protein 1 (HP1) which binds to H3K9me3 [6, 7] and the PHD finger of the chromatin-remodeling factor BPTF, which recognizes H3K4me3 [8].

### *Current State of the Art in Quantitative MS-Based Interactomics*

To identify interactions between proteins and histone PTMs, several approaches have been developed during the last couple of years. For example, candidate chromatin-"reading" domains can be expressed recombinantly and immobilized on arrays which are subsequently incubated with modified histone peptides, a method that was pioneered by the Bedford lab [9]. The disadvantage of this approach is that it is not unbiased but based on a selection of candidate domains that are then screened for putative interactions with modified histone peptides. To identify interactions with histone modifications in an unbiased manner, researchers typically make use of in vitro synthesized modified and non-modified histone peptides in pull-down experiments from crude nuclear or whole cell lysates. Following incubation and washes, proteins bound to the modified and non-modified immobilized histone peptide are then resolved using a SDS-PAGE gel and mass spectrometry is applied to identify the proteins in both samples [10]. However, identifying specific interactors in pull-downs from crude lysates is far from trivial, since these interactions are

usually masked by a large amount of high-abundant background proteins. In particular when making use of modern mass spectrometers, which are very sensitive and capable of sequencing thousands of peptides in complex samples in a matter of hours, researchers end up with a long list of identified proteins in the control and specific pull-down and it is often not immediately evident which are the PTM-specific binders. This approach therefore demands a quantitative filter that can be used to discriminate high abundant background proteins from specific interactors. In recent years, several methodologies have been developed that add a quantitative dimension to mass spectrometry measurements. Most of these methods rely on the introduction of stable isotopes in the proteins or peptides that are analyzed. This can be achieved through metabolic labeling during cell culture, the most popular method of which is called SILAC (stable isotope labeling by amino acids in cell culture) [11] or by chemical labeling at the protein or peptide level [12]. Differential labeling of proteins or peptides with "light" and "heavy" isotopes allows for a quantitative comparison of peptide and protein abundance between two experimental conditions. Prior to mass spec analysis, the light and heavy samples are combined. As a result, every peptide that is identified in the mass spectrometer has a light and a heavy peak and the ratio between these two peaks, which can be quantified using automated data analysis software, reveals the relative abundance of that peptide and the corresponding protein in the two different samples.

## *Identification of Histone PTM Readers*

In the context of PTM-dependent interactions, this quantitative filtering principle can also be applied. In this approach, in vitro synthesized peptides that are either unmodified or carry the PTM of interest are immobilized on a solid resin (Fig. 2). These peptides are separately incubated with light or heavy-labeled extracts. Following washes, beads from both pull-downs are combined and bound proteins are analyzed by LC-MS/MS. The quantitative abundance ratio of every peptide and corresponding protein in the mass spectrometer indicates whether this protein is a background protein (*H*/*L* ratio close to 1) or a specific reader of the PTM (*H*/*L* ratio significantly deviating from 1). This method was first applied to identify phosphotyrosine-dependent interactions in signal transduction pathways [13]. We adapted this approach to identify specific interactions with H3K4me3 using a SILAC-based histone peptide pull-down approach and discovered that the basal transcription factor TFIID binds to this mark with a high affinity [14]. This discovery is highly relevant given the genome-wide correlation between H3K4me3 and active promoters. The interaction is mediated via a PHD finger in the C terminus of the TAF3 protein and the Kd is ~0.16 μM, stronger than any of the other reported interactions with H3K4me3 [14]. Interestingly, the TFIID complex also contains a subunit (TAF1) that harbors a double bromodomain. Bromodomains specifically interact with acetylated lysines [15]. Given the general co-occurrence of H3K4me3 and acetylation of certain lysine residues such as lysine 9 and 14 of histone H3 on

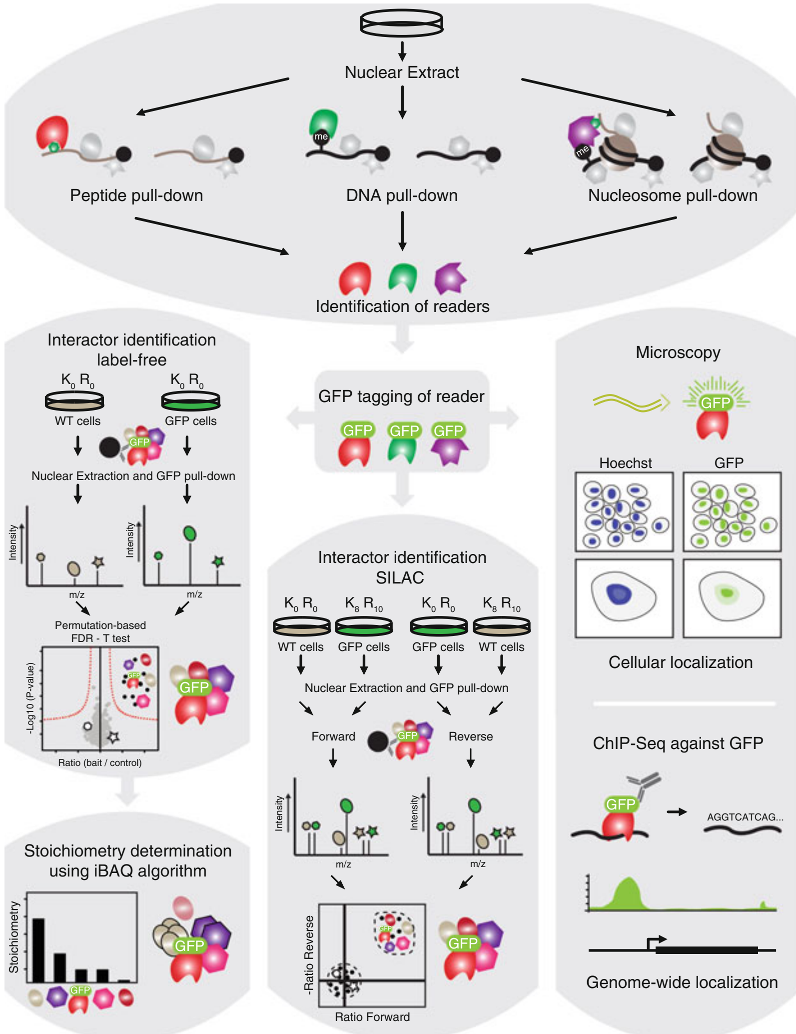



**Fig. 2** Integration of the multiple qMS-based interaction screens. Specific "readers" of (modified) histone tails, DNA strands, or nucleosomes can be identified using qMS, as described in the text. For further characterization, these readers are tagged with GFP using BAC TransgeneOmics. The interactors of GFP fusion proteins can be identified using two different workflows; a SILAC-based AP-qMS can be used, in which GFP and WT cells are differentially isotopically labeled and nuclear extracts obtained from these cells are applied to GFP-AP, mixed afterwards, and analyzed by LC-MS/MS. Alternatively, a label-free AP-qMS workflow can be applied, in which GFP and WT cells are both cultured in normal medium. Nuclear extracts from these cells are subjected to GFP-AP followed by LC-MS/MS. Furthermore, the stoichiometry of the "reader" complex can be determined using iBAQ-based calculations on label-free AP-qMS data. In addition to its use in mass spectrometry-based workflows, the GFP tag can also be used for cellular localization studies using microscopy and genome-wide localization studies using chromatin-immunoprecipitation followed by deep sequencing (ChIP-Seq)

active promoters, this implies a combinatorial agonistic recognition of these histone modifications by TFIID.

To investigate such potential cross talk between modifications occurring in close proximity on histone tails, we made use of a so-called triple pull-down approach. Cells can be SILAC labeled with two different stable isotope-containing versions of heavy lysine and arginine (lysine 4 and 8; arginine 6 and 10). Together with a third culture labeled with light amino acids, this allows incubating three different immobilized histone peptides with three differentially SILAC-labeled nuclear extracts. Every SILAC-labeled peptide in the mass spectrometer now appears as a triplet and the abundance of each of these three peaks indicates the relative affinity of that peptide and corresponding protein for each of the three baits. Using this approach, we were indeed able to show that H3K4me3 acts agonistically with H3K9 and H3K14 acetylation to anchor the TFIID complex on active promoters, which generally carry these modifications. Conversely, the triple pull-down approach was also used to show that another modification on the histone H3 tail, the asymmetric dimethylation of H3R2 (H3R2me2a), acts to prevent TFIID from binding to H3K4me3.

These initial encouraging observations led us to screen five major lysine trimethylation sites on histone H3 (H3K4me3, H3K9me3, H3K27me3, and H3K36me3) and H4 (H4K20me3) for novel readers [16]. In this study, in addition to performing the SILAC-based histone peptide pull-down as described above (called a “forward” pull-down), we also performed pull-downs using a SILAC label swap experiment in which the unmodified immobilized histone peptide is incubated with heavy SILAC-labeled extract whereas the modified histone peptide is incubated with the light extract. This is called a “reverse” pull-down. In this setup, PTM-specific readers that have a high ratio in the forward pull-down, for example a ratio of 10, will have a low ratio in the reverse experiment (0.1 ideally). Eventually all the identified and quantified proteins in the forward and reverse pull-downs are plotted against each other in a two-dimensional plot. Background proteins cluster around the origin of the figure, whereas the PTM-dependent interactors group together in one quadrant. Similarly, non-SILAC-labeled contaminants appear together in a single quadrant (low forward and low reverse ratio) and proteins for which binding to the peptide is abolished by the PTM also cluster (low forward ratio, high reverse ratio). Using this approach, we identified a large number of novel readers for each of the epigenetic trimethyl lysines on histone H3 and H4. For example, we discovered that the human SAGA complex, which is a major transcriptional coactivator complex, binds to H3K4me3 via a tudor domain in its subunit Sgf29. H3K9me3 is mainly read by HP1 isoforms, Polycomb proteins, and CDYL and CDYL2. Polycomb proteins also interact with H3K27me3, whereas proteins carrying a PWWP domain mainly recognize H3K36me3. Finally, the origin recognition complex interacts with H4K20me3. In fact, this protein complex reads all three repressive epigenetic modifications (H3K9me3, H3K27me3, and H4K20me3). In this study we further made use of the triple pull-down approach to show that the binding of TFIID, SAGA, PHF8, and BPTF to H3K4me3 is stimulated by H3K9 and H3K14 acetylation. Apparently, multiple proteins and multiprotein complexes have evolved to harbor a combination of domains that can specifically recognize these different epigenetic modifications that are commonly present on promoters of genes that are actively

transcribed. The triple pull-down approach was also used to show that phosphorylation of H3S10 and H3S28 selectively inhibits the binding of proteins to H3K9me3 and H3K27me3, respectively. Thus, multiple micromolar affinity histone PTM interactions together with specific DNA interactions eventually result in a very high affinity interaction of chromatin readers to their target genes. Furthermore, these interactions can be "tweaked" by adding or removing a certain reader within the complex or by adding an inhibitory modification in close proximity to the trimethylated lysine residue.

### *Identification of DNA and Nucleosome Readers*

The approach described above is not restricted to modified peptides but can also be used to identify specific DNA interactions [17]. In this case synthetic biotinylated and immobilized oligonucleotides are used as baits in affinity pull-downs from nuclear extracts. This method can be used to identify proteins binding to a particular transcription factor-binding site or a single nucleotide polymorphism that may be linked to a certain disease or phenotype [18, 19]. In the context of epigenetics, this method can also be used to identify proteins that specifically recognize methylated or hydroxymethylated DNA [17, 20, 21, 22].

Finally, complete in vitro reconstituted modified nucleosomes can be used as affinity baits. Using this approach one can investigate histone PTM cross talk between modifications occurring on different core histones as well as studying the interplay between DNA and histone modifications [23, 24]. In the future this approach can also be used to investigate the combinatorial effects of transcription factor-binding sites on DNA and PTMs on core histones.

## Identification of Protein–Protein Interactions

The approaches described in the previous section can be used to identify readers of epigenetic histone and DNA modifications. However, these experiments do not reveal any information about the protein complexes that these readers assemble in. Most cellular proteins perform their function in protein complexes, consisting of multiple stable core subunits and transient, substoichiometric interactors. Chromatin-associated proteins are no exception to this rule. Well studied chromatin-associated complexes, such as the Polycomb repressive complexes (PRC1 and PRC2) [25], the nucleosome remodeling and deacetylase (NuRD) complex [26], and mixed lineage leukemia (MLL) complexes [27, 28], harbor multiple (dynamic) subunits with different chromatin-binding domains and/or enzymatic activities towards histones and DNA. To understand the biological function of identified chromatin "readers," it is therefore crucial to identify their protein–protein interactions (PPIs). These experiments also help to discriminate direct PTM-mediated interactions from indirect, PPI-mediated binding to an epigenetic modification.

## *Current State of the Art*

Mass spectrometry has recently become an important platform to comprehensively identify PPIs of proteins of interest. The earliest approaches relied on tandem affinity purification (TAP) of the bait protein and its interactors. Two introduced affinity tags were sequentially used for affinity purification under stringent washing conditions, leading to relatively pure protein complexes. These complexes were separated by SDS-PAGE gel and individual gel bands were cut out and identified by mass spectrometry [29, 30]. However, mass spectrometers became extremely sensitive in recent years, and TAP tag purifications typically result in the identification of hundreds of proteins using the current state of the art instrumentation and data analysis software, even if the complexes appear relatively pure on gel. To overcome this problem, novel quantitative approaches that have been introduced in the previous section were developed. In these approaches, affinity purified samples are compared to control samples in a quantitative manner and this enables separating background binders from specific interactors [31]. As a consequence, there is no need for extensive purification of protein complexes, allowing for single-step affinity purifications and less stringent washes. This facilitates the identification of transient, substoichiometric but potentially important interactors in addition to core complex subunits [31].

To obtain a quantitative dimension in the mass spectrometric analyses, the earliest studies introduced stable isotope labels. As described above, this can be done in a chemical manner, in which peptides or proteins are chemically labeled, or in a metabolic manner, in which cells are grown in the presence of stable isotope-labeled amino acids [32]. Both chemical and metabolic labeling strategies allow the mixing of samples prior to mass spec analysis, thereby enabling direct comparison of peptide abundance. The disadvantage of peptide labeling approaches is that this mixing occurs later in the workflow, which may induce more handling errors [32]. Recently, label-free quantification methods were developed in PPI screenings [33–37]. In these approaches, the obtained mass spectrometry intensities are used to compare protein levels in different mass spectrometry runs. However, label-free quantification relies on more complex computational analyses compared to isotopic labeling, in which ratios can be directly derived from the light and heavy peptide pair. Therefore, label-free approaches are often slightly less accurate compared to stable isotope labeling approaches, although in large high-throughput datasets, this lack of accuracy is compensated for by more robust statistics as a result of the large amount of samples.

During the last couple of years, multiple affinity purification methods combined with quantitative mass spectrometry (AP-qMS) have been developed. Several of these approaches make use of endogenous antibodies against proteins of interest. In QUantitative Immunoprecipitation Combined with Knockdown (QUICK), a control immunoprecipitation experiment is performed in a lysate in which the protein of interest is knocked down using RNA interference [38]. The protein that is knocked down and its interaction partners have a quantitative abundance ratio deviating from

the background population. Malovannaya and colleagues adopted endogenous AP-qMS using a large number of antibodies (>3,000 IPs) against nuclear proteins, thereby identifying over 10,000 proteins and even more PPIs to define the nuclear "complexome" [39]. In this approach, rather than using a control pull-down for each immunoprecipitation, the complete dataset serves as a control for each individual immunoprecipitation. Other AP-qMS methods rely on the introduction of tags into proteins, such as FLAG [35, 40] and GFP [41]. Importantly, workflows have been adapted such that PPIs can be detected in a gel-free manner and using a single LC-MS/MS run, thereby significantly increasing the throughput. A potential problem of tagging approaches is that these methods tend to result in an overexpression of the bait relative to the endogenous protein. This issue can be addressed by inducible expression systems that allow for a sophisticated fine tuning of expression levels [42]. An alternative is using BAC TransgeneOmics, in which GFP tagging of proteins in mammalian cells is accomplished at near endogenous level by bacterial artificial chromosomes [41]. We used this approach to identify PPIs for chromatin readers of interest, as will be discussed in detail next.

## *Assigning Chromatin Readers into Complexes*

The SILAC-based workflow of AP-qMS using GFP-tagged nuclear proteins is optimized and explained previously [33]. Briefly, cells expressing the fusion protein (GFP cells) and control cells (WT cells) without the fusion protein are differentially SILAC labeled (Fig. 2, Interactor identification—SILAC). Nuclear extracts obtained from these cells are applied to GFP-AP using GFP-nanotrap beads [43] and mixed afterwards followed by on-bead trypsin digestion and LC-MS/MS (forward experiment). A second experiment is performed in which the SILAC labels of the GFP and WT cells are swapped (reverse experiment). In both experiments, proteins that bind nonspecifically to the AP beads are found in similar levels in the WT and GFP cells. Peptides derived from these proteins will show a one to one (1:1) ratio in the mass spectrometer. In contrast, the GFP-tagged protein and its interactors will be more abundant in the GFP-AP compared to the WT pull-down and will have a ratio that is significantly deviating from the background population.

This approach was extensively used to characterize the protein complexes harboring chromatin readers. In our comprehensive analyses of readers for trimethylated lysines on histone H3 and H4, many new interactors were discovered. Quite a few of these readers were previously uncharacterized proteins, which were not known to be involved in chromatin structure and function. One of these proteins, LRWD1, was found to have affinity for three transcriptionally repressive histone trimethyl lysine marks (H3K9me3, H3K27me3, and H4K20me3). Interestingly, these marks are also bound by the origin recognition complex (ORC), suggesting that LRWD1 may be a novel interactor of the ORC complex. Indeed, GFP tagging and purification of ORC2 revealed that LRWD1 is a direct interactor of this complex [16, 23]. Independently LRWD1 was also identified by others as a novel ORC

complex subunit called ORCA [44]. The same approach was used to identify two novel subunits of the human BPTF/NuRF complex. We also identified a large number of novel HP1 interactors, many of which carry zinc fingers such as POGZ and Znf828. These proteins may serve to recruit HP1 isoforms to target sites in the genome.

The GFP-tag-based purification can also be applied in a label-free approach [45–47]. In this method, GFP and WT cells are both cultured in normal medium after which nuclear extract preparation and GFP-AP MS are done separately (Fig. 2, Interactor identification—label-free). After raw data analysis, the normalized intensity of a protein is compared between the GFP-AP and the control. To quantitatively distinguish PPIs from background proteins, an adapted *t*-test with a permutation-based false discovery rate (FDR) is performed using Perseus (MaxQuant software package [48]). A prerequisite of this *t*-test is that the specific and control pull-downs have to be performed at least in triplicate. We used this workflow to study PPIs of important chromatin-associated complexes. Purification of the PRC2 and MBD3/NuRD complex resulted in the identification of multiple novel interactors. For PRC2, these included two uncharacterized proteins, C17orf96 and C10orf12, which we hypothesize to affect PRC2 activity or play a role in recruiting PRC2 to target sites in the genome [43]. The NuRD complex interacts with multiple zinc finger proteins including ZMYND8 and Znf592 [43, 49]. The function of these proteins is not well understood, but they were recently identified as a major transcriptional coregulator complex [39]. These results therefore establish a new link between the NuRD complex and regulation of transcription.

## Label-Free Stoichiometry Determination of PPIs

All of the above-mentioned stable isotope labeling and label-free approaches provide high-confidence PPI data. These experiments, however, do not reveal any information about the stoichiometry of the detected interactions. The identification of core subunits is, of course, of major importance in order to determine the key affinities and enzymatic activities of the complex of interest. In order to obtain stoichiometry information, the relative abundance of interactors needs to be determined. In recent years, different strategies have been deployed to obtain (relative) quantification of the abundance of proteins in a particular sample [50–52]. Many of these strategies rely on a spike-in of isotope-labeled reference peptides of which the exact concentration is known. The use of these peptides is, however, expensive, not straightforward, and it is not suitable for high throughput screening. In a recent study, the abundance of proteins was estimated using an algorithm that normalizes the sum of peptide intensities of a protein for the theoretical number of its tryptic peptides, so-called intensity-based absolute quantification (iBAQ) [53]. Using this algorithm, it is now possible to combine label-free PPI identification with an estimation of the relative abundance of the protein of interest and its interactors [43, 44]. This facilitates the stoichiometry determination of protein complex subunits,

by scaling one of the interactors, for example, the GFP tagged protein to 1 (Fig. 2—Stoichiometry determination). In a recent study, we applied this stoichiometry determination methodology to the PRC2 and NuRD complexes [43]. To determine the MBD3/NuRD complex stoichiometry, we tagged MBD3 in HeLa cells and applied AP-qMS on nuclear extracts. The stoichiometry determination exposed a core complex of 1× CHD3/4, HDAC1/2 and MBD3, 2× GATAD2A/B and DOC1, 3× MTA1/2/3 and 6× Rbbp4/7. The newly identified zinc finger proteins Zmynd8 and Znf592 were found to be substoichiometric, which can be expected of interactors that serve to recruit the core complex to specific target genes. Since the structure of the NuRD complex is still unknown, this information is of high value and it might pave the path to recombinantly reconstitute and model this complex. For PRC2, we tagged EED with GFP in HeLa cells and performed AP-qMS. This revealed a core PRC2 complex, consisting of one molecule of EED, Suz12 and Ezh1/2. Recently published cryo electron microscopy data for PRC2 confirm these observations [54]. Other known interactors such as Rbbp4/7, PCL1/2/3, AEBP2, and Jarid2 bind substoichiometrically. The same holds true for the novel PRC2 interactors C17orf96 and C10orf12. This result implies a functional diversification of distinct PRC2 subcomplexes each containing different interactors as a stoichiometric component, which may serve to affect PRC2 complex activity or recruitment to target loci in a highly regulated spatiotemporal manner.

## Integrating MS-Based Interactomics Technology to Study Chromatin Structure and Epigenetics

To exemplify the different workflows described in this chapter and to emphasize the complementary nature of these experiments, we present an example in Fig. 3. In this case, interactors of methylated DNA were explored using immobilized DNA strands with or without CpG methylation. Nuclear extracts from light and heavy-labeled HeLa cells were separately incubated with the unmodified and CpG methylated DNA, respectively. Additionally, a reverse experiment was performed using a label-swap. Plotting the SILAC ratios of both experiments in a scatterplot revealed that the NuRD complex specifically binds to the methylated DNA, whereas CXXC domain-containing proteins and zinc finger proteins show a higher affinity for the non-methylated DNA (Fig. 3a). CXXC domain-containing proteins are known to have a preference for non-methylated CpGs [55]. To characterize the subunit composition of the NuRD complex, its core subunit DOC1, which was identified as a specific interactor in the methylated DNA pull-down, was tagged with GFP and its interactors were identified using label-free AP-qMS. Using the described permutation-based FDR *t*-test (FDR = 0.001 and s0 = 1), all known NuRD subunits are identified as interactors together with a zinc finger protein and CSB, which was recently discovered as a NuRD-associated factor at rRNA genes [56] (Fig. 3b). To distinguish core subunits from substoichiometric interactors and to characterize the stoichiometry of the complex, we performed the iBAQ-based calculations. After

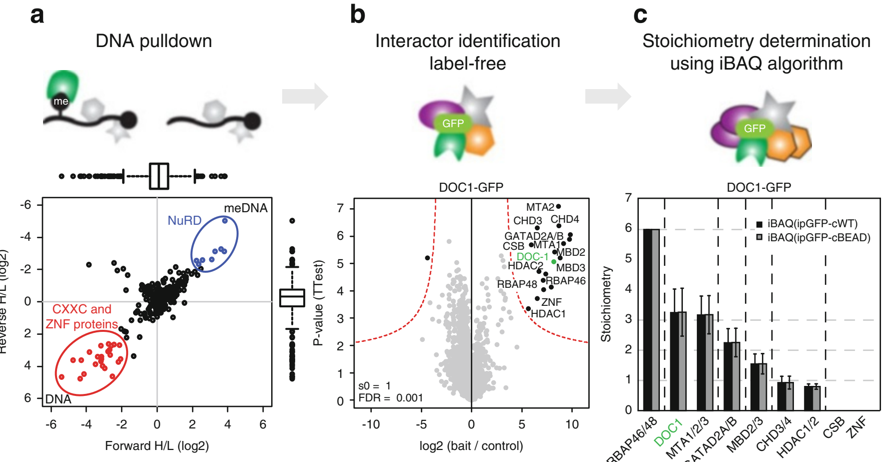



**Fig. 3** Identification of the NuRD complex as a methyl-CpG “reader” and characterization of the complex. (**a**) *Scatterplot* of a DNA pull-down using immobilized DNA strands containing non-methylated or methylated CpG dinucleotides. *Boxplot*-based outlier statistics reveal the NuRD complex as a methyl-CpG specific binder, whereas CXXC proteins preferentially bind to non-methylated DNA. (**b**) Further characterization of the NuRD complex, label-free GFP-based purification of NuRD subunit DOC1. The –Log(FDR) of the adapted *t*-test is plotted against the ratio of normalized intensity in GFP cells versus WT cells. This experiment clearly identifies all known NuRD complex subunits as specific DOC1 interactors. (**c**) The stoichiometry of the complex shown in (**b**) is determined using iBAQ-based calculations

scaling the relative abundance data to RBAP46/48, we obtain stoichiometries for the core subunits highly similar to previously published data [43] (Fig. 3c). Strikingly, both CSB and the zinc finger protein are highly substoichiometric, as one might expect from the fact that CSB, and probably the zinc finger protein as well, target the NuRD complex to only a subset of target genes [56].

## Summary and Outlook

In this chapter we have outlined recent developments in the field of quantitative mass spectrometry-based interactomics and we have illustrated how this technology can be used to answer important questions in the field of epigenetics, in particular to identify and characterize chromatin readers. In the future, such studies can be extended to virtually all organisms and cell types including embryonic stem cells or to cells arrested in a particular stage of the cell cycle to identify mitosis or S-phase-specific readers for epigenetic modifications of interest, for example. These quantitative approaches can be further developed to deduce the dissociation constants of detected chromatin reader–PTM interactions [57]. Furthermore, to complement these global interaction profiling experiments, techniques need to be developed that

can be used to characterize the proteome and epigenetic modification profile of particular genomic loci in a spatiotemporal and quantitative manner. Several technologies that can be used for this have recently been developed [58–61].

Due to the fact that epigenetic modifications are all reversible, the writers, readers, and erasers of these marks are attractive as potential drug targets. Therefore, identification and characterization of chromatin readers is not only important from a basic scientific interest but also from a clinical perspective. For example, bromodomain containing chromatin readers such as Brd2, Brd3, and Brd4 recently received a lot of attention, since small molecules that inhibit binding of these proteins to acetylated histones have been developed (JQ1 and I-BET) [62, 63]. These compounds have therapeutic potential and can be used for the treatment of MLL translocation-induced leukemia to inhibit aberrant expressed of proteins such as Myc in the tumor cells [64]. Such "epidrugs" may also be developed for other chromatin readers, including those interacting with trimethylated lysines on histone H3 and H4, to treat cancers that are characterized by aberrant lysine methylation patterns.

**Acknowledgments** We thank members of the Vermeulen lab for critical reading of the manuscript. Work in the lab is supported by grants from the Netherlands Organization for Scientific Research (NWO-VIDI), the Dutch Cancer Society (KWF), and the European Community's Seventh Framework Programme Project "4DCellFate."

## References

1. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
2. Campos EI, Reinberg D. Histones: annotating chromatin. Annu Rev Genet. 2009;43:559–99.
3. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol. 2007;14(11):1025–40.
4. Weake VM, Workman JL. Histone ubiquitination: triggering gene activity. Mol Cell. 2008;29(6):653–63.
5. Xiao A, Li H, Shechter D, Ahn SH, Fabrizio LA, Erdjument-Bromage H, et al. WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. Nature. 2009;457(7225):57–62.
6. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001;410(6824):120–4.
7. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001;410(6824):116–20.
8. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006;442(7098):91–5.
9. Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, et al. Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep. 2006;7(4):397–403.
10. Wysocka J. Identifying novel proteins recognizing histone modifications using peptide pull-down assay. Methods. 2006;40(4):339–43.

11. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1(5):376–86.
12. Eberl HC, Mann M, Vermeulen M. Quantitative proteomics for epigenetics. Chembiochem. 2011;12(2):224–34.
13. Schulze WX, Mann M. A novel proteomic screen for peptide-protein interactions. J Biol Chem. 2004;279(11):10756–64.
14. Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell. 2007;131(1):58–69.
15. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM. Structure and ligand of a histone acetyltransferase bromodomain. Nature. 1999;399(6735):491–6.
16. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell. 2010;142(6):967–80.
17. Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res. 2009;19(2):284–93.
18. Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet. 2012;8(9):e1002982.
19. Butter F, Kappei D, Buchholz F, Vermeulen M, Mann M. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO Rep. 2010;11(4):305–11.
20. Spruijt CG, Bartels SJ, Brinkman AB, Tjeertes JV, Poser I, Stunnenberg HG, et al. CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. Mol Biosyst. 2010;6(9):1700–6.
21. Bartels SJ, Spruijt CG, Brinkman AB, Jansen PW, Vermeulen M, Stunnenberg HG. A SILAC-based screen for Methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. PLoS One. 2011;6(10):e25884.
22. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013;152(5):1146–59.
23. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010;143(3):470–84.
24. Nikolov M, Stutzer A, Mosch K, Krasauskas A, Soeroes S, Stark H, et al. Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. Mol Cell Proteomics. 2011;10(11):M110.005371.
25. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011;469(7330):343–9.
26. Lai AY, Wade PA. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat Rev Cancer. 2011;11(8):588–96.
27. Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol. 2011;12(12):799–814.
28. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012;81:65–95.
29. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature. 2006;440(7084):637–43.
30. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084):631–6.
31. Vermeulen M, Hubner NC, Mann M. High confidence determination of specific protein–protein interactions using quantitative mass spectrometry. Curr Opin Biotechnol. 2008;19(4):331–7.
32. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005;1(5):252–62.

33. Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol. 2010;189(4):739–54.
34. Rinner O, Mueller LN, Hubalek M, Muller M, Gstaiger M, Aebersold R. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. Nat Biotechnol. 2007;25(3):345–52.
35. Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. Cell. 2009;138(2):389–403.
36. Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc Natl Acad Sci USA. 2008;105(5):1454–9.
37. Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods. 2011;8(1):70–3.
38. Selbach M, Mann M. Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). Nat Methods. 2006;3(12):981–3.
39. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, et al. Analysis of the human endogenous coregulator complexome. Cell. 2011;145(5):787–99.
40. Mak AB, Ni Z, Hewel JA, Chen GI, Zhong G, Karamboulas K, et al. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. Mol Cell Proteomics. 2010;9(5):811–23.
41. Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods. 2008;5(5):409–15.
42. Glatter T, Wepf A, Aebersold R, Gstaiger M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. Mol Syst Biol. 2009;5:237.
43. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics. 2008;7(2):282–9.
44. Shen Z, Sathyan KM, Geng Y, Zheng R, Chakraborty A, Freeman B, et al. A WD-repeat protein stabilizes ORC binding to chromatin. Mol Cell. 2010;40(1):99–111.
45. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. 2013;41(1):e28.
46. van Nuland R, Smits AH, Pallaki P, Jansen PW, Vermeulen M, Timmers HT. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. Mol Cell Biol. 2013;33(10):2067–77.
47. Hubner NC, Mann M. Extracting gene function from protein–protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods. 2011;53(4):453–9.
48. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–72.
49. Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M. A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol Cell. 2013;49(2):368–78.
50. Bennett EJ, Rush J, Gygi SP, Harper JW. Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. Cell. 2010;143(6):951–65.
51. Schmidt C, Lenz C, Grote M, Luhrmann R, Urlaub H. Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. Anal Chem. 2010;82(7):2784–96.
52. Wepf A, Glatter T, Schmidt A, Aebersold R, Gstaiger M. Quantitative interaction proteomics using mass spectrometry. Nat Methods. 2009;6(3):203–5.
53. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473(7347):337–42.

54. Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. Molecular architecture of human polycomb repressive complex 2. Elife. 2012;1:e00005.
55. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature. 2010;464(7291):1082–6.
56. Xie W, Ling T, Zhou Y, Feng W, Zhu Q, Stunnenberg HG, et al. The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. Proc Natl Acad Sci USA. 2012;109(21): 8161–6.
57. Sharma K, Weber C, Bairlein M, Greff Z, Keri G, Cox J, et al. Proteomics strategy for quantitative protein interaction profiling in cell extracts. Nat Methods. 2009;6(10):741–4.
58. Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep. 2012;2(1):198–205.
59. Fujita T, Fujii H. Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS One. 2011;6(10):e26109.
60. Lambert JP, Mitchell L, Rudner A, Baetz K, Figeys D. A novel proteomics approach for the discovery of chromatin-associated protein networks. Mol Cell Proteomics. 2009;8(4): 870–82.
61. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic loci. Cell. 2009;136(1):175–86.
62. Nicodeme E, Jeffrey KL, Schaefer U, Beinke S, Dewell S, Chung CW, et al. Suppression of inflammation by a synthetic histone mimic. Nature. 2010;468(7327):1119–23.
63. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, et al. Selective inhibition of BET bromodomains. Nature. 2010;468(7327):1067–73.
64. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12–27.

# Proteomic Interrogation of Human Chromatin Protein States

**Natarajan V. Bhanu and Benjamin A. Garcia**

**Abstract** Current epigenetic research encompasses studies on inheritance of stereotypic patterns of chromatin-associated histones and nonhistone proteins, their variant forms, modifications on the DNA as well as proteins and noncoding RNAs, relevant to nuclear-templated cellular events governing vertebrate gene function and differentiation. In recent years, the study of chromatin proteins and their interaction networks have evolved to grouping meaningful recurring marks into chromatin states. Major impetus in the study of chromatin states has been the outstanding improvements in analytical technology, including methodologies for efficient extraction of chromatin and its components, innovative preparatory chemistries, precision instrumentation, analytical layouts with better chemical discrimination, and retooled informatics capabilities. This chapter reviews some important recent developments in this growing field of interest and their potential for uncovering novel proteins and their modulatory roles in chromatin activity. We outline the global profiling of the most abundant of the chromatin proteins, the histones, and their modifications in several model organisms and cellular contexts, besides correlating particular histone marks and their turnover rates to specific chromatin states. A whole nuclear proteome description delineating an integrated workflow for protein preparation, interrogation of chromatin states, and data analysis is illustrated.

**Keywords** Chromatin • Proteomics • Histone • Mass spectrometry • Posttranslational modification (PTM) • Chromatin immunoprecipitation (ChIP)

---

N.V. Bhanu, Ph.D.
Epigenetics Program, Department of Biochemistry and Biophysics,
Perelman School of Medicine, University of Pennsylvania, 9-178 Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA

B.A. Garcia, Ph.D. (✉)
Epigenetics Program, Department of Biochemistry and Biophysics,
Perelman School of Medicine, University of Pennsylvania, 9-124 Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, PA 19104-6059, USA
e-mail: bgarci@mail.med.upenn.edu

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_8,
© Springer Science+Business Media New York 2014

# Spatial Chromatin States

## *Nucleosome Structure*

A diploid cell in the human body has about 2-m long DNA compacted in a nucleus that is 6 μm in diameter. This mind-boggling compression is made possible by coiling and supercoiling, starting with the wrapping of the negatively charged DNA around positively charged basic proteins called histones to form the *chromatin*. An octomeric ensemble consisting of a pair each of the core histones: H2A, H2B, H3, and H4 [1] winds 146 base pairs of DNA electrostatically (1.7 turns of DNA helix), and the next 20 base pairs are bound by the linker histone, the H1 protein, thereby spooling two full turns of DNA, into a structural and functional unit of chromatin called *nucleosome*. Histones have a carboxy-terminal tail, a globular domain, and a positively charged amino-terminal tail that is highly conserved as for the amino acid sequence [2].

## *Histone Variants*

All histones except H4 have been shown to have variations in the amino acid sequence at the carboxy terminal leading to *histone variants*. A comprehensive database of human histone proteins and histone-modifying enzymes, HIstome catalogs 55 histone variants, 106 distinct sites of modifications brought about by 152 modifying enzymes [3]. Histone variants seem to be correlated with particular chromatin states. H2A variants, viz., H2A.X, H2A.Z, H2AvD, macroH2A1, macroH2A2, and H2A. Bbd seem to act in a distinct pattern during gene activation or silencing ([4, 5]; reviewed by [6]). Similarly, H3 variants such as H3.1, H3.2, H3.3, and Cenp-A have been found to have unique functional associations. In addition to variants, different subtypes can be encodes by two genes as in macroH2A1 and macroH2A2 (abbreviated mH2A1 and mH2A2) encoded by *H2afy* and *H2afy2* [4]. Any number of alternative splicing of the histone gene can result in histone isoforms, such as seen in H2A isoforms, mH2A1.1 and mH2A1.2 encoded from the same gene, *H2afy* but differing in the nonhistone region [7]. In mammals, a number of these histone forms have evolved in a tissue-specific manner [8] and seem to carry out specialized functions [9, 10].

## *Histone Posttranslational Modifications*

Histones are modified at specific amino acids by multiple functional groups giving rise to *posttranslational modifications* (PTMs) such as lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, lysine ubiquitination,

sumoylation, etc. A recent study had extended the list to over 130 histone marks, comprising functional groups and modified amino acid sites, in a systematic analysis of PTMs on human histones [11]. Studies from our lab have confirmed that the turnover rates of PTMs vary depending upon the modification status and sequence variant [12, 13]. These PTMs modify the chromatin compaction by altering the charge on the amino acid, as seen with histone acetylation neutralizing the charge on lysine [14]; they also influence higher-order chromatin structure by organizing chromatin into distinct transcriptionally active and silent states [15], besides regulating chromatin remodeling, chromosomal condensation and segregation, mechanisms of replication, recombination, repair, senescence, and aging by modifying internucleosomal and histone–DNA contacts [16]. Further, histone PTMs act as docking sites for recruiting chromatin-associated proteins [17, 18] as well as allosterically regulate the chromatin complex at the binding domains (reviewed by [19]). Just as histone variants define distinct chromatin states (e.g. H3.3 with transcriptional activation and H3.2 with gene silencing), particular modifications of the same variant also differentially modulate chromatin activity (H3.1 acetylation with gene activation and H3.1 dimethylation to gene silencing) [20].

### *Histone Code*

The range of histone PTMs includes not only a variety of modifying groups but also degrees of modification of each group (mono, di, and tri) and together the specific patterns of amino-terminal histone modifications on a chromosome domain generate discrete combinatorial codes, often referred to as the *histone code* [2]. These differential forms anchoring chromatin-binding factors provide further functional diversity to each site and thereby enhance the informative potential of the genetic code. Histone PTM patterns seem to be unique across a broad variety of organisms [12] and even within the same organism, appear to have distinct kinetics at different physiological states [13]. Modifications, including acetylation and phosphorylation, are reversible and dynamic and are often associated with inducible expression of individual genes. Other modifications, such as methylation are found to be more stable, can be either activating or repressive and are involved in the long-term maintenance of the expression status of regions of the genome.

## Epigenetic Factors Influence Functional Chromatin States

Broadly, eukaryotic cells exist in two functional chromatin states during key cellular events such as proliferation, differentiation, apoptosis, death, and regeneration: *euchromatin* with open conformation allowing gene expression and the *heterochromatin* with condensation of DNA to inactivate genes. In humans, regions surrounding centromeres and telomeres are constitutively heterochromatic, whereas large

parts of the chromosome arms consist of transcriptionally competent *euchromatin*. Heterochromatin may be facultative as in X-inactivation in females, where the silent X-chromosome reverts back to euchromatin if transmitted to the male progeny. The chromatin states are modulated by proteins and enzymes that bind to the DNA. Current consensus is that transcriptionally silent CpG islands have hypoacetylated chromatin histone sites that are recognized by de novo DNA methyl transferases (DNMTs) that methylate the CpG regions and lock the gene promoter in a repressive state; proteins such as MeCP2 selectively bind to methylated DNA via a methyl-CpG sequence-binding domain (MBD) and recruit histone deacetylases (HDAC) to repress chromatin and prohibit transcription factor binding [21, 22]. However, the CpG islands within gene promoters are marked by monomethylated histone H3K4 that prevents de novo DNA methylation [23, 24] and help recruit transcription factors such as Sp1 and CTCF, which along with RNA polymerase II may block DNMT3a/b interaction with sites of transcriptional initiation [25–27]. DNMT1, itself a methylase has been shown to be directly methylated by the histone methylase, SET7/9 and demethylated by LSD1 [28]. Similar influences of histone distribution on the nature of DNA methylation have been endorsed in several large-scale studies of model organisms and humans, employing chromatin immunoprecipitation–microarray hybridization (ChIP-on-chip) analyses and expression reporter assays [29–32], highlighting the importance of deciphering global histone patterns for understanding chromatin dynamics.

## Interrogation of Chromatin States by Mass Spectrometry

### *Challenges in Proteomics*

The increased need for sensitive and accurate global identification of aberrant chromatin states and epigenetic determinants of disease is one of the many driving forces behind mass spectrometry (MS)-based proteomics. Most cancer cells manifest a global loss of H4K16ac, the active mark H3K4me3, and the repressive mark H4K20me3 and a gain in the repressive marks H3K9me and H3K27me3, due to the aberrant expression of histone acetyltransferases (HATs), histone deacteylases (HDACs), histone methyltransferases (HMTs), and histone demethylases (HDMs) [33]. In order to tackle the multifactorial molecular profiling of cells from complex biological samples, shotgun proteomics relying on mass spectrometry is performed for large-scale analyses. This is performed by digesting proteins into peptides, ionizing and sequencing them using tandem mass spectrometry followed by automated database searching. The success of this approach depends largely on two parameters: (1) the uniformity of peptide preparation across a variety of samples in different experiments to give reproducible results and (2) simplifying sample complexity so that tandem MS for important peptides are acquired. Compounding to these requirements, unique constraint in the identification of individual proteins within a

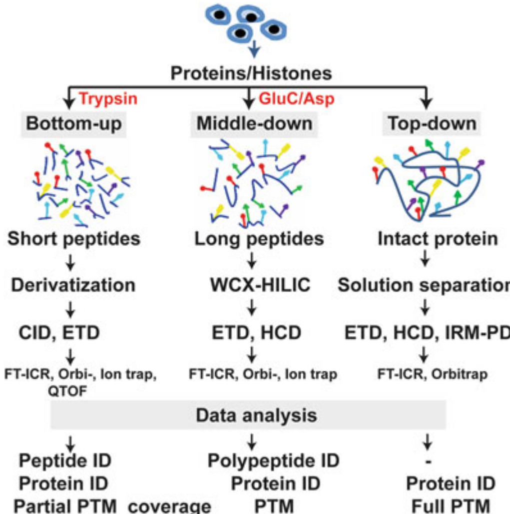



**Fig. 1** Workflow for quantitative proteomics using bottom-up, middle-down, and top-down MS analysis of posttranslational modifications. *WCX-HILIC* weak cation exchange-hydrophilic interaction liquid chromatography, *CID* collision-induced dissociation, *ETD* electron transfer dissociation, *HCD* higher energy collisional dissociation, *IRM-PD* infrared multiple photon dissociation, *FT-ICR* Fourier transform ion cyclotron resonance, *QTOF* quadrupole time of flight, *PTM* posttranslational modification

complex biological sample, limitations in dynamic range impeding detection of low-abundance proteins, and evaluation of all potentially useful information from the raw data have spurred a number of technical innovations.

## *Types of Mass Spectrometry*

Mass spectrometry-based approaches measure mass/charge ratio ($m/z$) of the ions fragmented from a peptide in helium or argon gas for mapping the chromatin proteins and their modifications. Electron capture dissociation (ECD) and electron transfer dissociation (ETD) are suitable for fragmenting ions of large peptides and proteins while preserving the PTMs stably during the ionization process, whereas collision-induced dissociation (CID) works best for short peptides (Fig. 1). Different PTMs result in unique mass shifts depending on the functional groups attached, such as acetylation on lysine residue having a mass shift of 42.011 Da over the unmodified peptide. However trimethylation (42.047 Da) may also have an identical mass shift and by using a high mass accuracy Orbitrap mass spectrometer, these masses differing by a mere 0.036 $m/z$ can be discriminated with reliable sensitivity (Fig. 2). Depending on the choice of gas phase fragmentation and the extent of the protein modification, a method that can both preserve and detect PTMs can be

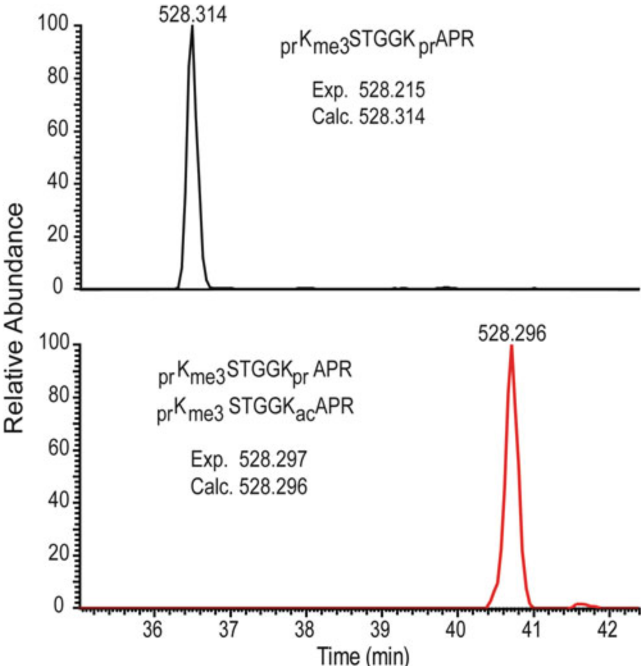



**Fig. 2** Accurate discrimination of the trimethyl and acetyl variants of histone H3 peptide (9–17 residues). (**a**) Mass spectrum shows H3K9me3 eluting at 36.4 min with *m*/*z* 528.314. (**b**) Mass spectrum shows coelution of H3K9ac and H3K14ac at 40.8 min *m*/*z* 528.296. The H3 peptides with analyte modifications in subscript (*pr* propionylation, *me3* trimethylation, *ac* acetylation) are shown *inset* in the chromatogram. Generally, H3K9me3 is associated with heterochromatin or silent euchromatin and H3K9ac has been correlated with transcriptional activation, while H3K14ac is critical for DNA damage checkpoint activation

adopted for characterizing chromatin proteins by "bottom-up," "middle-down," and "top-down" approaches (Fig. 1). Bottom-up proteomics relies on enzymatic protein digestions before liquid chromatography-coupled tandem MS analysis (LC-MS/MS), using trypsin for cleaving proteins highly specifically at arginine and lysine residues, so that peptides with an average size of 700–1,500 Da (typically ~8–25 residues long) and C-terminal arginine or lysine can be easily sequenced. On the other hand, a partial trypsinization or digestion with AspN preferentially cleaving proteins at N terminus of aspartic acid and cysteine or GluC cutting at the C terminus of glutamic and aspartic acid, ideally generating 2–20 kDa peptides is preferred for middle-down MS. Usually, a combination of enzymes or alternate enzymes is used to cleave peptides into sizes amenable for deep sequence coverage. Alternatively, top-down proteomics focuses on complete characterization of undigested proteins and is dependent on efficient separation of intact proteins for deducing combinatorial PTM information. The higher charge states of intact proteins in

the electrospray ionization (ESI) may often result in complex MS and MS/MS spectra and necessitates fragmentation by ETD and the use of high resolution tandem mass analyzers.

## *Sample Preparation for Mass Spectrometry*

In order to understand the chromatin states, it is imperative to develop rapid and reliable analytical methods that can be cost effectively performed routinely and in high throughput by exploiting naturally endowed features of some chromatin proteins. For example, histone purification in biological samples is greatly eased by their abundance in the nuclear compartment and by their basic nature that has allowed extraction in sulfuric acid. Phosphatase–deacetylase inhibitor cocktails and reducing agents preserve the PTMs on the peptides intact during isolation by mechanical disruption or nonionic detergents. Extraction of chromatin proteins in a relatively pure state is still not possible for the different species, cell types within the organism, and physiological states of the cell. In most cases, a compromise between biological specificity and broadness of characterization needs to be made. In a large-scale chromatin proteomic study by our group, three chromatin extraction procedures (total chromatin extraction using Triton X-100, salt extraction, and micrococcal nuclease digestion) were used to maximized the protein coverage; based on the protein quantity, MNase digestion fared better than both total and salt extraction, whereas the number of proteins identified were best with salt extraction [34]. This study also demonstrated the utility of simple area peak measurements of MNase-digested label-free cultured cells yielding higher specificity for chromatin proteins in eu- and heterochromatin fractions. However, only 25 % of the proteins (487 hits) were purified across all three methods and roughly only half of the proteins seemed exclusive to a single preparation, underlining the importance of multiple methods of sample preparation or the time-conserving isotope or metabolic-labeling strategy to enhance data retrieval in global proteomic analyses.

Affinity purification have been used to purify proteins of interest for MS and combined with pull-down protocols help to identify protein–protein interactors. These methods have to be used with caution as they may nonspecifically copurify contaminating proteins and inadvertently remove low abundance/low affinity specific partner proteins along with nonspecific proteins. While no one method meets all the requirements for interrogating chromatin states, the combination of several techniques have been proven to be useful. Immunoprecipiting Brd- and HP1-FLAG fusion proteins from asynchronously growing HEK (human embryonic kidney) cells, we performed tandem MS to map multiple and distinct PTMs on HP1a [35, 36]. Similar methods have been validated by other workers [37, 38]. Approaches such as affinity purification using engineered recombinase excision [39] or biochemical enrichment [40] though useful for single proteins and modifications, are not done at a whole-proteomic level and lack specificity of protein interactions.

## *Protein Quantification in Mass Spectrometry*

### Label-Free Quantification Format

The simplest label-free methods for the relative quantification of proteins in complex biological samples depend on the use of ion intensity, the number of unique peptides assigned to a given protein, and spectral counts as measures of relative abundance for individual proteins in a complex sample; subsequently, the masses of eluting cationic peptide precursors are measured in a MS scan. The most abundant precursors can also be selected in series for successive tandem MS events (MS/MS), called the *data-dependent acquisition* throughout the entire chromatographic separation and mass spectra of peptides generated can be then mapped to peptide or protein sequence databases such as the Sequest or Mascot. Relative expression levels (fold increase or decrease) can be deduced with algorithms such as MaxQuant, Proteome Discoverer, or Mascot Distiller. We routinely target for *m/z* of about 250–1,650 in full MS spectra and sequence seven most intense ions by CID for high-throughput proteome analysis. This straightforward MS setting had sensitively identified 1,912 unique proteins in chromatin-enriched fractions from HeLa cells [34]. Similar unlabeled middle-down approach identified over 150 PTMs on the histone H3.2-(1–50) peptide in asynchronously grown and butyrate-treated HeLa cells using off-line weak cation HILIC and infusion of HILIC fractions into an FTICR mass spectrometer with ECD fragmentation [41]. We found reproducible results employing both CID for short peptides (bottom-up and middle-down MS) and ETD for long protein (top-down) to confirm the paucity of histone PTMs in the most abundant of the chromatin proteins, HMGA1 [42].

### Quantification Through Chemical Derivatization with Stable Isotope Labeling

Some of the challenges in histone PTM detection by bottom-up MS such as sequence coverage are ensured by detecting peptides that are sufficiently small to bear modifications but long and hydrophobic enough to bind to the column. Presence of lysines and arginines too close to each other may prevent the digestion at one or the other residue, generating missed-cleavage events that are prone to poor reproducibility in MS. In addition, a difference in peptide sequence and modification state, unmodified and trimethylation for instance, may lead to differences in the ionization efficiencies of both peptides. Digestion of proteins into very small peptides result in poor chromatographic retention and this limitation has been remedied by chemical derivatization of the histones prior to proteolysis using acetic or propionic anhydride [43, 44]. We routinely derivatize the free ε-amino group in the N terminus and endogenously unmodified or monomethylated internal lysines, so that large and reproducible tail peptides containing histone modifications can be generated by trypsin digestion that only cleaves histones at the C-terminal to arginine residues

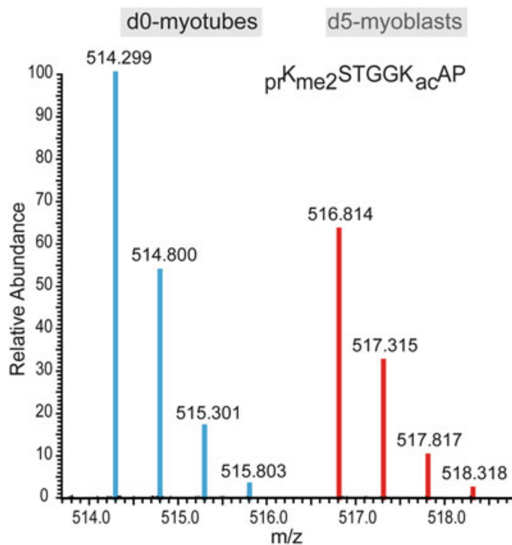



**Fig. 3** Chemical derivatization coupled to LC-MS/MS can relatively quantify proteins from two cell types. Full mass spectrum for the 2+ charge state of the modified 9–17 residue-long H3 peptide, $_{pr}K_{me2}STGGK_{ac}AP$, showing relative amounts in human myotubes (*blue*, derivatized with propionic anhydride, $d_0$) and myoblasts (*red*, derivatized with heavy-propionic anhydride, $d_5$) by quantitative bottom-up MS. *pr* propionylation, *me3* trimethylation, *ac* acetylation

generating slightly longer tryptic-like peptides ending in Arg residues that facilitate subsequent ionization, detection, and fragmentation by MS, with a mass shift of +56 Da. As a further innovation, we use a double derivatization method where a secondary derivatization using heavy $d_{10}$-propionic anhydride is performed after trypsin digestion on the second sample so that a pair of equivalent peptides derived from two samples that contain identical modifications will be detected as a doublet separated by 5 Da (the mass difference between $d_5$- and $d_0$-propionylated peptides) in MS spectra (Fig. 3). By comparing the intensity of individual peaks, the relative abundance of specific modifications in the two samples can be determined [36, 45].

## Metabolic Labeling for Characterizing Dynamic Changes in Chromatin Proteins

Label-free techniques based on protein spectral counts often have a limited dynamic range that is offset by metabolic/isotope labeling techniques that can provide more accurate quantification at low signal-to-noise and reduce errors introduced during sample preparation prior to mass spectrometry analysis. A metabolic labeling method, Stable Isotope Labeling by Amino acids in Cell culture (SILAC), which allows the incorporation of stable isotope amino acid residues into proteins has been greatly eased quantitative proteomic analysis. Matched cultures of cells are grown in identical media except that one medium contains a “light” and the other a “heavy” form of a selected amino acid (e.g., $^{12}C$- and $^{13}C$-labeled L-leucine, respectively). Metabolic incorporation of stable isotope amino acids results in pairs of chemically identical peptides that can be detected by MS. The relative abundance of the two proteins can be accurately determined by the ratio of MS peak intensities for such

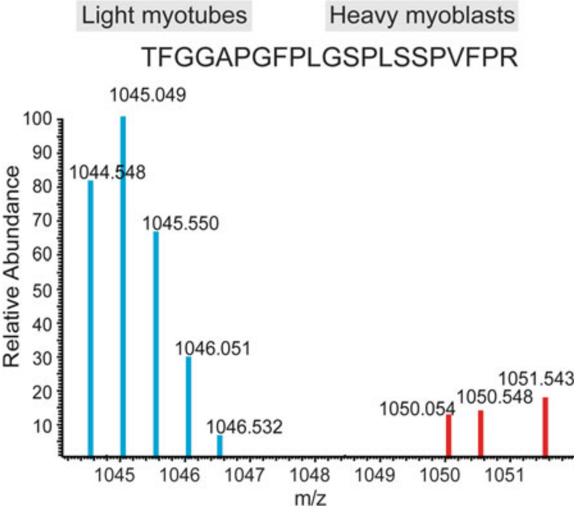



**Fig. 4** SILAC labeling enables comparative analysis of multiple samples. LC-MS of unlabeled myotubes (*blue*) and $^{13}C_6$,$^{15}N_4$-labeled myoblasts (*red*) shows differential expression of skeletal muscle marker, desmin (TFGGAPGFPLGSPLSSPVFPR) (2+)

peptide pairs (Fig. 4). Because of its high sequence coverage of identified proteins, high labeling efficiency, and simplicity of incorporation, this strategy has become one of the most popular methods for quantitative characterization of differentially expressed proteins and posttranslational modifications.

Pulsed SILAC method entails adding isotope-labeled "heavy" amino acids to the growth medium for only a short period of time. In heavy-methyl SILAC method, the cells were cultured in media with $^{13}CD_3$-methionine and can be converted to $^{13}CD_3$-adenosyl methionine, a sole biological methyl donor, thereby labeling all newly modified histones to heavy methyl group. The identification and relative quantification of histone methylation can be achieved by measuring relative peak intensities of methyl-modified peptide pairs. The advantages of this method include increased confidence in methylation site mapping as well as quantification and distinction between trimethylation and acetylation. Using isotope-labeled reference proteome, we have tracked the turnover of histone variants and PTMs in unsynchronized growing HeLa cells, pulsing in media containing exclusively$^{13}C_6{}^{15}N_2$-lysine, so that newly synthesized histones incorporated the isotopically "heavy" $^{13}C_6{}^{15}N_2$-lysine residues [46]. We found that in contrast to the highly dynamic acetylation, lysine methylation kinetics was observed to be much slower on the N-terminal tails of newly synthesized histones. In an earlier study, combining in vivo labeling with SILAC and top-down MS, Pesavento et al. [47] showed that almost all K20 methylations occurred progressively and targeted newly synthesized H4 histones during the cell cycle. Further expanding information on the temporal origins and methylation pathways for histone posttranslational modifications during cell cycle using

$^{12}CH_3$-methionine labeling, we found that H3K9me3 and H3K27me3 may be the "prompts" for histone synthesis and DNA replication in cycling cells [48].

SILAC-based technology has been employed for genome-wide and modest throughput analysis of chromatin proteins in multiple samples in a temporal fashion, such as for understanding the pathogenesis of yeast chromatin [49], cell cycle entry–exit in human cells [50], breast cancer-specific epigenetic signature [51], and many more. Migliori et al. [50] reported a previously unknown chromatin mark, H3R2me2s favoring euchromatin state during cell cycle withdrawal and differentiation in human cells. Using five different stable isotopic forms of arginine, a five-plex SILAC-based approach identified 144 nuclear proteins consistently upregulated during adipocyte differentiation [52]. Mann and colleagues demonstrated the highest ratio of modified to unmodified residues of any protein viz. 57 posttranslational modifications mapped to 36 residues in chromatographically enriched NUCKS in HeLa cells grown synchronously in SILAC medium [53]. SILAC-based quantitative MS in affinity purified HeLa S3nuclear extracts revealed the interactome of two critical chromatin marks, H3K4me3 (enriched with 32 proteins) and H3K9me3 (enriched with 40 proteins) [54]. In a recent study, Mann and coworkers reported the highest coverage, identifying 5,111 proteins in SILAC-labeled mouse embryonic stem cells and detected predominantly active H3K4me3, repressive H3K27me3, and bivalent histone marks (H3K4me3 together with H3K27me3). The power of SILAC technology resides in the number of amino acids that can be differentially labeled, which also imposes a limitation for some proteins ending up not being quantified.

## Isotope Labeling During Sample Preparation for Relative Quantification

Accurate quantification of proteins that are low abundance, hydrophobic, or highly charged can also be addressed by innovations in incorporation of mass tags during sample preparation (isobaric Tags for Relative and Absolute Quantification, iTRAQ; $^{18}O$ labeling; Isotope-coded affinity tag, ICAT; and Tandem Mass Tags, TMT). In these cases, relative quantification of peptides is based on ratios of reporter ions in the low *m/z* region of spectra produced by precursor ion fragmentation and comparison to its unlabeled counterpart, without additional analytical load. Using affinity purification, iTRAQ and mass spectrometry methods, Montes de Oca et al. [55] identified 11 candidates that constitute the proteome of an abundant nuclear protein, BAF in mediating DNA damage responses, genome replication, epigenetic control, and chromatin organization. Recently, Dephoure and Gygi [56] validated a rapid hyperplex approach with 2-day turnaround time and threefold increased protein coverage, using three metabolic labels (light, medium, and heavy labeling of lysine $^{13}C$, $^{15}N$, and D) and six isobaric tags (TMT) to profile the effect of rapamycin on yeast proteome and consistently identified >2,200 in four or more time points in three biological replicates. A recently developed variation of iTRAQ, mTRAQ (Applied Biosystems) performs absolute quantification based on the isotope-dilution mass spectrometry (IDMS) principle and is specially designed to take advantage of the MRM mode. The mTRAQ labels are nonisobaric and maximize possible

differences in the MRM transitions [57]. Even though iTRAQ can be 8-plexed, drawbacks including biases due to labeling efficiency and limited fractionation options at protein and subcellular levels restrict its use for nuclear proteomic interrogation.

Isotope-coded affinity tag (ICAT) labeling uses cleavable ICAT reagents consisting of a biotin tag, a cleavable linker, the isotope-coded tag with nine $^{12}C$ (light form) or nine $^{13}C$ atoms (heavy form), respectively, and a protein reactive group. The stable isotopes incorporate into the proteins postisolation by selective alkylation of cysteines. Following SCX, tryptic digestion and separation of cysteine-containing peptides in an avidin column, the complexity of the mixture of proteins was reduced. ICAT-based technology has been used for describing the chromatin state in rat fibroblasts [58], human B lymphocytes [59], and skeletal myocytes [60]. However, besides being expensive, time consuming, and tedious to perform, since the technique targets the low-abundance cysteine, PTMs on cysteine interfere with ICAT incorporation. ICAT also has low peptide coverage as its mass (442 Da) in the fragmented ion, interfering with reverse-phase retention times and MS/MS sequencing; also, derivatization at two sites may increase the molecular weight by 884 Da, thereby generating more complex MS/MS spectra. Another innovation called the Mass Coded Affinity Tagging (MCAT) quantifies protein by chemically converting C-terminal lysine into homoarginine (42 Da heavier). But the differential chemical reactivity of peptides to guanidination can be of critical concern in the acquisition of data and these tags have not been employed to study chromatin states as yet.

## *Advances in Workflow Instrumentation*

Technological advances to increase resolution of mass spectrometry by ionization techniques such as matrix-assisted laser desorption/ionization (MALDI) and ESI has been concurrent to evolution of nanoscale extractions, workflow refinement, mass-based multiplexing, and high throughput capabilities, thereby greatly enhancing the scope and potential of proteomic research in recent years. Shotgun proteomic interrogation of complex biological samples using liquid chromatography (LC)-electrospray ionization tandem MS (ESI-MS/MS) on 3D- or linear ion traps as well as quadrupole time of flight (TOF) instruments has now graduated to coupling multidimensional separation approaches such as offline strong cation exchange (SCX) chromatography, online SCX as in MudPIT strategy, by combining SDS-PAGE to LC-MS (GeLC-MS), and isoelectric focusing (IEF). In a large-scale proteomic analyses of chromatin and posttranslational modifications of chromatin protein, we coupled nanoliquid chromatography to LTQ-Orbitrap mass spectrometer for MS–MS to be performed on all the peptide components at the same time in parallel without the need to serially select each peptide precursor ion for fragmentation and identified over 1,900 unique proteins, besides global histone codes in eu- and heterochromatin fractions [34]. These novel proteomic analytical tools enable broad characterization, high throughput, improved dynamic range, high resolution,

and specificity of protein identification. However, in order to mature from being academic curiosities to reliable and efficient approaches, these methodologies need validation using simple, cost-effective, and user-friendly workflows.

### *Data Analysis Following Mass Spectrometry*

Early nuclear proteomic profiling of HeLa by MALDI TOF and nanoES-MS relied on peptide matching to human-expressed sequence tags [61]. In recent years, the availability of comprehensive proteomic/genomic databases as well as computational tools has augmented the protein identification tremendously. From the simplest correlation of chromatographic peak areas detected by MS to signal intensities to using data analysis pipelines that compare the experimental spectra to a selected protein database using "search engine" algorithms (e.g., Mascott, X!Tandem, Sequest, OSSMA), a plethora of free access software are available to a modestly bioinformatics-savvy researcher. Although no one software suite has been created to handle all data analysis steps, the desired analysis can be achieved by using combinations of multiple platforms. We recently developed the PILOT Protein for rapid computation of proteins and their PTMs in complex biological samples (HeLa cells) and evaluated with five algorithms (VEMS, SEQUEST, InsPecT, X!Tandem, and ProteinProspector), the spectral data acquired from five LC MS/MS instruments (QTOF, QStar, LTQ-FT Orbitrap, and hybrid Orbitrap/ion trap) of samples extracted by three procedures (total extraction, a salt extraction, or a micrococcal nuclease (MNase) digestion of Hela cells) [62]. Lastly, parsing data from different instruments, settings and search engines into publicly available databases is presently a work in progress in the mass spectrometry community.

## Antibody-Based Methods for Interrogating Chromatin States

While mass spectrometry is valuable for protein identification and quantification, the specific configurations of epigenetic mechanisms that determine chromatin states and the hierarchy of chromatin-proteins are examined at a molecular level using two strategies: (1) chromatin-centric approach for quantification of locus-specific proteins, their PTMs, and the associated DNA segments and (2) genome-centric approach for bound factors. Chromatin immunoprecipitation (ChIP) is currently used to map known molecular targets and histone modifications at the nucleosome level but has several drawbacks that limit its widespread utility. ChIP is heavily dependent on availability of epitope-specific antibodies and can characterize the genome-binding behavior of already known proteins and thus does not help to identify new, previously unknown proteins, nor does it help to reveal the identities of additional interacting proteins that are associated with particular genomic regions of interest. Further, ChIP can only effectively examine single histone PTMs or proteins and cannot simultaneously profile multiple targets, thereby being time

consuming and failing to shed light on the coincidence of particular histone PTMs, limitations shared by Western blotting and protein arrays as well. In recent years, some of these issues are being addressed by complementing immunoprecipitation, cDNA amplification, qPCR, and microarray with application of LC-MS for global quantitative proteome analysis.

Current immunoprecipitation methods for identifying protein complexes bound to chromatin aim to reduce cellular protein contamination while maximizing number of chromatin-associated proteins purified. We found about 15 % cytosolic, mitochondrial, ribosomal, and cytoskeletal proteins in a combined MS data set obtained after analyzing samples following three separate chromatin extraction procedures [34]. However some of these associations may be *bona fide* as in the case of $Ca^{2+}$/calmodulin interacting with H2AX complex during DNA repair and ionizing radiation-induced cell cycle arrest [63] and our observation of non-nuclear proteins like vimentin in chromatin preparations [34] may be explained by their emerging role in gene regulation [64] that warrants careful validation for their nuclear involvement. Another drawback is low abundance of epigenetic forms of proteins as we found in the case of Arg-modified peptides of HMGA1a, probably due to their expression in specialized nuclear events [42]. Thirdly, subtle biologically meaningful proteomic changes may be masked when the background association of a protein with chromatin is high as in the case of unaltered levels of Pol α-primase binding to replisome [65].

In the last few years, lack of availability of specific antibodies to pull-down all nuclear proteins had prompted ex vivo approaches for rapidly identifying and characterizing proteins that interact with the genome at locations of interest. It has been shown that sequence-specific DNA-binding proteins can be isolated using nucleic acid affinity capture [66] that "baited" DNA sequence outside of its endogenous context or an in vitro capture approach [67–70] and this strategy used along with SILAC [70] and can yield sufficient material for MS identification. A novel method called Proteomics of Isolated Chromatin segments (PICh) uses the endogenous DNA to retrieve the protein information. The DNA/formaldehyde-crosslinked chromatin regions were processed to release the proteins bound to those loci and identified using mass spectrometric analysis. This method is the reverse of ChIP, which uses protein antigens to capture the associated proteins. To determine low abundance proteins from protein–DNA complexes formed in vivo, Butala et al. [71] used a low copy number plasmid containing the sequence of interest and LacI to facilitate extraction. Dejardin and Kingston [40] used locked nucleic acid (LNA) probes to capture highly repetitive telomeric sequences to obtain sufficient bacterial protein for identification. With further ingenuity, Wu and colleagues subjected cross-linked DNA–protein segments to partial exonuclease digestion to generate single-stranded regions of the DNA and captured the complex on a solid support grafted with complementary oligonucleotides, an approach they termed as GENECAPP (Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics) [72].

Antibody-based protein arrays have been developed for phosphorylated proteins; while they have a utility for high-throughput screening, finer details such as the exact site of phosphorylation on each protein cannot be determined by this approach.

Other limitations of commercial histone antibodies include preferential recognition of epitopes multiply enriched for a given modification as against single modifications, failure to distinguish methyl-lysine states (mono-, di-, and trimethylation) and to recognize off-target PTMs that are altered by neighboring PTMs [73–76]. As conventional immunological methodologies have limited success in detecting multiple PTMs and sequence variants, it is important to develop site-specific histone antibodies with good discriminatory abilities for intended targets and validate the PTMs, DNA, and protein partners that define particular chromatin states.

From the perspective of identifying DNA sequences that are directly or indirectly bound to proteins of interest throughout the genome, chromatin immunoprecipitation with subsequent analysis on DNA arrays (ChIP-Chip) [77], or by DNA sequencing (ChIP-Seq) [29, 78, 79] are popularly used. ChIP-chip (Chromatin IP and microarray) that have been used to study histone modifications either detect the distribution of histone modifications using antibodies specially targeting these modifications [80] or locate the enzymes that catalyze the histone modification reactions [81]. Combining ChIP-chip with expression profiling has established the correlation of histone markers with transcription activity [80, 82]. In humans, CpG island array showed strong correlation between CpG methylation and histone modifications [83]. Similarly, cDNA array has provided new information on the distribution of histone methylation patterns in the coding regions of human genes [84] and the tiling array mapping of H3 markers (dimethyl-K4 and trimethyl-K4, acetyl-H3K9, and acetyl-H3K14) to nonrepetitive regions of human chromosomes 21 and 22 [29], enabling wider perspective about correlates of chromatin states.

Coupling ChIP and quantitative MS (ChIP-qMS), we estimated the relative abundance of histones in human mononucleosomes bound by the bromodomain-containing proteins Brd2, Brd3, and Brd4 and by the chromodomain-containing heterochromatin proteins HP1 β and HP1 α [36]. Similar approach had been reported using Agilent 1100 series nanoHPLC-Chip/MS system to identify 39 nuclear factors in *Toxoplasma gondii* that control the activity of bradyzoite-specific *ENO1* promoter; TgNF3was targeted for ChIP-Seq and subjected to qPCR to identify genes important in proliferation, virulence, differentiation, and cyst formation, through modulation of nucleosome assembly and/or disassembly [85]. Further, RNA sequencing has been used to elucidate the development-specific hepatic expression of 142 epigenetic modifiers, including enzymes involved in DNA de/methylation, histone de/acetylation, histone de/methylation, histone phosphorylation, and chromosome-remodeling factors in male C57BL/6 mice [86].

## Interrogation of Posttranslational Modification of Chromatin Proteins

Approaches targeting chromatin marks have identified a number of interacting proteins [87–90]. The differential enrichment of these DNA-binding proteins and their PTMs along with histone modifications may further increase the complexity of the

chromatin marks. Unlike the DNA that only gets methylated, chromatin-binding proteins bear all conceivable covalent modifications in combinatorial patterns reminiscent of the histones, albeit at low levels. The complex cross talk between PTMs of both histones and DNA-associated proteins is exemplified by ubiquitination of H2A at K119 that seems to be essential for PcG-mediated gene repression and is facilitated by the E3 ligase, a PRC1 protein ring finger protein 2, RNF2, which itself is a ubiquitin-conjugated protein and alludes to multiple regulatory feedback loops [91–93].

## *Heterochromatin Protein 1 (HP1)*

We studied the diversity of PTMs of two notable chromatin proteins in order to understand the translation of this plurality into combinatorial specificities, their mechanism of action and functional consequences in appropriate cellular responses. The heterochromatin protein 1, HP1 initially recognized to be important in heterochromatin formation, is part of the chromodomain protein superfamily that was identified originally in *Drosophila* to mediate position effect variegation [94]. It exists as three isoforms: α, β, and γ and contain a N-terminal chromodomain (CD) that binds to histone PTMs such as H3K9me3, as well as a C-terminal chromoshadow domain (CSD). During transcription, mammalian HP1γ and H3K9me3 seem to associate with the elongating RNA polymerase II (PolII), possibly involving the histone chaperone, FACT [95–98]. HP1 interacts with a number of other chromatin proteins such as kinase, KAP-1/Tif1β, histone deposition chaperone, CAF-1 complex, an ATP-dependent chromatin-remodeling complex, the BRG1 and the SET domain containing H3K9-specific HMTs, Suv39h1, SETDB1, and G9a/GLP [99–102]. Fodor et al. [103] have demonstrated the functional role of a protein hydroxylase at pericentric heterochromatin in mammalian cells by monitoring lysine trimethyl states of H3K9me3. We and others have reported PTMs on HP1, including acetylation, methylation, ubiquitination, and sumoylation using antibody-based methods [35, 104–106]. We have comprehensively characterized known as well as novel sites of phosphorylation, methylation, acetylation, and formylation by nanoflow liquid chromatography tandem MS, following immunoprecipitation of HP1 fusion proteins using anti-HP1-FLAG antibodies in asynchronously grown HEK (human embryonic kidney) cells [35]. The localization of these PTMs seem functionally strategic since phosphorylation sites were in the hinge region of the HP1 protein, modulating the protein binding at the origin recognition complex, while rest of the PTMs were detected in the chromo/chromoshadow domain alluding to interactions with histones. Our study distinguished several PTM marks to typify certain chromatin states and contexts such as phosphorylation at Ser-83 (Ser-93) as a euchromatic mark of HP1γ, Ser-95 for heterochromatin population, and Thr-51 phosphorylation on HP1β during DNA damage by casein kinase II. Further, HP1 methylation and acetylation were often at the same sites, possibly functioning as "transcriptional switches" similar to those involving acetylation/methylation in the case of H3K9ac (associated with gene activation), and H3K9me3 (associated

with transcriptional repression), thereby regulating chromatin states. Interestingly, we identified a newly recognized modification, the formylation in the chromodomains on HP1that bind particular histone marks as well as co-occur with methylation, acetylation, and phosphorylation on the same residues and may have erstwhile unknown roles in the regulation of chromatin states.

### *High Mobility Group A1 (HMGA1) Protein*

The second case in point is the widely studied chromatin-associated High Mobility Group A1 (HMGA1) protein enriched in both naked DNA and nucleosomes, especially in the vicinity of H3, H2A, and H2B and particularly seems to be expressed at high levels in testis and cancer cells in the adult, besides during embryonic development [107, 108]. HMGA family seems to have wide-ranging PTMs in 39 residues out of the 106 amino acids. Using bottom-up propionylation approach, Sgarra et al. [109] found relatively high levels of phosphorylated forms of the well-known C-terminal serine residues on HMGA1a compared to most modifications; a variety of kinases including CDC2, PKC, CK2, and HIPK2 phosphorylate the residues, especially at serine and threonine residues that are in close proximity to or within the DNA/chromatin binding motifs, possibly fine tuning their chromatin remodeling ability [109]. While bottom- up analysis is suitable for proteins rich in Lys/Arg generating short peptides, for those that have long stretches of Lys/Arg-poor regions, middle-down is the strategy of choice that characterizes the PTMs on peptides in higher charge states (>3+), besides giving a bird's eye view of the combinatorial patterns of PTMs and their relationships. Complementing bottom-up MS with middle-down analysis of HMGA1a in asynchronously grown HeLa S3 cells, we demonstrated all degrees of lysine methylation, lysine acetylation, arginine methylation, and phosphorylation of threonine and serine residues at varying abundance levels combinations and novel sites in HMGA1 [42]. We also found phosphorylation of Ser35 and Thr52 and additionally, we identified two novel but low abundance monomethylations on Lys64 and Lys70 of HMGA1a in these 1–23 amino acid long peptides, besides the previously reported acetylations. Performing relative quantification of all the PTMs, we found a general low abundance of histone PTMs, with di-phosphorylations and single mono-acetylations predominating among different peptides while di-/tri-modifications, multiple PTMs, and any sort of methylation were rare, leading us to speculate that HMGA1a may have unique tissue-restricted expression patterns in specific cellular and functional contexts.

### *Whole Nuclear Proteome Profiling*

Thirdly, in our large-scale proteomic cataloging of human chromatin, we identified 1,912 unique proteins, of which 45 % were nuclear [34]. Analyzing mono- and oligo nucleosomes, we characterized histone PTMs in the euchromatin or

heterochromatin-enriched samples using a stable isotope labeling quantitative MS method. We found a total of 274 unique proteins (out of 530) in the euchromatin and 103 (out of 384) in the heterochromatin fractions. Our quantitative proteomics analyses found that all acetylations including H3K9ac, H3K16ac, H3K18ac, H3K23ac, H4K8ac3, and H4K12ac3, besides methylations such as H4K16me2, H4K16me3, H4K20me1, H3K36me3, unmodified K27, and K36 (on H3.2) favor euchromatic state. This is not so straightforward as it appears, since both unmodified and modified forms can act as active marks if their neighbors (H3K18ac1K23un and H3K18unK23ac are euchromatic marks), are indicating that combinatorial codes are more important than any single modification alone. On the other hand, modifications such as H3K16ac, H3K12ac2, H3K16ac2, H3K36me2, H3K27me3, H3K9me3, (more in H3.1 and H3.2 compared to H3.3), H3K4me1 (more on H3.3 variant), H3K27me1, and H3K9me1 (more enriched on the H3.2 variant) seem to indicate heterochromatin state. Further, variant-specificity of PTMs seems to add another layer of complexity to the already bewildering epigenetic landscape. We also found histone H3.1 localizing to constitutive heterochromatin, histone H3.2 to facultative heterochromatin, and H3.3 to euchromatin, firmly supporting the H3 "barcode" hypothesis [110]. As we delve deeper into nuclear proteogenomics, it is apparent that gene activation or silencing may not be constituted by discrete on/off states in terms of histone PTM patterns and that combinations of histone PTMs may play a larger role in modulating transcriptional states than any single modification alone.

## Defining Chromatin States by Histone PTM-Based Interactome

Global interpretation of the histone code is meaningless unless combinatorial histone marks and their interactomes are fully described. Some chromatin marks such as acetylation are found to be pervasively activating, while the histone methylation seems to are either activating/repressing effect depending on the lysine they modify. Methylation of H3K4, H3K36, and H3K79, along with hyperacetylation of H3K9 were frequently encountered at transcription start sites (TSSs) and promoter regions of active genes while methylation of H3K9, H3K27 and H4K20 correlated with heterochromatin and transcriptional silencing [111–113]. Co-occurrences of chromatin marks such as the mutually antagonistic H3K4me3 and H3K27me3 as seen in embryonic stem cells and T cells seem to create "bivalent domains" that potentially fine-tune differentiation [113, 114]. Similarly, some repressive modifications such as H3K9 methylation may activate genes in a context-specific manner [98]. The interpretation of the chromatin marks seem to be very complex and it is evident that there are missing links in our understanding of the "histone language." To some degree, examination of chromatin marks in simpler model organisms and across species may provide insights about finer details of histone signatures [12], but these inventories need to be updated in the context of new knowledge, sites, and modifications. Lennartsson and Ekwall [115] summarized particular histone marks

predominating during various degrees of gene activation in human CD4+ T-cells: H4K20me1, H2AK9ac, H2BK5me, H3K79me1, H3K79me2, H3K79me3, H4K16ac mark strong gene expression, H4K16ac, and H3K36me3 maintain intermediate expression, while H2AZ, H3K27me, H3K9me1, H3K4me1, H3K4me2, and H3K4me3 are associated with low transcriptional activity. These patterns cannot be extrapolated to all human cell types, as the human embryonic stem cells seem to have H3K9ac, H3K14ac, and H3K4me3 marking high expression, while low levels of H3K4me3 correlated with subdued gene activity. It is not yet clear to what extent the "histone language" is exclusive and justifies the need for the exhaustive cataloging in all species, cell types, and functional stages.

Recently, chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) imparted an understanding of the chromatin states in large data sets. Systematically mining genome-wide data sets of 38 different histone methylation and acetylation marks, the histone variant H2AZ, RNA polymerase II (PolII) and CTCF at 200 base pair resolution in human CD4 T-cells, Ernst and Kellis [116] recognized 51 chromatin states after analyzing gene annotations, expression information, evolutionary conservation, regulatory motif instances, compositional biases, genome-wide association data, transcription-factor binding, DNaseI hypersensitivity, and nuclear lamina maps. The first group of 11 chromatin states (1–11) had high frequency of H3K4me3 in common, as well as significant enrichments for DNaseI hypersensitive sites, CpG islands, evolutionarily conserved motifs, and bound transcription factors differed however in the presence and levels of other associated marks, primarily H3K79me2/3, H4K20me1, H3K4me1/2, and H3K9me1, and of numerous acetylations leading to varying strength of the aforementioned functional enrichments, and varying expression levels of the DNA hypersensitivity sites, CpG islands, evolutionarily conserved motifs, and bound transcription factors; the second group of 17 chromatin states (12–28) was defined by combinations of seven marks, H3K79me3, H3K79me2, H3K79me1, H3K27me1, H2BK5me1, H4K20me1, and H3K36me3; the third class of 11 chromatin states (29–39) were associated with higher frequencies for H3K4me1, H2AZ, numerous acetylation marks, and/or CTCF and with lower frequencies for other methylation marks; the next group of states (40–45) marked large-scale repressed and heterochromatic regions, with H3K27me3 and H3K9me3; the final group of six states (46–51) showed strong and distinct enrichments for specific repetitive elements $(CA)_n$, $(TG)_n$ or $(CATG)_n$ (44, 45 and 302-fold, respectively), some with higher frequencies of H4K20me3 and H3K9me3 were heavily enriched for satellite repeat elements.

Computing recurring combination of a restricted panel of nine chromatin marks across nine cell types using the same approach, Ernst and coworkers [117] defined 15 chromatin states corresponding to repressed, poised, and active promoters, strong and weak enhancers, putative insulators, transcribed regions, and large-scale repressed and inactive domains. It is interesting to note that the definition of chromatin state is being expanded astronomically with the availability of high-resolution tools. The automated computational system that enabled this high-throughput pipeline, ChromHMM assigns chromatin state association based on ten criteria: expression level of downstream genes, transcription factor binding and motif

enrichments, spliced exon enrichments, elongating versus resting PolII enrichments relative to an IgG control, dinucleotide percentages, chromatin state enrichments for each chromosomal staining band for all human chromosomes, staining band genome-wide enrichments for each state, Gene Ontology (GO) enrichments for states with the most transcription start sites, histone deacetylase (HDAC) inhibition response enrichments and RepeatMasker class and family enrichments [116].

## Chromatin and Cancer

The epigenetic basis for tumorigenesis has espoused a new and active field of cancer epigenomics. Extensive reviews have summarized activating and inactivating mutations of some key histone modulators in malignancies implying their role in tumor initiation and suppression, respectively. Specific histone marks have been associated with cancer prognoses, clinical outcome, and relapse and have spawned epigenetic therapies based on HDAC/HAT and HMT/HDM/HAT and HMT/HDM. This promising novel alternative therapeutic approach of targeting transitory epigenetic changes compared to the stable genetic mutations for reversal of diverse cellular pathogenesis relies heavily on comprehensive understanding of the epigenetic consequence of genetic aberrations as key to medical intervention. A common diagnostic marker for cancer, hyperacetylation can result from a straightforward overexpression of HAT activity as in the case of steroid-receptor coactivator, amplified in breast cancer (A1B1) leading to hyperacetylation characterizing breast and ovarian cancer [118], or may result from complex interactions of aberrations as in the case of acute myeloid leukemia triggered by chromosomal translocations that creates a fusion CBP-MOZ chimera with protein-interacting domains for HAT, resulting in hyperacetylation [119]. HMTs and HDACs associate with transcription factors, tumor suppressors, and oncogenes to regulate cellular functions and their dysregulation alters the recruitment of complexes and alter transcriptional activity. Thus, while HDAC2 overexpression leads to familial-adenomatosis-polyposis-induced tumors [120], HDAC6 expression predicts better survival in breast cancer [121]. Translational research has elucidated common neoplastic denominators such as lysine methyl transferase, NSD2 implicated in the pathogenesis of the hematologic malignancy multiple myeloma (MM) [122], when knocked down in vitro resulted in disruption of H3K36me2 organization, initiating oncogenic programming in many cell types [123]. Unless phylogeny and functional pathways of these upcoming targets are well understood, the prospect for developing epigenetic anticancer drugs will be limited.

## Concluding Remarks and Perspectives

Unlike the cellular proteins, chromatin-associated regulatory factors controlling transcription and replication are generally expressed at very low levels and are difficult to extract from the nucleus, hindering their identification by mass

spectrometry. Label-free quantitation between biological samples, as well as samples with differential treatments or disease states to compare relative abundance of proteins between the samples while good for multiplexing any number of primary samples, are not sensitive to small changes. Method validation using structural analogues as internal standards may yield better assay performance results for quantitative bioanalytical LC/MS assays; but their identical chemistries may cover up assay problems with stability, recovery, and ion suppression.

As the repertoire of novel histone modifications and sites is being expanded and different technical approaches are integrated to understand the histone language, it is worthwhile to interrogate old data and revisit old questions to understand the scope of their combinatorial effects on chromatin states. By enriching phosphoproteins by IMAC, we discovered 19 sites of phosphorylation using tandem mass spectrometry on six H1 isoforms from asynchronously grown HeLa cells, with several of these being found on non-(S/T)P*XZ*-containing sequences [124]. The same strategy in conjunction with SILAC permitted identification of novel phosphorylation sites, besides establishing the hierarchy of phosphorylation [125, 126]. As ChIP is increasingly coupled into all analytical strategies, the unique problems that arise with using antibodies need to be addressed. Clearly, the rate of discovering novel modifications, new sites, and combinations by mass spectrometer has outpaced the availability of antibodies. Even if they were available, presence of different epitopes of modification-specific antibodies may introduce ambiguity in the results and may not in be agreement with mass spectrometric identification.

The concept of "histone code" that was discovered by proteomic efforts is emerging to encompass a more holistic understanding of the functional interplay of combinatorial histone marks called "histone language." With the biological interpretation of histone language using proteogenomic tools such as ChIP-Seq and ChIP-chip, widely held paradigms about epigenetic mechanisms, ambi/multivalence of chromatin marks, etc., need to be revisited. The recent spate of chromatin state descriptions using ChromHMM indicate that the number of chromatin states discerned depends on the chromatin marks and functional genomic characteristics included in the study [116, 117]. It is logical that the resolution of mapping increased due to multiple parameters in these studies. However, it is imperative to chart a universal classification for chromatin state with biological validation for defining characteristics lest a burgeoning class of nuances of these states confound this promising field of knowledge with incoherent data. Standardization of methods and availability of reagents such as the 38 antibodies for methylation and acetylation marks used by Ernst and Kellis [116] to describe 51 human T-cell chromatin states in a public repository will help reproducibility by other workers and comparison of new data sets. To parse the unabated information flow, integrated proteomic and PTM database should be created for discovering the relationships between genes and proteins with their modifications from the biomedical literature. The eventual promise of proteomics is understanding biology; while technology is the vehicle to convey us to this goal, it is the science that drives the process.

## References


1. Luger K, Mäder AW, Richmond RK, et al. Crystal structure of the nucleosome core particle at 2.8A resolution. Nature. 1997;389:251–60.
2. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293:1074–80.
3. Khare SP, Habib F, Sharma R, et al. HIstome—a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 2012;40:D337–42.
4. Chadwick BP, Willard HF. Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum Mol Genet. 2001;10:1101–13.
5. Tanabe M, Kouzmenko AP, Ito S, et al. Activation of facultatively silenced Drosophila loci associates with increased acetylation of histoneH2AvD. Genes Cells. 2008;13:1279–88.
6. Arnaudo AM, Molden RC, Garcia BA. Revealing histone variant induced changes via quantitative proteomics. Crit Rev Biochem Mol Biol. 2011;46:284–94.
7. Rasmussen TP, Huang T, Mastrangelo MA, et al. Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. Nucleic Acids Res. 1999;27:3685–9.
8. Garcia BA, Thomas CE, Kelleher NL, et al. Tissue-specific expression and post-translational modification of histone H3 variants. J Proteome Res. 2008;7:4225–36.
9. Kamakaka RT, Biggins S. Histone variants: deviants? Genes Dev. 2005;19:295–310.
10. Talbert PB, Henikoff S. Histone variants: ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol. 2010;11:264–75.
11. Tan M, Luo H, Lee S, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell. 2011;146:1016–28.
12. Garcia BA, Hake SB, Diaz RL, et al. Organismal differences in post-translational modifications in histones H3 and H4. J Biol Chem. 2007;282:7641–55.
13. Zee BM, Levin RS, Dimaggio PA. Global turnover of histone post-translational modifications and variants in human cells. Epigenetics Chromatin. 2010;3:22–33.
14. Shogren-Knaak M, Ishii H, Sun JM, et al. Histone H4K16 acetylation controls chromatin structure and protein interactions. Science. 2006;311:844–7.
15. Martin C, Zhang Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol. 2005;6:838–49.
16. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128:693–705.
17. Campos EI, Reinberg D. Histones: annotating chromatin. Annu Rev Genet. 2009; 43:559–99.
18. Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. Curr Opin Cell Biol. 2003; 15:172–83.
19. Rando OJ. Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr Opin Genet Dev. 2012;22:148–55.
20. Hake SB, Garcia BA, Duncan EM, et al. Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem. 2006;281:559–68.
21. Jones PL, Veenstra GJ, Wade PA, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet. 1998;19:187–91.
22. Nan X, Ng HH, Johnson CA, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 1998;393:386–9.
23. Okitsu CY, Hsieh CL. DNA methylation dictates histone H3K4 methylation. Mol Cell Biol. 2007;27:2746–57.
24. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.
25. Fan S, Fang F, Zhang X, et al. Putative zinc finger protein binding sites are over-represented in the boundaries of methylation-resistant CpG islands in the human genome. PLoS One. 2007;2:e1184.
26. Guccione EF, Martinato G, Finocchiaro L, et al. Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol. 2006;8:764–70.

27. Turker MS. The establishment and maintenance of DNA methylation patterns in mouse somatic cells. Semin Cancer Biol. 1999;9:329–37.
28. Pradhan S, Chin HG, Estève PO, et al. SET7/9 mediated methylation of non-histone proteins in mammalian cells. Epigenetics. 2009;4:383–7.
29. Bernstein BE, Kamal M, Lindblad-Toh K, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell. 2005;120:169–81.
30. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306:636–40.
31. Gerstein MB, Lu ZJ, van Nostrand EL, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science. 2010;330:1775–87.
32. Rakyan VK, Hildmann T, Novik KL, et al. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol. 2004;2:e405.
33. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28:1057–68.
34. Torrente MP, Zee BM, Young NL, et al. Proteomic interrogation of human chromatin. PLoS One. 2011;6:e24747.
35. LeRoy G, Weston JT, Zee BM. Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. Mol Cell Proteomics. 2009;8:2432–42.
36. Leroy G, Chepelev I, Dimaggio PA, et al. Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. Genome Biol. 2012;13:R68.
37. Byrum SD, Taverna SD, Tackett AJ. Quantitative analysis of histone exchange for transcriptionally active chromatin. J Clin Bioinforma. 2011;1:17–22.
38. Trinkle-Mulcahy L, Boulon S, Lam YW. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol. 2008;183:223–39.
39. Griesenbeck J, Boeger H, Strattan JS, et al. Affinity purification of specific chromatin segments from chromosomal loci in yeast. Mol Cell Biol. 2003;23:9275–82.
40. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic loci. Cell. 2009;136:175–86.
41. Garcia BA, Pesavento JJ, Mizzen CA, et al. Pervasive combinatorial modification of histone H3 in human cells. Nat Methods. 2007;4(6):487–9.
42. Young NL, Plazas-Mayorca MD, DiMaggio PA, et al. Collective mass spectrometry approaches reveal broad and combinatorial modification of high mobility group protein A1a. J Am Soc Mass Spectrom. 2010;21:960–70.
43. Garcia BA, Busby SA, Shabanowitz J, et al. Resetting the epigenetic histone code in the MRL-lpr/lpr mouse model of lupus by histone deacetylase inhibition. J Proteome Res. 2005; 4:2032–42.
44. Garcia BA, Mollah S, Ueberheide BM, et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc. 2007;2:933–8.
45. Plazas-Mayorca MD, Bloom JS, Zeissler U, et al. Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. Mol Biosyst. 2010;6:1719–29.
46. Zee BM, Levin RS, Xu B. In vivo residue-specific histone methylation dynamics. J Biol Chem. 2010;285:3341–50.
47. Pesavento JJ, Yang H, Kelleher NL, et al. Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol Cell Biol. 2008;28:468–86.
48. Zee BM, Britton LM, Wolle D, et al. Origins and formation of histone methylation across the human cell cycle. Mol Cell Biol. 2012;32:2503–14.
49. Schwertman P, Lagarou A, Dekkers DH, et al. UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. Nat Genet. 2012;44:598–602.
50. Migliori V, Müller J, Phalke S, et al. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. Nat Struct Mol Biol. 2012;19:136–44.
51. Cuomo A, Moretti S, Minucci S, et al. SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. Amino Acids. 2011;41:387–99.

52. Molina H, Yang Y, Ruch T, et al. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. J Proteome Res. 2009;8:48–58.
53. Wiśniewski JR, Zougman A, Krüger S, et al. Constitutive and dynamic phosphorylation and acetylation sites on NUCKS, a hypermodified nuclear protein, studied by quantitative proteomics. Proteins. 2008;73:710–8.
54. Nikolov M, Stützer A, Mosch K, et al. Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. Mol Cell Proteomics. 2011;10:M110.005371.
55. Montes de Oca R, Shoemaker CJ, Gucek M, et al. Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. PLoS One. 2009;4:e7050.
56. Dephoure N, Gygi SP. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. Sci Signal. 2012;5:rs2.
57. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004; 3:1154–69.
58. Shiio Y, Donohoe S, Yi EC. Quantitative proteomic analysis of Myc oncoprotein function. EMBO J. 2002;19:5088–96.
59. Shiio Y, Eisenman RN, Yi EC, et al. Quantitative proteomic analysis of chromatin-associated factors. J Am Soc Mass Spectrom. 2003;14:696–703.
60. Himeda CL, Ranish JA, Hauschka SD. Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. Mol Cell Biol. 2008;28:6521–35.
61. Andersen JS, Lyon CE, Fox AH, et al. Directed proteomic analysis of the human nucleolus. Curr Biol. 2002;12:1–11.
62. Baliban RC, Dimaggio PA, Plazas-Mayorca MD, et al. PILOT PROTEIN: identification of unmodified and modified proteins via high-resolution mass spectrometry and mixed-integer linear optimization. J Proteome Res. 2012;11:4615–29.
63. Du YC, Gu S, Zhou J, et al. The dynamic alterations of H2AX complex during DNA repair detected by a proteomic approach reveal the critical roles of Ca(2+)/calmodulin in the ionizing radiation-induced cell cycle arrest. Mol Cell Proteomics. 2006;5:1033–44.
64. Ivaska J, Pallari HM, Nevo J, et al. Novel functions of vimentin in cell adhesion, migration, and signaling. Exp Cell Res. 2007;313:2050–62.
65. Kubota T, Hiraga S, Yamada K. Quantitative proteomic analysis of chromatin reveals that Ctf18 Acts in the DNA replication checkpoint. Mol Cell Proteomics. 2011;10:M110.005561.
66. Kadonaga JT, Tjian R. Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci USA. 1986;83:5889–93.
67. Griffin TJ, Aebersold R. Advances in proteome analysis by mass spectrometry. J Biol Chem. 2001;276:45497–500.
68. Himeda CL, Ranish JA, Angello JC, et al. Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. Mol Cell Biol. 2004; 24:2132–43.
69. Nordhoff E, Krogsdam AM, Jorgensen HF, et al. Rapid identification of DNA-binding proteins by mass spectrometry. Nat Biotechnol. 1999;17:884–8.
70. Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res. 2009;19:284–93.
71. Butala M, Busby SJ, Lee DJ. DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. Nucleic Acids Res. 2009;37:e37.
72. Wu CH, Chen S, Shortreed MR. Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. PLoS One. 2011;6:e26217.
73. Bock I, Dhayalan A, Kudithipudi S, et al. Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays. Epigenetics. 2011;6:256–63.
74. Fuchs SM, Krajewski K, Baker R, et al. Influence of combinatorial histone modifications on antibody and effector protein recognition. Curr Biol. 2011;21:53–8.

75. Rothbart SB, Krajewski K, Strahl BD, et al. Peptide microarrays to interrogate the "histone code". Methods Enzymol. 2012;512:107–35.
76. Rothbart SB, Lin S, Britton LM, et al. Poly-acetylated chromatin signatures are preferred epitopes for site-specific histone H4 acetyl antibodies. Sci Rep. 2012;2:489.
77. Iyer VR, Horak CE, Scafe CS, et al. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature. 2001;409:533–8.
78. Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316:1497–502.
79. Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. 2007;4:651–7.
80. Bernstein BE, Humphrey EL, Erlich RL, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci. 2002;99:8695–700.
81. Reid JL, Iyer VR, Brown PO, et al. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol Cell. 2000;6:1297–307.
82. Kurdistani SK, Tavazoie S, Grunstein M. Mapping global histone acetylation patterns to gene expression. Cell. 2004;117:721–33.
83. Kondo Y, Shen L, Yan PS, et al. Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. Proc Natl Acad Sci USA. 2004;101:7398–403.
84. Miao F, Natarajan R. Mapping global histone methylation patterns in the coding regions of human genes. Mol Cell Biol. 2005;25:4650–61.
85. Olguin-Lamas A, Madec E, Hovasse A, et al. A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence. PLoS Pathog. 2011;7:e1001328.
86. Lu H, Cui J, Gunewardena S, et al. Hepatic ontogeny and tissue distribution of mRNAs of epigenetic modifiers in mice using RNA-sequencing. Epigenetics. 2012;7:914–29.
87. Bua DJ, Kuo AJ, Cheung P, et al. Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. PLoS One. 2009;4:e6789.
88. Chan DW, Wang Y, Wu M. Unbiased proteomic screen for binding proteins to modified lysines on histone H3. Proteomics. 2009;9:2343–54.
89. Liu H, Galka M, Iberg A, et al. Systematic identification of methyllysine-driven interactions for histone and nonhistone targets. J Proteome Res. 2010;9:5827–36.
90. Vermeulen M, Eberl HC, Matarese F. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell. 2010;142:967–80.
91. Ben-Saadon R, Zaaroor D, Ziv T, et al. The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. Mol Cell. 2006;24:701–11.
92. Buchwald G, Stoop P, Weichenrieder O, et al. Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J. 2006;25:2465–74.
93. Wang H, Wang L, Erdjument-Bromage H, et al. Role of histone H2A ubiquitination in polycomb silencing. Nature. 2004;431:873–8.
94. Eissenberg JC, James TC, Foster-Hartnett DM, et al. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 1990;87:9923–7.
95. De Lucia F, Ni JQ, Vaillant C, et al. HP1 modulates the transcription of cell-cycle regulators in *Drosophila melanogaster*. Nucleic Acids Res. 2005;33:2852–8.
96. Kwon SH, Florens L, Swanson SK, et al. Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. Genes Dev. 2010;24:2133–45.
97. Lin CH, Li B, Swanson S, et al. Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Mol Cell. 2008;32:696–706.

98. Vakoc CR, Mandat SA, Olenchock BA, et al. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell. 2005;19:381–91.
99. Fritsch L, Robin P, Mathieu JR, et al. A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol Cell. 2010;37:46–56.
100. Murzina N, Verreault A, Laue E, et al. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Mol Cell. 1999;4:529–40.
101. Nielsen AL, Sanchez C, Ichinose H, et al. Selective interaction between the chromatin-remodeling factor BRG1 and the heterochromatin-associated protein HP1alpha. EMBO J. 2002;21:5797–806.
102. Pak DT, Pflumm M, Chesnokov I, et al. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell. 1997;91:311–23.
103. Fodor BD, Kubicek S, Yonezawa M. Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. Genes Dev. 2006;20:1557–62.
104. Ayoub N, Jeyasekharan AD, Bernal JA, et al. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. Nature. 2008;453:682–6.
105. Huang DW, Fanti L, Pak DT, et al. Distinct cytoplasmic and nuclear fractions of Drosophila heterochromatin protein 1: their phosphorylation levels and associations with origin recognition complex proteins. J Cell Biol. 1998;142:307–18.
106. Lomberk G, Bensi D, Fernandez-Zapico ME, et al. Evidence for the existence of an HP1-mediated subcode within the histone code. Nat Cell Biol. 2006;8:407–15.
107. Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer. 2007;7:899–910.
108. Shah SN, Resar LM. High mobility group A1 and cancer: potential biomarker and therapeutic target. Histol Histopathol. 2012;27:5675–9.
109. Sgarra R, Maurizio E, Zammitti S, et al. Macroscopic differences in HMGA oncoproteins post-translational modifications: C-ter phosphorylation of HMGA2 affects its DNA binding properties. J Proteome Res. 2009;8:2978–89.
110. McKittrick E, Gafken PR, Ahmad K, et al. Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci USA. 2004;101:1525–30.
111. Kim TH, Barrera LO, Zheng M, et al. A high-resolution map of active promoters in the human genome. Nature. 2005;436:876–80.
112. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by polycomb in human embryonic stem cells. Cell. 2006;125:301–13.
113. Roh TY, Cuddapah S, Cui K. The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci USA. 2006;103:15782–7.
114. Bernstein BE, Mikkelsen TS, Xie X, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125:315–26.
115. Lennartsson A, Ekwall K. Histone modification patterns and epigenetic codes. Biochim Biophys Acta. 2009;1790:863–8.
116. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010;28:817–25.
117. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43–9.
118. Anzick SL, Kononen J, Walker RL, et al. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science. 1997;277:965–8.
119. Borrow J, Stanton Jr VP, Andresen JM, et al. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. Nat Genet. 1996;14:33–41.
120. Zhu P, Martin E, Mengwasser J, et al. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. Cancer Cell. 2004;5:455–63.
121. Zhang Z, Yamashita H, Toyama T, et al. HDAC6 expression is correlated with better survival in breast cancer. Clin Cancer Res. 2004;10:6962–8.

122. Anderson KC, Carrasco RD. Pathogenesis of myeloma. Annu Rev Pathol. 2011;6:249–74.
123. Kuo AJ, Cheung P, Chen K, et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol Cell. 2011;44:609–20.
124. Garcia BA, Busby SA, Barber CM, et al. Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. J Proteome Res. 2004;3:1219–27.
125. Garcia BA, Joshi S, Thomas CE, Chitta RK, et al. Comprehensive phosphoprotein analysis of linker histone H1 from *Tetrahymena thermophila*. Mol Cell Proteomics. 2006;5:1593–609.
126. Garcia BA, Platt MD, Born TL, et al. Protein profile of osteoarthritic human articular cartilage using tandem mass spectrometry. Rapid Commun Mass Spectrom. 2006;20: 2999–3006.

# Quantitative Proteomics Characterization of Chromatin-Remodeling Complexes in Health and Disease

**Mahadevan Lakshminarasimhan and Michael P. Washburn**

**Abstract** Recent advances in the field of chromatin remodeling have elucidated its role in various cellular processes beyond transcription. The intricate dynamics and interplay between several chromatin-remodeling complexes has been shown to be responsible for events ranging from cell differentiation, epigenetic regulation, and human diseases. One of the biggest challenges in understanding the function of these large protein complexes is dissecting their assembly, interactions with other proteins, and identifying the role of individual components. Technological advances in quantitative proteomics make it one of the most sought after technique in identifying and analyzing multiprotein complexes and posttranslational modifications. In particular, multidimensional protein identification technology and spectral counting-based quantitative proteomic analysis is a popular choice for analyzing chromatin remodeling complexes. The reason being they are straightforward approaches and are able to identify and quantify low abundant proteins in a label-free manner. This chapter highlights the recent findings of chromatin-remodeling complexes with respect to cellular processes and disease states and the role of quantitative proteomics has played in these findings.

**Keywords** Chromatin remodeling • Cancer • Posttranslational modification • Mass spectrometry • Multidimensional protein identification technology • Spectral counting • Quantitative proteomics • Protein interaction networks

---

M. Lakshminarasimhan, Ph.D. • M.P. Washburn, Ph.D. (✉)
Stowers Institute for Medical Research, 1000 E. 50th Street,
Kansas City, MO 64110, USA
e-mail: MPW@stowers.org

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_9,
© Springer Science+Business Media New York 2014

## Abbreviations

| | |
|---|---|
| CTCL | Cutaneous T cell lymphoma |
| Da | Daltons |
| dNSAF | Distributed normalized spectral abundance factors |
| DSB | Double-strand break |
| ESI | Electron spray ionization |
| HATs | Histone acetyltransferases |
| HPLC | High-performance liquid chromatography |
| ICAT | Isotope coded affinity tag |
| iTRAQ | Isobaric tags for absolute and relative quantitation |
| LC | Liquid chromatography |
| MS | Mass spectrometry |
| MudPIT | Multidimensional protein identification technology |
| *m/z* | Mass to charge |
| NSAF | Normalized spectral abundance factor |
| PFL | Protein frequency library |
| PTMs | Posttranslational modifications |
| rDNA | Ribosomal DNA |
| RP | Reversed phase |
| SAHA | Suberoylanilide hydroxamic acid |
| SCX | Strong cation exchange |
| SILAC | Stable isotope labeling by amino acids in cell culture |
| SNAP | SILAC nucleosome affinity purifications |
| TAP | Tandem affinity purification |
| TMT | Tandem mass tags |

## Introduction

The eukaryotic genome is tightly packed into a stable entity called chromatin. The packing and access of the genome involves changes in the chromatin structure. This highly dynamic and coordinated effort is performed by chromatin-remodeling machineries which can be broadly classified into to two main types; the histone-modifying complexes and the ATP-dependent complexes belonging to the SWI/SNF family [1]. Histones, the basic repeating unit of chromatin, are part of the nucleosomal core on which the DNA is wrapped around, are heavily modified by posttranslational modifications (PTMs). Examples of histone modifications include acetylation, methylation, phosphorylation, sumoylation, and ubiquitination [2]. The most common histone-modifying complexes are histone acetyltransferases (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs), and histone demethylases (HDMTs) [3]. The discovery of PTMs on histones and their role in epigenetics and development lead to the “histone code” hypothesis, which states

that the modifications on histones play a role in gene regulation and passage of genetic information [4]. Histone modification by PTMs can influence chromatin structure directly or via chromatin-remodeling complexes [3].

The association and interplay between complexes involved in chromatin remodeling and histone modification are dynamic and often lead to alterations in chromatin structure and gene regulation. The interactions between these complexes are manifested in various cellular functions and disease states. It is also becoming clear that histone PTMs themselves influence each other and the cross talk between them is more significant than an individual PTM alone [5]. Recent advancements in this field suggests that site-specific modifications in histones could lead to several outcomes depending not just on this dynamics and interplay but also on the cellular context where the modifications take place, thus adding more complexity to the "histone code" leading to coining of terms such as "chromatin language" [5] and "histone language" [6]. Therefore, it is important to systematically characterize the associations and interactions between the chromatin-remodeling complexes and identify the individual components that mediate such interactions, especially in a context-dependent manner.

The complexity of the chromatin-remodeling system requires technologies that can handle and analyze complexity. Protein mass spectrometry and quantitative proteomics have had a major impact on the analysis of chromatin-remodeling complexes and their dynamics. Many approaches and technologies have been developed to generate novel insights into chromatin-remodeling systems. In this review, we will highlight the chromatin-remodeling complexes successfully studied using these approaches and we provide recent examples where quantitative proteomics technologies have led to novel insights into these systems.

## Multidimensional Protein Identification Technology

Affinity purification coupled with protein mass spectrometry has proved out to be a powerful technology to study protein–protein interactions, protein dynamics, and to identify novel multiprotein complexes (reviewed in [7]). As an example, multidimensional protein identification technology (MudPIT) is a powerful methodology that couples two-dimensional chromatography of peptides to tandem mass spectrometry resulting in the identification of 1,000s of proteins from a complex mixture [8]. The key advantage of MudPIT is its ease of use and the ability to analyze the whole cellular proteome and protein complexes using a shotgun proteomics approach (reviewed in [9]). Although MudPIT was initially established to analyze the whole proteome in a comprehensive manner [8], it is also being extensively used to discover novel protein complexes and has been a major contributor in identifying and analyzing chromatin remodeling complexes [10–12]. The procedures and protocols involved in performing a MudPIT run have been extensively described [13]. Briefly, proteins to be analyzed in a MudPIT run can either comprise whole cell lysates or multiprotein complexes purified using affinity purification such as

immunoprecipitation and or other established purification techniques. The proteins are then precipitated and digested using a common protease like trypsin to obtain peptides. The peptides are then loaded on to a triphasic microcapillary column that serves as an electrospray ionization (ESI) source. The column is packed with materials consisting of reversed phase (RP), strong cation exchange (SCX) followed by RP resin. The column is directly placed "in line" with an HPLC (high-performance liquid chromatography) coupled to a tandem mass spectrometer. The peptides first bind the RP phase for buffer exchange and then they are moved to the SCX by RP gradient [13]. This is followed by a salt bump to transfer some of the peptides to the second RP material, prior to a second RP gradient to elute the peptides directly into the mass spectrometer for fragmentation and analysis. Programs such as SEQUEST [14], DTASelect, and Contrast [15] are then used to interpret the resulting mass spectra and identify proteins present in the sample.

## *Quantitative Proteomics*

Due to the dynamic nature of chromatin-remodeling events, quantitative MS methodologies are required to determine the relative amounts of chromatin-remodeling complexes and histone PTMs. This is required for example to compare proteins between healthy and diseased cells [16], to understand the influence of a drug on the whole cell [17] or on a particular complex [18, 19], to identify the selectivity of a drug, or to understand the dynamics between two multiprotein complexes at a given point in a cell cycle. Several quantitative proteomics approaches have been used to characterize chromatin remodeling complexes as summarized in Table 1.

### Stable Isotope Labeling-Based Quantitative Proteomics

Quantification of proteins samples by MS was first developed based on incorporation of a label to a protein or its peptide before analysis. The method is based on the fact that a labeled peptide will have similar chemical properties to its unlabeled counterpart but differs in its mass. This would allow easy comparison of two different types of samples in one MS experiment. Several label-based approaches have been employed in recent years such as chemical labeling and metabolic labeling [26]. Chemical labeling approaches involve labeling of amino acids by affinity tags to quantify proteins. For example, in the ICAT approach, proteins from two different cells are compared by labeling an amino acid with an isotopic linker, where one sample contains the heavy isotope and the other the light isotope, and a tag [27]. The differences in peptide size arising due to linker are exploited to quantitate the proteins from two different samples [27]. Shiio and colleagues used the ICAT approach and identified changes in chromatin-remodeling factors upon Myc oncoprotein expression in human B cell lymphocytes [20]. The ICAT approach is useful to study proteins involved in cysteine-based redox processes but is not suited for studying

**Table 1** Overview of various MS-based quantitative approaches used in analyzing chromatin remodeling complexes and representative examples

| Quantitative approach | Advantages | Disadvantages | Examples |
|---|---|---|---|
| Chemical labeling | | | |
| ICAT, iTRAQ and TMT | Fast, any sample can be tagged at the protein level, simultaneous analysis of 6–8 samples is possible (iTRAQ and TMT approach) | Label introduced at the peptide level leading to high variation. Poor proteome coverage (only cysteines modified in ICAT approach). Chemical side products can obstruct identification of rare PTMs | Changes in chromatin-remodeling factors due to Myc oncoprotein expression [20]. Mapping interconnec tions among epigenetic modifications [21] |
| Metabolic labeling | | | |
| $N^{15}$, $C^{13}$, SILAC | Label is incorporated at the organism level leading to low variation and high accuracy | More expensive and does not apply to all samples. Isomer discrimination is difficult. Higher complexity during analysis due to mixing two or more samples | Identification of chromatin remodeling complexes binding to histone trimethylation marks [22]. Histone PTM comparative analysis between normal and cancerous cells of the breast tissue [23] |
| Label free | | | |
| Ion intensity, spectral counting | Low level of complexity, whole proteome analysis possible. Applicable to all types of samples. Comparison of multiple states and high dynamic range | Run to run variations and lack of internal standard leads to greater error in individual datasets | Regulation of RNA Pol I transcription by human Sirt7 (class III HDAC) through interactions with chromatin remodeling complexes [24]. Identification of proteins associated with H3K4me2, K3K9me2, H3K9ac PTMs [25] |

proteins with low cysteine content. In an isobaric-labeling approach, peptides are derivatized with various chemical groups that are isobaric (same mass), separated using liquid chromatography, and analyzed by tandem MS/MS to obtain peptides with different masses that depends on the reporter ions present on the tag. The reporter ions can then be compared to obtain quantitative information. Examples of isobaric labeling include the isobaric tags for absolute and relative quantitation (iTRAQ) approach where up to eight different samples can be analyzed simultaneously [28] and tandem mass tags (TMT) approach where up to six different samples can be analyzed simultaneously [29].

Metabolic labeling is a second stable isotope labeling-based approach. The principle of metabolic labeling procedure is similar to chemical labeling, except

that the label is introduced to the whole cell or organism via the growth medium [30, 31]. In metabolic labeling, the entire proteome is labeled and can be easily mixed with unlabeled samples [30, 31]. Metabolic labeling can be introduced either globally such as replacing all nitrogen atoms by $N^{15}$ or oxygen atom by $C^{13}$ or by a more targeted approach such as stable isotope labeling by amino acids in cell culture (SILAC), where amino acids such as arginine and lysine are replaced by their heavy counterparts [32]. When cleaving the SILAC-labeled proteins using protease such as Trypsin (cleaves at the C-terminal of arginine or lysine), at least one heavy amino acid is incorporated into the peptide, permitting relative quantification [32]. SILAC has emerged as a popular method for studying protein dynamics in particular chromatin-remodeling dynamics. For example; using SILAC method Vermeulen and colleagues were able to identify that human SAGA complex binds to H3K4me3 site via the double tudor domain of the SAGA subunit Sgf29 and hence established a link between human SAGA complex and H3K4me3 [22]. SILAC-based quantitative proteomics was also used in analyzing lysine acetylation and methylation patterns at fourteen distinct sites on histone H3 and H4 from breast cancer cell lines [23]. This study revealed that significant changes take place in these sites in cancer cells, which may represent a "breast cancer-specific epigenetic mark" and could prove useful in targeted therapeutic strategies [23].

### Label-Free Quantitative Proteomics

Label-free quantitative proteomics is an alternative approach that is growing rapidly in use. There are essentially two types of approaches used in label-free proteomics; the ion intensity method and the spectral counting method (reviewed in [33, 34]). In an LC-MS analysis, an ion with a given mass to charge ratio is detected and recorded at a particular time with a particular intensity. This enables determination of relative peptide levels between different samples directly from these peak intensities. Although, peptide quantification between two samples could be achieved by directly comparing each peptide ion obtained from LC-MS, there are certain practical limitations. For example, run to run variations resulting from sample preparation can result in different peptide intensities, making the quantification more challenging, although data normalization and statistical analysis can help in overcoming this issue (reviewed in [33, 34]). Since large amount of data are generated, sensitive algorithms and accurate mass spectrometers are required for automated peak alignment and comparison.

Spectral counting is another method that is used to carry out label-free quantitative proteomic analysis (reviewed in [33, 34]). In this approach, the total number of identified spectra from the same protein is compared between different datasets to obtain relative quantification (reviewed in [33, 34]). The approach takes advantage of the observation that, the more abundant a protein, more unique peptides result from its digestion, giving rise to more spectra [8, 35, 36]. In recent years, the spectral counting approach has been refined to account for differences in protein size since larger proteins usually contribute more distinct peptides and hence more

**Table 2** Representative examples of multiprotein complexes identified and characterized using the MudPIT and dNSAF approach

| Protein complexes | References |
|---|---|
| Identification of subunit composition of the mammalian mediator complex | [41] |
| The mammalian HAT complex TRRAP/TIP60 shares subunit YL1 with the SRCAP complex | [42] |
| Identification and characterization of yeast INO80 like chromatin remodeling complexes in mammals | [43–45] |
| Discovery of novel subunits in yeast SWI/SNF and SAGA complexes | [46] |
| In *Drosophila melanogaster*, H3K36 demethylation by KDM4A requires interaction with HP1a (heterochromatin protein 1a) | [47] |
| Regulation of chromatin structure by preventing histone exchange at coding regions by association of ISW1 and CHD1 with H3K36me3 | [48] |
| Organization and network architecture of the yeast SAGA and ADA complex | [10] |
| Analysis of RNA polymerase complexes | [40] |
| Identification of novel interaction partners of MLL-fusion proteins | [49] |
| Identification of proteins from the ELL (eleven-nineteen lysine-rich leukemia) complex of *Drosophila melanogaster* | [50] |

spectra than their smaller counterparts [37] and for taking into account peptides that are shared between multiple proteins [38]. This has led to the development of normalized spectral abundance factor (NSAF) [37] and distributed normalized spectral abundance factor (dNSAF) [38] approaches. The NSAF approach has been successfully used to estimate the relative levels of each subunit within a protein complex [39, 40]. Examples of the use of NSAF/dNSAF approaches for the analysis of protein complexes is given in Table 2.

## Chromatin Remodeling: Molecular Mechanism, Role in Cell Function and Diseases

### *Molecular Mechanisms Behind Chromatin Remodeling*

The ATPase subunit of the chromatin-remodeling complexes hydrolyses ATP to generate the energy needed to switch, slide, and or evict nucleosomes. The complexes employ a common basic mechanism to remodel chromatin; which is disrupting histone–DNA interactions. Nevertheless, the method of disruption and the effects of each complex on the position and stability of nucleosomes differ. To explain the mechanism of chromatin remodeling, the “loop capture” hypothesis has been put forward where a loop or bulge is created likely from the translocase activity of the complexes [51, 52]. The “loop capture” hypothesis is proposed based on a model where the ATPase subunit of the SWI/SNF binds to DNA at a specific location, loosens the histone-DNA contacts, and exerts its 3′ to 5′ translocase activity in conjunction with a DNA-binding domain to draw DNA from one end and pump it

to the other end in a unidirectional mode [51, 53]. Although, the binding of the SWI/SNF complex is required to disrupt histone–DNA interactions, hydrolysis of ATP is not required for this process [54]. Direct observation of DNA distortion by RSC (chromatin structure remodeling) complex [55] and conformational changes observed during the ISW2 ATPase cycle [56] further support the "loop capture" model. ISWI family of remodelers, which are smaller than the SWI/SNF complexes, seem to differ in their mode of binding and energetics. ISWI binds to nucleosomes using lesser number of contacts and their binding seems to require ATP, leading to conformational changes [57, 58]. Following ATP hydrolysis by the second ATPase domain of ISWI, template commitment occurs. The ISWI complexes bind nucleosomes as a dimeric motor allowing bidirectional movement in contrast to the SWI/SNF complexes. This mechanism of ISWI corroborates well with its function, such as establishing boundary elements within the chromatin, where the complex moves on both ways of the nucleosome [59, 60]. Despite similar mechanisms and domains, the remodeling complexes seem to differ in their function, which could be explained by their associated subunits.

## *Role of Chromatin Remodeling in Cellular Function*

Chromatin remodeling complexes carry out a plethora of biological functions which range beyond transcription regulation [61]. SWI/SNF family members are involved in both activation and repression of transcription, DSB (double-strand break) repair, cell signaling, spindle-assembly checkpoint, chromosome segregation, and cohesion [62–66]. The ISWI family members participate in assembly of chromatin, nucleosome spacing, embryonic development, and differentiation [67, 68]. The CHD family of proteins plays essential roles during cell development [69]. The INO80 complexes are involved in repression and activation of transcription [70], DNA repair and check point regulation [71], and replication fork progression and stabilization [72], whereas the SWR1 complexes are involved in heterochromatin establishment [73], chromosome segregation [74], and cell cycle progression [75]. Cross talk also exists between different types of chromatin-remodeling complexes to perform a particular function. For example, the transcription of yeast gene INO1 requires both INO80 and SNW/SNF to be present at the promoter site and presence of INO80 is required for the recruitment of SWI/SNF, indicating conditionality in the recruitment process [76].

Histone modifications by complexes such as HATs and HDACs not only play a major role in gene regulation but also carry out several cellular functions. For example, the human HATs, TIP60, and yeast Esa1 acetylate H4K12 leading to DNA repair and apoptosis [77, 78] whereas, Hat1 is involved in histone deposition [79]. Histone deacetylation of specific lysine residues have also been shown to affect cellular functions. HDAC1 and 2 together participate in DNA damage response by being recruited at the site of damage and deacetylating H3K56 and H4K16 leading to

DSB repair [80]. The histone methylase Dot1 is involved in methylating H3K79 leading to transcriptional elongation [81], normal cardiac function [82], and DSB repair [83]. Similar to HDACs, demethylases also participate in functions other than gene regulation. An example is the demethylation of H3K20me1 by the demethylase PHF8, which leads to cell cycle regulation [84].

### *Defects in Chromatin Remodeling Leads to Disease States*

Genes belonging to the SWI/SNF complexes have been identified as tumor suppressors, since mutations in this complex have been implicated in several types of cancer such as breast, prostate, lung, and Rhabdoid tumors (reviewed in [1]). The different subunits are associated with different types of cancer (reviewed in [1]). The NuRD (nucleosome remodeling and deacetylase) complex which is part of the CHD subfamily is also implicated in several cancers including breast cancer since its cell type-specific subunit MTA proteins been observed in metastatic carcinomas [85].

Histone-modifying complexes are also implicated in cancer. The human MLL1 gene encodes for a SET1 family H3K4 methyltransferase, which is part of a large complex called COMPASS-like and is involved in epigenetic regulation of Hox genes (reviewed in [86]). Chromosomal translocations in the MLL1 gene have been shown to cause aggressive types of leukemia in both infants and adults [87]. Similar to methylation, histone modifications involving acetylation have also been implicated in several diseases. Missense and truncating mutations in p300 leading to loss of its acetyltransferase activity has been shown to occur in epithelial cancers such as breast and prostate [88]. Human MOF, a HAT which specifically acetylates H4K16 has been shown to be reduced in medulloblastoma cells, both at mRNA and protein levels [89]. HDACs have also been implicated in several cancers and their role seems to be either due to activity or loss of it [90]. HDACs can deacetylate histones and downregulate gene expression either by translocating to certain regions of chromosome as fusion proteins or their expression itself is upregulated as found in gastric, colon, prostate, and breast cancer [90]. Loss of HDAC2 activity due to mutation was identified in colorectal tumors, indicating that HDACs might also function as tumor suppressor by downregulating oncogenes [90].

Mutations in chromatin-remodeling complexes are also implicated in diseases other than cancer. For example, the CHD4 gene which codes for the ATPase subunit of the NuRD complex has been found to have missense mutations in patients suffering from Schizophrenia, although no molecular basis has been assigned to it [91]. Similarly, mutations in the ATRX gene which encodes the SWI/SNF ATPase subunit has been shown to cause inherited mental retardation disorders such as alpha-thalassemia [92]. Interestingly, these mutations have been shown to cause aberrant changes in DNA methylation patterns at regions containing repetitive DNA and subtelomeric repeats [93], indicating cross talk between chromatin remodeling complexes and PTMs.

### *Interplay Between Chromatin-Remodeling Complexes*

The majority of histone-modifying proteins function as complexes. The complexes often associate and or share subunits between themselves or with chromatin remodelers. This enables them to be recruited together in a spatial and temporal manner to carry out cell functions. The subunits shared between complexes often times dictate the enzyme activity, substrate preference, nucleosome target, etc. [3]. As mentioned above, TIP60 is recruited at DSB, but it also requires the chromatin-remodeling complex NuA4 and its associated p400 ATPase activity to create an open conformation for efficient DSB repair [94]. The yeast Set1 protein, which is part of the COMPASS complex, is responsible for trimethylation of H3K4. Set1 is recruited to DSB by the RSC complex, again indicating cross talk between histone modification and chromatin-remodeling complexes [95]. A cross talk between phosphorylation and acetylation has also been established. During the DSB repair mechanism of *Drosophila melanogaster*, dTip60 is recruited to the DSB site by phosphorylated H2Av (γH2Av), which in turn acetylates γH2Av at K5 leading to exchange of unmodified H2Av by the Domino/dTip60 complex [96]. The ATAC HAT complex of *Drosophila melanogaster* contains two distinct HATs (Gcn5/KAT2 which preferentially acetylates H3K9 and H3K14 and Atac2/KAT14 which acetylates H4K16) and shares four subunits with SAGA [97]. ATAC coordinates interactions between upstream kinases and their target genes by functioning as a cofactor; another instance where cross talk takes place between remodeling events. ATAC complex is a cofactor for c-jun-dependent transcription at the JNK target gene site Jra and chickadee [98] and acts as a positive transcriptional coactivator. At the same time, under osmotic stress the ATAC complex colocalizes with c-jun, recruits upstream kinases such as JNK, and suppresses JNK-dependent gene transcription [98] indicating a dual role which depends on the context and interaction partners. As we learn more about chromatin remodeling, it becomes clear that a high degree of complexity is involved in this process and required systematic analysis. The development and advances in fields such as mass spectrometry and protein interaction networks play a greater role in such analyses.

## Quantitative Proteomics Analysis of Chromatin Remodeling Complexes

### *Protein Interaction Networks*

The identification of cross talk between multiprotein complexes involved in chromatin remodeling leads to questions such as how are these complexes organized and how do they interact with each other? Recent evidence shows that subunits shared between the complexes can mediate such cross talk [6, 99, 100]. One potential approach to begin to obtain a better understanding of the links between distinct

chromatin-remodeling complexes is to use affinity purification coupled to mass spectrometry-based protein interaction network technologies. Initially, protein interaction networks of protein complexes were large-scale studies attempting to cover an entire proteome [101, 102]. More recently, methods have been developed and applied to study smaller scale and focused interactomes [11]. The SILAC approach combined with computational methods has also been used to develop a strategy called the protein frequency library (PFL) [103]. The PFL method could be used to analyze transient and dynamic interactions between protein complexes, which are prevalent in chromatin-remodeling events. Specifically related to chromatin networks, approaches have been developed to analyze chromatin-associated proteomes by coupling chromatin immunoprecipitation to proteomics approaches [104–106]. These approaches may provide important insights into chromatin biology when they are used for comparative analyses.

## *Analysis of Multiprotein Complexes*

Several chromatin remodeling complexes, protein–protein interactions, and complex subunit organization have been identified and characterized using the affinity purification-MudPIT approach coupled with quantification using the NSAF method (Table 2). In this approach, the protein of interest, called the bait, is usually either tagged and integrated into the genome of the host or overexpressed using an expression vector. Commonly used tags are tandem affinity purification (TAP) tag [40], Halo tag [107], and FLAG tag [11]. Once purified, the bait protein can be analyzed using techniques such as MudPIT. The spectra obtained from MudPIT are processed and quantified to provide information about the proteins present in the complex and their relative abundance [39, 40]. Once interaction partners or novel subunits of a multiprotein complex are obtained, the partner proteins are tagged and purification followed by MudPIT analysis is performed to verify bona fide interactions [11].

As an example of a complex studies using the MudPIT and NSAF approaches, the yeast class I HDAC Rpd3 exists as two distinct complexes with varying molecular weights namely, the 1.2 MDa Rpd3L and 0.6 MDa Rpd3S complexes [108, 109]. Rpd3 regulates a wide range of genes by performing both transcription activation and repression. Rpd3 is recruited to the promoter either by interaction of its subunit Sin3 with sequence-specific transcription repressor factor Ume6 or by sequence-independent binding [110]. Carrozza and colleagues performed TAP tag purification followed by MudPIT analysis on the Rpd3 complexes Rpd3L and Rpd3S and identified novel subunits [108]. The MudPIT analysis also revealed subunits that are shared between the two complexes, as well as unique. Rpd3, Sin3, and Ume1 are part of a shared subunit core between the complexes, whereas Eaf3 and Rco1 subunits are specific for Rpd3S [108]. In additional studies, Sardiu et al. were able to generate a probabilistic deletion network for the yeast Rpd3 HDAC complex and predict interactions between their subunits [12]. This study determined the organization of proteins within the large and small Rpd3 complexes. Using an expanded approach, a deletion network analysis of the SAGA/ADA complexes determined that the

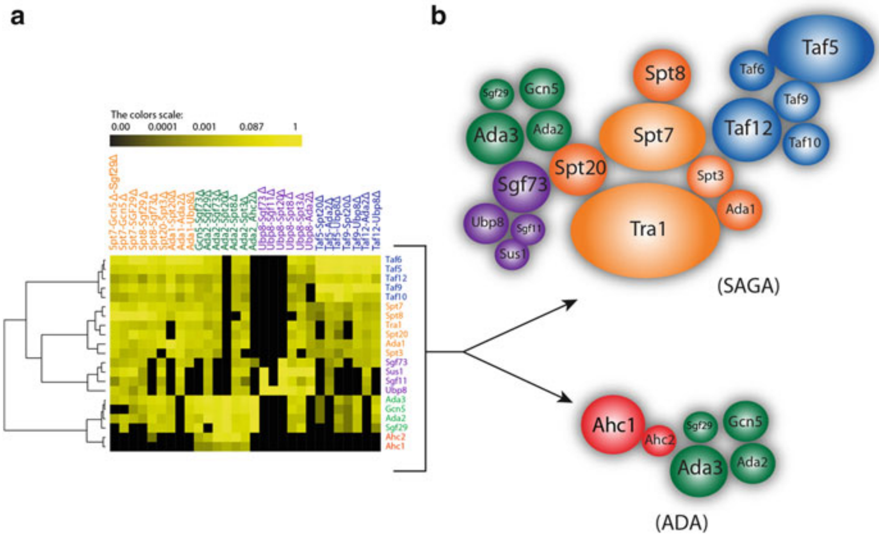



**Fig. 1** Deletion network analysis and complex reconstruction of the SAGA and ADA chromatin-remodeling complexes. (**a**) Each *column* represents an isolated TAP in a different deletion strain, and each *row* represents an individual protein (prey). The *color intensity* represents protein abundance (dNSAF) normalized on the subunits of the SAGA/ADA complexes with the *brightest yellow* indicating highest abundance and decreasing intensity indicating decreasing abundance. *Black* indicates that the protein was not detected in a particular purification. (**b**) Based upon all deletion purifications, all proteins of the SAGA/ADA complexes were organized into modularity and consequently a macromolecular model was assembled. The size of the *inset circle* correlates to the molecular weight of each illustrated protein. From [10]. Reprinted with permission from Nature Publishing Group

SAGA complex is composed of five distinct modules and the HAT core of the SAGA complex is also present in the ADA complex, which contains the distinct Ahc1 and Ahc2 proteins [10]. Figure 1 demonstrates the approach used in this study where AP-MS of TAP-tagged yeast proteins is carried out in *Saccharomyces cerevisiae* strains deleted for certain genes in the SAGA complex. It is particularly interesting that there are large and small subunits in the Rpd3 network [108, 109] and the SAGA/ADA network [10]. Whether or not this will be a widespread feature of chromatin-remodeling complexes remains to be seen.

In other quantitative proteomic analyses of chromatin-remodeling complexes, Hah et al. used affinity purification followed by SILAC-based proteomic approach and identified that BAF57, a subunit of the human SWI/SNF complex, is required for cell cycle progression and for maintenance of the SWI/SNF complex [111]. In addition, a SILAC-based proteomic approach called SILAC nucleosome affinity purifications (SNAP) was established to study cross talk between chromatin-modifying complexes [112]. In the SNAP approach, modified histone H3.1 was reconstituted with recombinant H2A, H2B, and H4 to generate modified octamer,

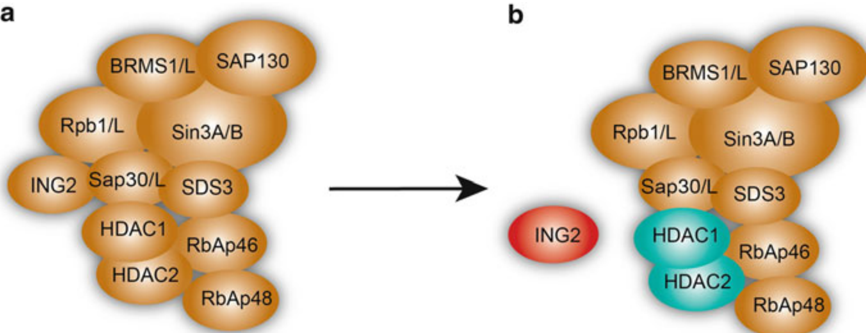



**Fig. 2** HDAC Inhibitor-Induced Dynamics of the Human Sin3 Complex. Protein complexes can change when subjected to a stimulus. The decrease of Ing2 with the Sin3 complex is shown as an example of the way quantitative proteomics can be used to study changes in protein complexes [18]. In this study, the Sin3 complex was analyzed by MudPIT and spectral counting in the absence (**a**) and presence (**b**) of the HDAC inhibitor SAHA [18]. Upon incubation with SAHA, the protein ING2 was shown to dissociate from the Sin3 complex (**b**) [18]. As an HDAC inhibitor, SAHA is likely binding to and inhibiting the enzymatic activity of the HDAC1 and HDAC2 proteins, but this study demonstrates that additional effects of inhibitors on multiprotein complexes can occur. From [115]. Reprinted with kind permission from Elsevier Limited

which was then mixed with DNA to generate modified nucleosomes [112]. These modified nucleosomes were used as baits to perform a SILAC experiment [112]. Using this approach, the authors identified several instances of cross talk between histone modifications and DNA methylation towards recruitment of proteins to chromatin [112].

## *Applications in Therapeutics*

Defects in chromatin remodeling have been implicated in cancer [1, 90]. Unlike genetic mutations, which are irreversible, chromatin-remodeling events such as histone modifications can be controlled due to their reversible nature. Moreover, chromatin-remodeling events often involve multiprotein complexes whose subunits may be modulated by small molecules in a specific manner, to yield desirable results. Since drug targets are predominantly proteins, many of which exist as complexes and are tightly networked with other proteins, MS-based proteomics is a valuable tool in the drug discovery process. HDACs in particular have generated great interest as cancer therapeutic targets [113, 114] and HDAC inhibitors have been studied using quantitative proteomics approaches [18, 33]. Smith et al. used MudPIT and the NSAF approach to show that suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor used in treatment cutaneous T cell lymphoma [114], induces conformational changes to the SIN3 complexes upon binding, resulting in dissociation of its subunit ING2 (Fig. 2) [18, 115]. This leads to disruption of the

HDAC complex and alters its chromatin targeting, indicating that small molecules can mediate their effects by more than one mechanism [18]. In an extensive study, Bantscheff et al. used affinity capture combined with a quantitative MS based chemoproteomics strategy to measure binding affinities of small molecules to large complexes in a native-like environment [19]. Using this approach, they discovered new targets for HDAC inhibitors, showed selectivity exists between the inhibitors and HDAC complexes, and also identified a novel HDAC complex called MiDAC, which assembles during mitosis [19].

## Conclusion

Our understanding of chromatin-remodeling events has significantly advanced in the past two decades. Nevertheless, there are new and exciting discoveries waiting to be made, since such events involve several multiprotein complexes and PTMs, many of which are still uncharacterized. Moreover, the dynamics and interplay between the complexes also present significant complexity and require further research. Improvements in MS-based proteomic methods have played a major role in answering several questions concerning chromatin remodeling. Thanks to the rapid advancement in both technology and method development in quantitative proteomics; it is now possible to analyze multiprotein complex samples and whole cell lysates in a quantitative and accurate manner. This provides important information about novel PTMs and the relative abundance of protein complexes involved in a particular remodeling event. MS-based proteomics combined with molecular biology techniques lead to systematic analysis of multiprotein complexes and protein–protein interaction networks. This provides a better understanding of the dynamics and spatiotemporal interactions that occur during chromatin remodeling events in both normal and disease states.

**Acknowledgments** The authors would like to thank the members of Washburn laboratory for critical reading of the manuscript and thoughtful insights. This work was supported by the Stowers Institute for Medical Research.

## References

1. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011;11(7):481–92.
2. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381–95.
3. Suganuma T, Workman JL. Signals and combinatorial functions of histone modifications. Annu Rev Biochem. 2011;80:473–99.
4. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293(5532):1074–80.
5. Berger SL. The complex language of chromatin regulation during transcription. Nature. 2007;447(7143):407–12.

6. Lee J-S, Smith E, Shilatifard A. The language of histone crosstalk. Cell. 2010;142(5): 682–5.
7. Gavin AC, Maeda K, Kuhner S. Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. Curr Opin Biotechnol. 2011;22(1):42–9.
8. Washburn MP, Wolters D, Yates 3rd JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19(3):242–7.
9. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. Chem Rev. 2007;107(8):3654–86.
10. Lee KK, Sardiu ME, Swanson SK, Gilmore JM, Torok M, Grant PA, et al. Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol Syst Biol. 2011;7:503.
11. Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc Natl Acad Sci USA. 2008;105(5):1454–9.
12. Sardiu ME, Gilmore JM, Carrozza MJ, Li B, Workman JL, Florens L, et al. Determining protein complex connectivity using a probabilistic deletion network derived from quantitative proteomics. PLoS One. 2009;4(10):e7310.
13. Swanson SK, Florens L, Washburn MP. Generation and analysis of multidimensional protein identification technology datasets. Methods Mol Biol. 2009;492:1–20.
14. Eng J, McCormack A, Yates J. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 1994;5(11):976–89.
15. Tabb DL, McDonald WH, Yates 3rd JR. DTASelect and contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res. 2002;1(1): 21–6.
16. Geiger T, Cox J, Mann M. Proteomic changes resulting from gene copy number variations in cancer cells. PLoS Genet. 2010;6(9):e1001090.
17. Fournier ML, Paulson A, Pavelka N, Mosley AL, Gaudenz K, Bradford WD, et al. Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. Mol Cell Proteomics. 2010;9(2):271–84.
18. Smith KT, Martin-Brown SA, Florens L, Washburn MP, Workman JL. Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. Chem Biol. 2010;17(1):65–74.
19. Bantscheff M, Hopf C, Savitski MM, Dittmann A, Grandi P, Michon AM, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. Nat Biotechnol. 2011;29(3):255–65.
20. Shiio Y, Eisenman RN, Yi EC, Donohoe S, Goodlett DR, Aebersold R. Quantitative proteomic analysis of chromatin-associated factors. J Am Soc Mass Spectrom. 2003;14(7): 696–703.
21. Xiong L, Darwanto A, Sharma S, Herring J, Hu S, Filippova M, et al. Mass spectrometric studies on epigenetic interaction networks in cell differentiation. J Biol Chem. 2011;286(15):13657–68.
22. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell. 2010;142(6):967–80.
23. Cuomo A, Moretti S, Minucci S, Bonaldi T. SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. Amino Acids. 2011;41(2): 387–99.
24. Tsai YC, Greco TM, Boonmee A, Miteva Y, Cristea IM. Functional proteomics establishes the interaction of SIRT7 with chromatin remodeling complexes and expands its role in regulation of RNA polymerase I transcription. Mol Cell Proteomics. 2012;11(2):M111.015156.
25. Chan DW, Wang Y, Wu M, Wong J, Qin J, Zhao Y. Unbiased proteomic screen for binding proteins to modified lysines on histone H3. Proteomics. 2009;9(9):2343–54.

26. Schulze WX, Usadel B. Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol. 2010;61(1):491–516.
27. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17(10):994–9.
28. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3(12):1154–69.
29. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003;75(8):1895–904.
30. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat Biotechnol. 2001;19(4):379–82.
31. Washburn MP, Ulaszek R, Deciu C, Schieltz DM, Yates 3rd JR. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. Anal Chem. 2002;74(7):1650–7.
32. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1(5):376–86.
33. Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem. 2012;404(4):939–65.
34. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics. 2011;11(4):535–53.
35. Liu H, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76(14):4193–201.
36. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. Anal Chem. 2005;77(19):6218–24.
37. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. J Proteome Res. 2006;5(9):2339–47.
38. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal Chem. 2010;82(6):2272–81.
39. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, et al. Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. Proc Natl Acad Sci USA. 2006;103(50):18928–33.
40. Mosley AL, Sardiu ME, Pattenden SG, Workman JL, Florens L, Washburn MP. Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes. Mol Cell Proteomics. 2011;2011:10(2).
41. Sato S, Tomomori-Sato C, Parmely TJ, Florens L, Zybailov B, Swanson SK, et al. A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol Cell. 2004;14(5):685–91.
42. Cai Y, Jin J, Florens L, Swanson SK, Kusch T, Li B, et al. The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. J Biol Chem. 2005;280(14):13665–70.
43. Jin J, Cai Y, Yao T, Gottschalk AJ, Florens L, Swanson SK, et al. A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. J Biol Chem. 2005;280(50):41207–12.
44. Cai Y, Jin J, Yao T, Gottschalk AJ, Swanson SK, Wu S, et al. YY1 functions with INO80 to activate transcription. Nat Struct Mol Biol. 2007;14(9):872–4.
45. Chen L, Cai Y, Jin J, Florens L, Swanson SK, Washburn MP, et al. Subunit organization of the human INO80 chromatin remodeling complex. J Biol Chem. 2011;286(13):11283–9.

46. Lee KK, Prochasson P, Florens L, Swanson SK, Washburn MP, Workman JL. Proteomic analysis of chromatin-modifying complexes in *Saccharomyces cerevisiae* identifies novel subunits. Biochem Soc Trans. 2004;32(Pt 6):899–903.
47. Lin C-H, Li B, Swanson S, Zhang Y, Florens L, Washburn MP, et al. Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Mol Cell. 2008;32(5):696–706.
48. Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, et al. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat Struct Mol Biol. 2012;19(9):884–92.
49. Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, et al. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol Cell. 2010;37(3):429–37.
50. Smith Edwin R, Lin C, Garrett Alexander S, Thornton J, Mohaghegh N, Hu D, et al. The little elongation complex regulates small nuclear RNA transcription. Mol Cell. 2011;44(6):954–65.
51. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annu Rev Biochem. 2009;78:273–304.
52. Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. 2011;21(3):396–420.
53. Saha A, Wittmeyer J, Cairns BR. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. Nat Struct Mol Biol. 2005;12(9):747–55.
54. Lorch Y, Maier-Davis B, Kornberg RD. Mechanism of chromatin remodeling. Proc Natl Acad Sci. 2010;107(8):3458–62.
55. Lia G, Praly E, Ferreira H, Stockdale C, Tse-Dinh YC, Dunlap D, et al. Direct observation of DNA distortion by the RSC Complex. Mol Cell. 2006;21(3):417–25.
56. Fitzgerald DJ, DeLuca C, Berger I, Gaillard H, Sigrist R, Schimmele K, et al. Reaction cycle of the yeast Isw2 chromatin remodeling complex. EMBO J. 2004;23(19):3836–43.
57. Gangaraju VK, Bartholomew B. Dependency of ISW1a chromatin remodeling on extranucleosomal DNA. Mol Cell Biol. 2007;27(8):3217–25.
58. Gangaraju VK, Prasad P, Srour A, Kagalwala MN, Bartholomew B. Conformational changes associated with template commitment in ATP-dependent chromatin remodeling by ISW2. Mol Cell. 2009;35(1):58–69.
59. Racki LR, Yang JG, Naber N, Partensky PD, Acevedo A, Purcell TJ, et al. The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. Nature. 2009;462(7276):1016–21.
60. Blosser TR, Yang JG, Stone MD, Narlikar GJ, Zhuang X. Dynamics of nucleosome remodelling by individual ACF complexes. Nature. 2009;462(7276):1022–7.
61. Saha A, Wittmeyer J, Cairns BR. Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol. 2006;7(6):437–47.
62. Sudarsanam P, Iyer VR, Brown PO, Winston F. Whole-genome expression analysis of snf/swi mutants of *Saccharomyces cerevisiae*. Proc Natl Acad Sci. 2000;97(7):3364–9.
63. Davie JK, Kane CM. Genetic interactions between TFIIS and the Swi-Snf chromatin-remodeling complex. Mol Cell Biol. 2000;20(16):5960–73.
64. Chai B, Huang J, Cairns BR, Laurent BC. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. Genes Dev. 2005;19(14):1656–61.
65. Angus-Hill ML, Schlichter A, Roberts D, Erdjument-Bromage H, Tempst P, Cairns BR. A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. Mol Cell. 2001;7(4):741–51.
66. Huang J, Laurent BC. A role for the RSC chromatin remodeler in regulating cohesion of sister chromatid arms. Cell Cycle. 2004;3(8):971–3.
67. Varga-Weisz PD, Wilm M, Bonte E, Dumas K, Mann M, Becker PB. Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature. 1997;388(6642):598–602.

68. Stopka T, Skoultchi AI. The ISWI ATPase Snf2h is required for early mouse development. Proc Natl Acad Sci. 2003;100(24):14097–102.
69. Unhavaithaya Y, Shin TH, Miliaras N, Lee J, Oyama T, Mello CC. MEP-1 and a homolog of the NURD complex component Mi-2 act together to maintain germline-soma distinctions in *C. elegans*. Cell. 2002;111(7):991–1002.
70. Shen X, Mizuguchi G, Hamiche A, Wu C. A chromatin remodelling complex involved in transcription and DNA processing. Nature. 2000;406(6795):541–4.
71. Morrison AJ, Highland J, Krogan NJ, Arbel-Eden A, Greenblatt JF, Haber JE, et al. INO80 and γ-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. Cell. 2004;119(6):767–75.
72. Shimada K, Oma Y, Schleker T, Kugou K, Ohta K, Harata M, et al. Ino80 chromatin remodeling complex promotes recovery of stalled replication forks. Curr Biol. 2008;18(8):566–75.
73. Rangasamy D, Berven L, Ridgway P, Tremethick DJ. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J. 2003;22(7): 1599–607.
74. Krogan NJ, Baetz K, Keogh M-C, Datta N, Sawa C, Kwok TCY, et al. Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. Proc Natl Acad Sci USA. 2004;101(37):13513–8.
75. Dhillon N, Oki M, Szyjka SJ, Aparicio OM, Kamakaka RT. H2A.Z functions to regulate progression through the cell cycle. Mol Cell Biol. 2006;26(2):489–501.
76. Ford J, Odeyale O, Shen C-H. Activator-dependent recruitment of SWI/SNF and INO80 during INO1 activation. Biochem Biophys Res Commun. 2008;373(4):602–6.
77. Ikura T, Ogryzko VV, Grigoriev M, Groisman R, Wang J, Horikoshi M, et al. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell. 2000;102(4): 463–73.
78. Bird AW, Yu DY, Pray-Grant MG, Qiu Q, Harmon KE, Megee PC, et al. Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. Nature. 2002;419(6905): 411–5.
79. Parthun MR, Widom J, Gottschling DE. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell. 1996;87(1):85–94.
80. Miller KM, Tjeertes JV, Coates J, Legube G, Polo SE, Britton S, et al. Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. Nat Struct Mol Biol. 2010;17(99):1144–51.
81. Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, et al. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell. 2003;11(3):721–9.
82. Nguyen AT, Xiao B, Neppl RL, Kallin EM, Li J, Chen T, et al. DOT1L regulates dystrophin expression and is critical for cardiac function. Genes Dev. 2011;25(3):263–74.
83. Huyen Y, Zgheib O, DiTullio Jr RA, Gorgoulis VG, Zacharatos P, Petty TJ, et al. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature. 2004;432(7015):406–11. doi:10.1038/nature03114.
84. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, Liu WT, et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature. 2010;466(7305): 508–12.
85. Denslow SA, Wade PA. The human Mi-2//NuRD complex and gene regulation. Oncogene. 2007;26(37):5433–8.
86. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012;81:65–95.
87. Bernt KM, Armstrong SA. Targeting epigenetic programs in MLL-rearranged leukemias. Hematology Am Soc Hematol Educ Program. 2011;2011(1):354–60.
88. Gayther SA, Batley SJ, Linger L, Bannister A, Thorpe K, Chin SF, et al. Mutations truncating the EP300 acetylase in human cancers. Nat Genet. 2000;24(3):300–3.
89. Pfister S, Rea S, Taipale M, Mendrzyk F, Straub B, Ittrich C, et al. The histone acetyltransferase hMOF is frequently downregulated in primary breast carcinoma and medulloblastoma and

constitutes a biomarker for clinical outcome in medulloblastoma. Int J Cancer. 2008;122(6):1207–13.
90. Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. Mol Oncol. 2007;1(1):19–25.
91. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet. 2011;43(9):860–3.
92. Huang C, Sloan EA, Boerkoel CF. Chromatin remodeling and human disease. Curr Opin Genet Dev. 2003;13(3):246–52.
93. Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, et al. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet. 2000;24(4):368–71.
94. Xu Y, Sun Y, Jiang X, Ayrapetov MK, Moskwa P, Yang S, et al. The p400 ATPase regulates nucleosome stability and chromatin ubiquitination during DNA repair. J Cell Biol. 2010;27:2010.
95. Faucher D, Wellinger RJ. Methylated H3K4, a transcription-associated histone modification, is involved in the DNA damage response pathway. PLoS Genet. 2010;6(8):e1001082.
96. Kusch T, Florens L, MacDonald WH, Swanson SK, Glaser RL, Yates JR, et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science. 2004;306(5704):2084–7.
97. Suganuma T, Gutierrez JL, Li B, Florens L, Swanson SK, Washburn MP, et al. ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nat Struct Mol Biol. 2008;15(4):364–72.
98. Suganuma T, Mushegian A, Swanson SK, Abmayr SM, Florens L, Washburn MP, et al. The ATAC acetyltransferase complex coordinates MAP kinases to regulate JNK target genes. Cell. 2010;142(5):726–36.
99. Bao Y, Shen X. INO80 subfamily of chromatin remodeling complexes. Mutat Res. 2007;618(1–2):18–29.
100. Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, et al. Histone cross-talk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell. 2007;131(6):1084–96.
101. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature. 2006;440(7084):637–43.
102. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084):631–6.
103. Boulon S, Ahmad Y, Trinkle-Mulcahy L, Verheggen C, Cobley A, Gregor P, et al. Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. Mol Cell Proteomics. 2010;9(5):861–79.
104. Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep. 2012;2(1):198–205.
105. Lambert JP, Fillingham J, Siahbazi M, Greenblatt J, Baetz K, Figeys D. Defining the budding yeast chromatin-associated interactome. Mol Syst Biol. 2010;6:448.
106. Lambert JP, Mitchell L, Rudner A, Baetz K, Figeys D. A novel proteomics approach for the discovery of chromatin-associated protein networks. Mol Cell Proteomics. 2009;8(4):870–82.
107. Daniels DL, Mendez J, Mosley AL, Ramisetty SR, Murphy N, Benink H, et al. Examining the complexity of human RNA polymerase complexes using HaloTag technology coupled to label free quantitative proteomics. J Proteome Res. 2012;11(2):564–75.
108. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005;123(4):581–92.
109. Keogh M-C, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, et al. Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell. 2005;123(4):593–605.

110. Kurdistani SK, Grunstein M. Histone acetylation and deacetylation in yeast. Nat Rev Mol Cell Biol. 2003;4(4):276–84.
111. Hah N, Kolkman A, Ruhl DD, Pijnappel WW, Heck AJ, Timmers HT, et al. A role for BAF57 in cell cycle-dependent transcriptional regulation by the SWI/SNF chromatin remodeling complex. Cancer Res. 2010;70(11):4402–11.
112. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010;143(3):470–84.
113. Smith KT, Workman JL. Histone deacetylase inhibitors: anticancer compounds. Int J Biochem Cell Biol. 2009;41(1):21–5.
114. Prince HM, Bishton MJ, Harrison SJ. Clinical studies of histone deacetylase inhibitors. Clin Cancer Res. 2009;15(12):3958–69.
115. Washburn MP. Driving biochemical discovery with quantitative proteomics. Trends Biochem Sci. 2011;36(3):170–7.

# Proteome Analysis of Chromatin Complexes in Differentiating Stem Cells

**Ariane Watson and Gerard Cagney**

**Abstract** Regulation of gene expression by proteins associated with chromatin is a major, yet poorly understood, feature of differentiation and development. Recent genomic studies have highlighted the role of chromatin regulatory proteins in pathologies affecting cellular proliferation and cell cycle, such as cancer. Mass spectrometry-based proteomics approaches have, in the last decade, provided a wealth of information on the dynamic nature of the proteome during cellular differentiation. Label-based approaches have predominated the literature, however, with the development of increasingly sensitive mass spectrometers and liquid chromatographic systems; label-free techniques offer a compelling alternative. Using these approaches, a vast repertoire of proteins have been identified in the proteome of undifferentiated and differentiating stem cells, including transcription factors, chromatin-modifying complexes, histone-modifying enzymes and signalling proteins which act in concert to regulate gene expression. Given the recent correlation between mutations in epigenetic machinery and the development and progression of various cancers, application of these approaches to the study cancer cell proteomes could provide valuable insights into the role of epigenetic reregulation in tumourogenesis.

**Keywords** Embryonic stem cells • NTERA-2 • Differentiation • Proteomics • Mass spectrometry • Chromatin • Epigenetics • Polycomb group proteins • Pluripotency • Self-renewal

A. Watson, B.Sc. • G. Cagney, Ph.D., B.Sc. (✉)
University College Dublin, The Conway Institute, School of Biomolecular and Biomedical Science, Belfield, Dublin, 4, Ireland
e-mail: ariane.watson@ucdconnect.ie; gcagney@gmail.com

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_10,
© Springer Science+Business Media New York 2014

# Epigenetic Regulation in Stem Cells and Cancer

## *Stem Cell Properties*

Stem cells have been a focus of intense research for several decades. In spite of this there are many unanswered questions regarding their regulation and the mechanisms by which they direct single-lineage (homogenous) differentiation. Stem cells exhibit two key properties: pluripotency, which is the ability to transform from an undifferentiated state [1] into any type of cell in the adult body and self-renewal, which constitutes the ability to maintain the undifferentiated phenotype through cell division and replication. Clearly, dysfunction of genes and proteins involved in these pathways are a potential cause of diseases of development and differentiation, and in particular diseases of proliferation such as cancer [2–4]. While stem cells represent a powerful source of cells with the potential to treat human disease through regenerative medicine [5, 6] (Fig. 1), in order to design disease-specific therapies and address safety concerns, a more profound understanding of stem cell biology is required.

Genomic DNA sequence in different cell types of an individual organism are almost identical; hence, it is hypothesised that epigenetic factors may explain much of the difference between cells that have highly varying roles in the body [7].

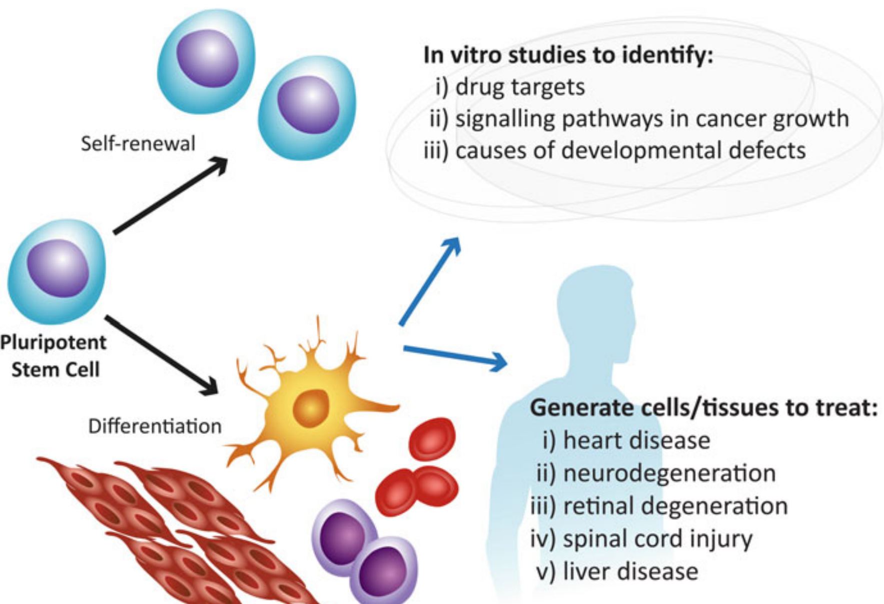



**Fig. 1** Stem cells have great potential for use in regenerative medicine and for studying human development and cancers

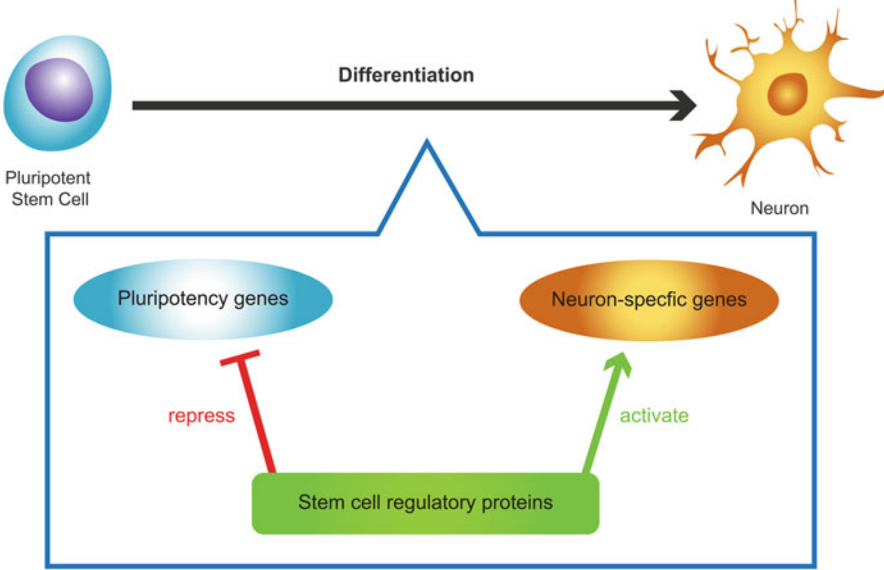



**Fig. 2** A simple model of differentiation

Epigenetic modifications are heritable changes which do not change the DNA sequence, but influence gene expression by controlling the access of transcriptional machinery to gene promoters, through alterations to the chromatin structure. Differentiation of a stem cell into any defined cell type involves specific alterations to gene expression, in order to repress pluripotency-associated genes and activate cell type-specific genes (Fig. 2); this is driven by epigenetic modifications ultimately brought about by protein factors and complexes, making proteomics an appropriate approach for studying the process of differentiation.

## *Epigenetics, Stem Cell Phenotype and Cancer*

Cellular (or transcriptional) memory involves the epigenetic modification of chromatin structure to regulate gene expression in a stably heritable manner. Extensive evidence suggests a role for protein complexes in either directly or indirectly altering the structure of nucleosomes, in order to regulate the access of the transcriptional machinery to gene promoters [8]. Nucleosomes are the fundamental unit of chromatin and comprise approximately 147 bp of DNA wound around an octamer of core histone proteins, H2A, H2B, H3 and H4. Chromatin structure can be altered through two types of epigenetic alterations: (a) histone modification or (b) ATP-dependent chromatin remodelling.

Covalent modifications to specific histone residues (especially on the N-terminal tails of histone proteins) alter chromatin structure and function. They include methylation, acetylation, phosphorylation and ubiquitination. The unique combination of these modifications has been proposed to form a readable "histone code" [9]. These modifications can alter the histone charge which may alter the DNA–histone association; alternatively, they can act as recognition and binding sites for other chromatin-altering proteins. Specific protein complexes are responsible for the establishment, maintenance and removal of histone marks, leading to a general classification into readers, writers and erasers of the histone code. For example, Polycomb Repressor Complex 2 (PRC2) lays down the H3K27me3 mark (trimethylation of lysine 27 of histone H3, associated with gene silencing) and, along with the activating H3K4me3 mark at bivalent domains, is a key component of stem cell identity and the "poised" state of developmentally important genes [10]. Other complexes read and interpret this code. For instance, Polycomb Repressor Complex 1 (PRC1) stabilises the inactive transcriptional state imposed by PRC2 so that the transcriptional memory is maintained. Other complexes contain enzymes such as lysine demethylases that remove the marks thereby relieving transcriptional repression at those loci. Other important epigenetic mechanism relies on altering accessibility of the transcriptional machinery to repressed genes through energy-dependent remodelling of nucleosomes or replacement of histone molecules with variant forms. ATP-dependent chromatin remodelling complexes are characterised by an ATPase subunit which drives the remodelling process. The exact mechanism, or mechanisms, of this remodelling are not fully understood, but current proposed mechanisms include: the replacement of some or all of the histone octamer; the repositioning of the histone octamer relative to DNA (nucleosome sliding); and the ejection of part or all of the histone octamer [11]. Some remodelling complexes such as the NUcleosome Remodeling and Deacetylase (NuRD), which possesses histone deacetylating activity, are also capable of altering specific histone tail residues. Epigenetic modifiers, such as the Polycomb (PcG) complexes, play crucial roles in biological processes such as transcription, DNA replication and repair, embryonic development and the differentiation of adult somatic cells. Given the fundamental role of these proteins, it is perhaps not surprising that recent studies (reviewed in [12]) have implicated the mutation of many of these proteins as key contributors to the initiation and progression of tumourigenesis.

The development of cancer involves both direct aberrant genetic alterations to tumour suppressors and proto-oncogenes, and the deregulation of gene expression patterns established through the epigenetic modifications described above. A genetic alteration may initiate the carcinogenesis, with an epigenetic alteration leading to its progression. Alternatively, epigenetic changes may prime cells so that a subsequent genetic alternation promotes carcinogenesis. Genetic mutation in epigenetic modifiers can lead to the deactivation of tumour suppressors, through mechanisms such as aberrant DNA methylation, or to the overexpression of proto-oncogenes through hyperacetylation or the destabilisation of normal silencing marks on these proto-oncogenes. Members of PcG complexes have recently been implicated as key contributors to the initiation and progression of many different types of human

cancers, through their regulation of the INK4B and INK4A-ARF loci which encode important tumour suppressor genes [13]. Overexpression of the PcGs causes inappropriate silencing of these loci, and prevents proper cellular senescence from occurring, thus allowing the persistence of tumourigenic cells. Given the emerging crucial role of epigenetic modifiers in cancer, they have provided new promising targets for therapeutic intervention [4, 14].

## Mass Spectrometry-Based Approaches to Studying the Dynamic Proteomes of Differentiating Cells

### *Model Systems for Studying Stem Cells*

One goal of stem cell systems biology is to develop a predictive model of the human Embryonic Stem Cell (hESC). As an experimental model, however, they are not always the most straightforward: they can differentiate into multiple cell types, but it is not yet understood how to effectively manipulate this in vitro. Somatic (or "adult") stem cells also provide an alternative model for research. Compared to pluripotent embryonic stem cells, these cells can only differentiate into a subset of cell types (multipotent) which is dependent on the type of somatic stem cell involved. For instance, neural stem cells can give rise to neurons as well as astrocytes and oligodendrocytes, while mesenchymal stem cells can differentiate into various cell types such as osteocytes, chondrocytes, adipocytes and other related cells found in connective tissues. However, as with ESCs, these stem cells can be difficult to culture and manipulate, as well as being relatively costly. Another alternative is a well-characterised immortalised human embryonal cancer cell line, NTERA-2 (or NT2), that when treated with retinoic acid (RA) undergoes differentiation. A major advantage of NT2 cells is that they undergo homogenous differentiation—differentiating only along a single (neuronal) lineage (Fig. 3) [15]. NT2 cells are, in comparison to traditional stem cells, less expensive and easier to handle. However, given that these cells are originally sourced from cancer tissue, the possibility remains that they may exhibit abnormal expression of certain genes, so results need to be treated with caution.

### *Label-Based Versus Label-Free Proteomics Approaches*

In the past decade, mass spectrometry-based proteomics has emerged as a powerful tool capable of describing the properties (identity, abundance, sequence variation, post-translational modification) of thousands of proteins in a single experiment. Thus, aspects of cancer and stem cell biology, which cannot be explained by genomics alone, can be investigated using proteomic technologies, revealing an intricate

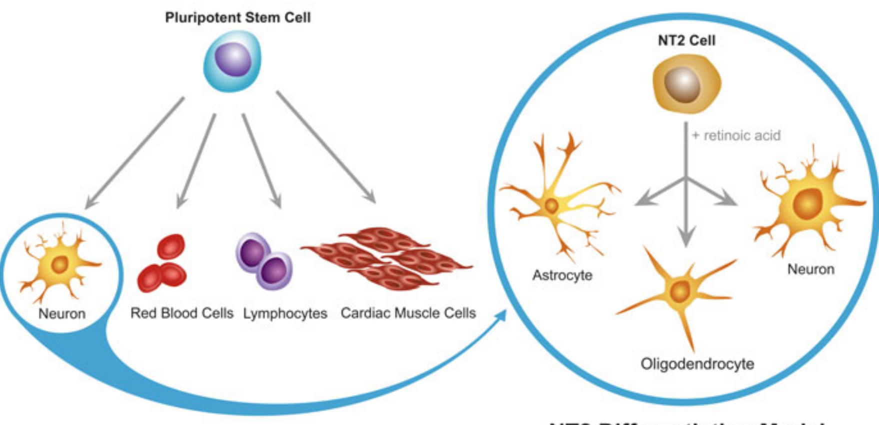



**Fig. 3** Comparison of heterogeneous ES cell differentiation vs. homogenous NT2 cell differentiation

network of signalling cascades, transcription programmes and histone and chromatin-modifying enzymes.

MS-based proteomics aims to achieve three levels of protein analyses: separation, quantification and identification. Traditionally, the separation approaches have largely involved two-dimensional (2D) separation methods such as the gel-based 2D gel electrophoresis (2DE) or 2D-differenitial gel electrophoresis (2D-DIGE) and more recently the quickly evolving 2D-liquid chromatography techniques. 2DE was one of the earliest proteomics approaches to be used, but suffers from a lack of reproducibility between gels, an issue which was improved upon by 2D-DIGE which allowed for the fluorescent labelling of samples, thus allowing mixing of samples on the same gel, circumventing the issue of gel-to-gel variation. Gel-free approaches, such as 2D-liquid chromatography, have become increasingly popular, being generally more sensitive than the gel-based counterparts, and more easily automated. Gel-free approaches also have an advantage when quantifying proteins, as multiple peptides from the same protein can be quantified whereas gel-based approaches rely on the use of a single spot's volume for quantifying that protein's abundance. However, these approaches are not necessarily mutually exclusive and are often used in a complementary fashion.

Protein identification in MS-based proteomics currently relies on the ability to measure peptides derived from tryptic digestion of the sample. Tryptic peptides share a number of properties (uniformity of mass and charge) that render them suitable for analysis by peptide mass fingerprinting or tandem mass spectrometry approaches [16]. Quantitation (or more specifically, determination of relative abundance) can be achieved either through the use of isotope-coded labels, e.g. metabolic or isobaric labelling, or through a label-free approach. Metabolic labelling, as used

in stable isotope labelling in cell culture (SILAC), involves incorporating heavy or light versions of a given amino acid into the proteins, by growing samples of cells on media which differs only in the presence or absence of the light or heavy form of a particular amino acid [17]. Using SILAC, an experimenter can compare two to three samples in a single run, as the "case" cells and control cells can be mixed directly after harvesting and combined together for protein extraction and all subsequent treatments and analyses. A popular alternative approach has been isobaric tag for relative and absolute quantification, or iTRAQ, which was developed in by Ross et al. [18]. iTRAQ involves the use of an isobaric tag which includes a reporter group with a different isotopic composition in each tag, thus allowing for discrimination between tags. Each reporter group has an associated balance group to prevent discrimination during MS protein identification; this balance group is separated from the reporter during MS/MS peptide fragmentation, allowing for the quantification of relative abundance between the different reporter tags. This technique has the advantage of being applicable to non-auxotrophically cultivable cells, a limitation of metabolic labelling, and allows for the comparison of up to eight different samples (or four different pairs of case/control samples).

The use of labels prior to sample preparation allows for the mixing of different samples and their simultaneous preparation, reducing any variability which may arise during the isolation and digestion of proteins. This improves the robustness of the experiment, as well as allowing for the simultaneous analysis of the samples by MS. Label-based approaches also come with limitations: the labels can be costly as well as time consuming to apply, and although sample pooling improves robustness, it often comes at the cost of resolving power when studying phenomena involving large numbers of proteins, such as stem cell differentiation. Many of the proteins fundamental to stem cell pluripotency and self-renewal (e.g. Oct4, Nanog) are exclusive to embryonic stem cells and present at low quantities. As such, the mixing of undifferentiated ESCs with differentiating cells easily results in the masking of these proteins by the more plentiful proteins present in the various differentiating samples.

Label-free MS-based proteomics provides a more widely applicable approach, only requiring that the sample preparation, chromatographic separation and data acquisition protocols are sufficiently robust to allow for comparison of different samples. There is no need for protein labelling and in theory no limitation on the number of samples which can be compared. However, given the inevitable variation that is likely to arise between different samples, due to factors such as pipetting error during sample preparation and variable chromatographic conditions, data normalisation is usually performed prior to comparison of results. The underlying principle upon which label-free MS quantification relies is that of the linear relationship between the MS peptide signal and protein abundance—in other words, the more prevalent the protein's representation in the mass spectra, the more abundant it is in the sample [19]. Built on from this principle is a strategy now termed Spectral Count (the frequency of tandem MS fragmentation events for a given peptide during an LCMS run); this strategy was investigated by Liu et al. [20] for use in measuring relative protein abundance, and validated through their statistical model for random sampling during MS/MS data acquisition. This model illustrated a direct relationship

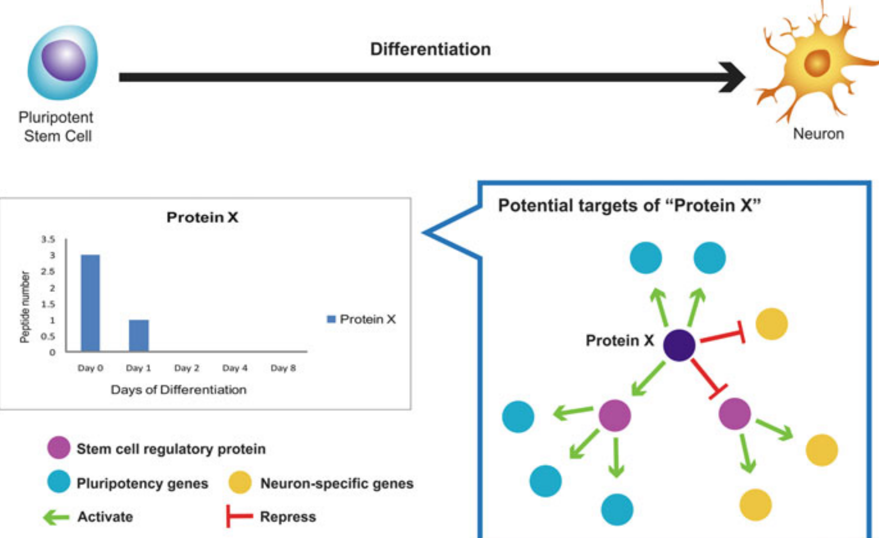



**Fig. 4** Hypothetical representation of information which can be derived through analysis of MS-based proteomic datasets

between a peptide ion's abundance in a peptide mixture and its probability of being identified in a MS/MS run, as long as all peptides in the mixture are analysed under the same conditions. Another approach, termed ion abundance, relies on estimating relative peptide quantities by integrating peptide signal across two dimensions (*m*/*z* ratio, chromatographic elution time). This approach was recently extended to three dimensions (by exploiting redundancies in the ion signal such as different isotopic forms of a peptide) [21]. Although these methods are normally confined to determination of "relative" abundance, they can report of absolute abundance if mass-coded (i.e. isotopic forms) reference peptides are added to the experiment. Both Spectral Count and Ion Abundance allow one to circumvent the use of stable isotope labels, greatly simplifying the experimental procedure.

## *Proteomic Approaches to Studying Epigenetic Regulation of Stem Cell Differentiation (Fig. 4)*

Early studies of stem cell epigenetic regulation have generally focused on mouse (mESCs) or human (hESCs) embryonic stem cells. Most examined small subsets of genes and/or proteins, and only a few investigated the proteome of differentiating cells. The firm establishment of several key pluripotency-inducing transcription factors by Yamanaka et al. [22, 23], including Oct4 and Sox2, provided a basis from which to study pluripotency and the mechanisms involved in its regulation. Their work

focused on the use of a few defined key factors to induce a pluripotent state onto already differentiated somatic cells; the resulting cells were coined induced pluripotent stem cells (iPSCs). Other groups such as van den Berg et al. [24] and Pardo et al. [25] aimed to determine the network of proteins regulating "stemness" by using FLAG-affinity proteomics approaches to examine the interactors of the key pluripotency factor Oct4 in ESCs. The results of these studies indicated a high degree of interaction between different transcription factors, with many showing co-occupancy of same gene promoters. Overlapping interaction of pluripotency factors with several remodelling complexes such as NuRD and hSWI/SNF was also observed, suggesting the coordinated activity of a complex network of proteins in regulating and maintaining an undifferentiated state.

Chaerkady et al. [26] used iTRAQ labelling of whole cell lysates to examine the proteome of hESCs over eight stages of differentiation. They identified and quantified 1,251 proteins using LC-MS/MS with a quadrupole time-of-flight (TOF) system. However, they were unable to identify any of the pluripotency factors (Oct4, Nanog, Sox2) which are only present in low abundances in the nucleus. The low resolution of their results was possibly due to a lack of pre-analysis subcellular fractionation, combined with a label-based approach which inherently requires the mixing of peptides from samples at different stages of differentiation. This severely dilutes the peptides of low abundance proteins, especially those (a) only present in the nucleus and (b) present only at the earliest stages of differentiation, as is the case with many pluripotency factors. Dilution to this extent would make detection of these peptides exceedingly unlikely.

More recently, Sarkar et al. [27] attempted to identify the proteomes of three subcellular fractions of undifferentiated hESCs: the nuclear, cytosolic and membrane fractions. Each fraction was separately SILAC labelled and combined with unlabelled protein samples from differentiating hESC cells, prior to analysis with a high-sensitivity LTQ-Orbitrap XL LC-MS/MS system. Using this approach they were able to identify 893 nuclear proteins, 1,397 cytosolic proteins and 1,185 membrane proteins, with an overall of ~70 % of the total identified proteins being unique to one fraction. Overall they achieved significant coverage (total of >3,000 proteins identified), likely due to their use of subcellular fractionation.

Jadaliha et al.[28] combined 8-plex iTRAQ with LC-MS/MS analysis to investigate the proteome of differentiating hESCs. They compared the proteome of undifferentiated hESCs to that of hESCs at different stages (days 6, 12 and 20) of spontaneous differentiation by embryoid body (EB) formation. The proteins in each sample were tagged using eight different iTRAQ labels, and subsequently pooled for trypic digestion into peptides which were fractionated in chromatographic columns prior to LC-MS/MS analysis. The group performed two technical replicates for each of their three biological replicates, and using this approach were able to identify 1,032 non-redundant proteins between all replicates. Of these, 156 proteins were deemed to exhibit statistically significant changes in abundace during differentiation, with 81 showing upregulation and 79 showing downregulation in at least one EB compared to the undifferentiated ESCs. The group carried out Western Blot analysis on five (ERP29, NPM1, HSC70, CALU and STMN1) of these 159 proteins

with the results of this supporting that of their MS results. The coverage achieved by the group included a wide range of proteins with differential abundance profiles during differentiation, including nucleic acid-binding protein, cytoskeletal proteins, ribosomal proteins, calcium-binding proteins and proteins involved in Integrin signalling. However, with the exception of Lin-28, the proteins identified did not include well-established proteins of the pluripotency network; given the ESC-exclusive expression and low abundance of proteins such as Oct4 and Nanog, it is possible that the pooling of differentiated and undifferentiated samples lead to the masking of these proteins by more abundant proteins in differentiating cells.

Pewsey et al. [29] examined the nuclear fraction of RA-induced differentiating NT2 cells, investigating samples from four different stages of differentiation, with a focus on the first 6 days. Both iTRAQ and ExacTag labelling approaches were used, with the respective samples pooled for analysis by LC-MS/MS with a quadrupole-ion trap system. Fifty four proteins were identified, including low abundance proteins such as Oct4 and Sox2 (but not Nanog), indicating that the sensitivity was relatively high. The combined results of the independent label-based experiments (iTRAQ and ExacTag) showed 37 proteins which exhibited statistically significant changes in abundance during the first 6 days of differentiation. These results were validated by both Western blotting and RT-PCR. Through the use of clustering and interaction analyses, they highlighted the presence of an intricate protein–protein interaction network in the regulation differentiation; the regulation likely occurring by the coordinated effort of numerous proteins, leading to the activation and/or silencing of specific target genes at defined stages of differentiation.

## *Label-Free Proteomics: An Optimised Approach*

In contrast to the studies described, our lab uses label-free MS analysis (Fig. 5). Sample preparation involves the subcelluar fractionation of harvested cells followed by a size-separation by simple one dimensional SDS-PAGE. Sample lanes are then excised and cut into a number of sections (usually a minimum of 10). Each of these sections are treated as separate sub-samples and subject to in-gel tryptic digestion and peptide extraction. Samples are further fractionated using a reverse-phase HPLC system prior to analysis by an LTQ Orbitrap Classic MS (Thermo Scientific), with a nano-electrospray ionisation source (Proxeon Biosystems). Peptides are analysed by LC-MS/MS with an initial survey scan collecting data continously in a data-dependent manner. Collision induced dissociation (CID) MS/MS scans use the ten most abundant ions from the survey scan. The resulting raw MS files are processed using a quantitative proteomics software application such as MAXQUANT. The principle of maximum parsimony [27] is applied in matching peptides to proteins in the database, with proteins identified in at least two of three experimental datasets being accepted. Identifications that are only based on one unique peptide, or two (or more) unique peptides in only one dataset, need to be manually validated by gauging the assignment of major peaks, the occurrence of continuous (at least three amino acids) y- or b-ion series, low mass error and the *p*-value. The False

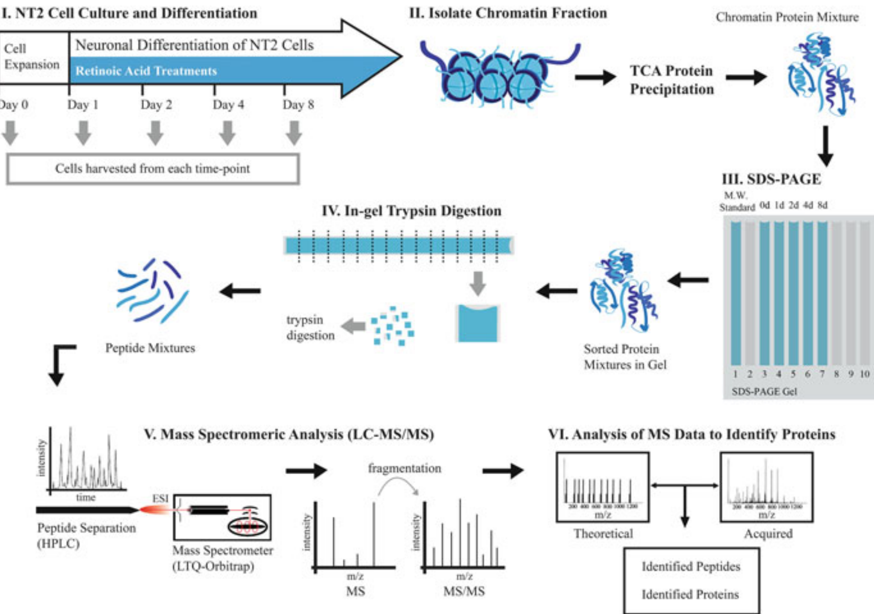



**Fig. 5** Workflow for a label-free proteomics approach to quantifying the changes in protein abundance during differentiation

Discovery Rate (FDR) is determined using a Reversed Sequence database, and Spectral Count (the frequency of productive MS/MS events for each protein) is used as an estimate of protein abundance [30].

Using an approach similar to the above, we have been able to identify over 2,400 proteins in the chromatin proteome of differentiating NT2 cells, using samples from five different time points: undifferentiated NT2 cells (day 0 of *all trans*-RA treatment) and four stages of RA-induced differentiation (days 1, 2, 4 and 8). Proteins identified included a large cohort of well-established plutipotency-related proteins such as Oct4, Nanog and Sox2. Markers for neuronal differentiation were also identified, including Nestin. Furthermore, a significant number of proteins from several chromatin- and histone-modifying enzyme complexes were identified, including members of PRC1, PRC2, NuRD, BRAF and Sin3. Data such as this, on changing protein abundance during differentiation, can provide the basis from which to extrapulate information on the regulatory networks involved in the controlling gene expression during celllular differentiation (see Fig. 4).

## Conclusion

Great advances have been made in proteomics technology in the last decade, in which a single experiment is capable of identifying and quantifying thousands of proteins. This capacity will be needed to address the enormous complexity of stem

cell biology, which uses pathways comprised of hundreds of individual proteins and dozens of protein complexes. Resolving the changes of abundance, location and activity of these protein activities across time and in different cell types still represents a formidable challenge. However, an improved understanding of the role of these protein networks in normal developmental processes is essential both for ensuring that human therapies based on stem cell biology are safe, and in other to uncover the basis of cancers induced by dysfunction in epigenetic regulator proteins.

## References

1. Odorico JS, Kaufman DS, Thomson JA. Multilineage differentiation from human embryonic stem cell lines. Stem Cells. 2001;19(3):193–204.
2. Bolli R, Chugh AR, D'Amario D, Loughran JH, Stoddard MF, Ikram S, et al. Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trail. Lancet. 2011;378(9806):1847–57.
3. Ruff CA, Fehlings MG. Neural stem cells in regenerative medicine: bridging the gap. Panminerva Med. 2010;52(2):125–47.
4. Takawa M, Masuda K, Kunizaki M, Daigo Y, Takagi K, Iwai Y, et al. Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. Cancer Sci. 2011;102(7):1298–305.
5. Wu SM, Hochedlinger K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat Cell Biol. 2011;13(5):497–505.
6. Strauss R, Hamerlik P, Lieber A, Bartek J. Regulation of stem cell plasticity: mechanisms and relevance to tissue biology and cancer. Mol Ther. 2012;20(5):887–97.
7. Vincent A, Van Seuningen I. Epigenetics, stem cells and epithelial cell fate. Differentiation. 2009;78(2–3):99–107.
8. Young RA. Control of the embryonic stem cell state. Cell. 2011;144(6):940–54.
9. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000; 403(6765):41–5.
10. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011; 469(7330):343–9.
11. Piatti P, Zeilner A, Lusser A. ATP-dependent chromatin remodeling factors and their roles in affecting nucleosome fiber composition. Int J Mol Sci. 2011;12(10):6544–65.
12. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1): 12–27.
13. Bracken AP, Kleine-Kohlbrecher D, Dietrich N, Pasini D, Gargiulo G, Beekman C, et al. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev. 2007;21(5):525–30.
14. Popovic R, Licht JD. Emerging epigenetic targets and therapies in cancer medicine. Cancer Discov. 2012;2(5):405–13.
15. Pal R, Ravindran G. Assessment of pluripotent and multilineage differentiation potential of NTERA-2 cells as a model for studying human embryonic stem cells. Cell Prolif. 2006;39(6):585–98.
16. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 2004 Sep;5(9):699–711.
17. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1(5):376–86.

18. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3(12):1154–69.
19. Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res. 2002;1(4):317–23.
20. Liu H, Sadygov RG, Yates 3rd JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76(14):4193–201.
21. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12): 1367–72.
22. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2006;126(4):663–76.
23. Takahashi K, Tanabe K, Ohnuki MA, Narita M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–72.
24. van den Berg DLC, Snoek T, Mullin NP, Yates A, et al. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell. 2010;6:369–81.
25. Pardo M, Lang B, Yu LA, Proser H, et al. An expanded Oct4 interaction network; implication for stem cell biology, development and disease. Cell Stem Cell. 2010;6:382–95.
26. Chaerkady R, Kerr CL, Marimuthu A, Kelkar DS, et al. Temporal analysis of neural differentiation using quantitative proteomics. J Proteome Res. 2009;8(3):1315–26.
27. Sarkar P, Collier TS, Randalf SM, Muddiman DC, et al. The subcellular proteome of undifferentiated human embryonic stem cells. J Proteomics. 2012;12:1–10.
28. Jadaliha M, Lee HJ, Pakzad M, Fathi A, Jeong SK, et al. Quantitative proteomic analysis of human embryonic stem cell differentiation by 8-plex iTRAQ labelling. PLoS One. 2012;7(6):e38532.
29. Pewsey E, Bruce C, Tonge P, Evans C, et al. Nuclear proteome dynamics in differentiating embryonic carcinoma (NTERA-2) cells. J Proteome Res. 2010;9:3412–26.
30. Cagney G, Park S, Chung C, Tong B, O'Dushlaine C, Shields DC, et al. Human tissue profiling with multidimensional protein identification technology. J Proteome Res. 2005;4(5):1757–67. PubMed PMID: 16212430.

# Structural Genomics and Drug Discovery for Chromatin-Related Protein Complexes Involved in Histone Tail Recognition

**Panagis Filippakopoulos and Stefan Knapp**

**Abstract** Recruitment of transcriptional regulators and enzymes that remodel chromatin structure is controlled by a complex pattern of post-translational modifications on histones and other chromatin binding proteins. These so-called epigenetic marks specifically recruit protein interaction modules that "read" the complex pattern of post-translational modifications resulting in assembly of protein complexes that alter chromatin structure and regulate gene transcription. Often, several diverse reader domains are present in nuclear chromatin modifying proteins acting synergistically to recognize post-translationally modified histones. In addition to this modular set of interactions, some reader domains simultaneously recognize combination of several post-translational marks, rather than isolated modifications. Due to the complexity and the large number of marks and their combinations, reader domains have evolved as large and diverse families of interaction modules that specifically recognize combinations of acetylated and methylated lysines, methylated arginines, phosphorylated serine, threonine and tyrosine residues as well as other modifications. High throughput protein crystallography has recently contributed significantly in our understanding of the structural mechanisms that govern reader–histone tail interactions. Established parallel expression and purification of recombinant reader domains have enabled screening technologies that evaluated the substrate specificity of entire families of these protein interaction modules. In addition, dysfunction of epigenetic mechanisms such as writing, erasing and reading of post-translational marks has been associated with the development of a large variety

P. Filippakopoulos, Ph.D.
Nuffield Department of Clinical Medicine, Medicine Division, Oxford University, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ, UK
e-mail: panagis.filippakopoulos@sgc.ox.ac.uk

S. Knapp, Ph.D. (✉)
Nuffield Department of Clinical Medicine, Medicine Division, Oxford University, Target Discovery Institute (TDI), Roosevelt Drive, Oxford, OX3 7BN, UK
e-mail: stefan.knapp@ndm.ox.ac.uk

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_11,
© Springer Science+Business Media New York 2014

of diseases, and reader domains have recently emerged as interesting targets for pharmaceutical intervention. Here, we will review our current knowledge of reader domain structural biology, the mechanisms of specific recognition of substrate sequences and emerging inhibitors that specifically disrupt binding of reader domains to histone tails. Due to space limitations we will limit our analysis in this chapter on reader domains that primarily recognize methylated lysine and arginine residues as well as acetyl-lysine readers of the bromodomain family.

**Keywords** Bromodomain • Acetyl-lysine reader domains • Methyl-lysine reader domains • MBP (malignant brain tumour) • PHD (plant homeodomain) • BAH (bromo adjacent homology) domains • Tudor • Chromodomain • PWWP • WD40 repeat proteins

## Architecture of Reader Domains

Acetylation of lysine residues changes significantly the physiochemical properties of the lysine side chain. Most notably the charge of the unmodified lysine at physiological pH is neutralized by this post-translation modification. In contrast, methylation of lysines and arginine residues maintains their positive charge making specific recognition of target sequences more challenging. It is therefore not surprising that at least some methyl-lysine reader domains are less stringent in differentiating between their methylated target sequences and unmodified or differently modified lysine side chains.

High resolution crystal structures are now available for several families of epigenetic reader domains, and high throughput structural genomics groups made a significant contribution to this effort in particular for bromodomains, PWWP and MBT domains (Fig. 1a) [1, 2].

Acetylation sites at ε-N-lysine residues are principally recognized by bromodomains, a family of 61 protein interaction modules present in 46 proteins [1, 3] that have been named after the Drosophila gene “brahma” for which the central bromodomain sequence motif was first recognized [4].

Despite the low level of sequence conservation, particularly in the terminal helices, all bromodomains share a conserved fold that comprises a left handed bundle of four alpha helices (αZ, αA, αB and αC). The four bromodomain core helices are linked by highly diverse loop regions (ZA and BC loops) that determine substrate specificity (Fig. 1b). Bromodomains harbour a large central cavity which constitutes the binding site for acetyl-lysine containing sequence motifs. The bromodomain acetyl-lysine binding site contains largely hydrophobic and aromatic residue in addition to a conserved asparagine that anchors the carbonyl group of acetyl-lysine by a hydrogen bond. However, a number of alternative residues exist that can potentially act as hydrogen bond donors, but no target sequences have been published so far for these bromodomains [1].

In contrast to the acetyl-lysine specific bromodomains, protein interaction modules that selectively recognize methyl-lysine and methyl arginine containing

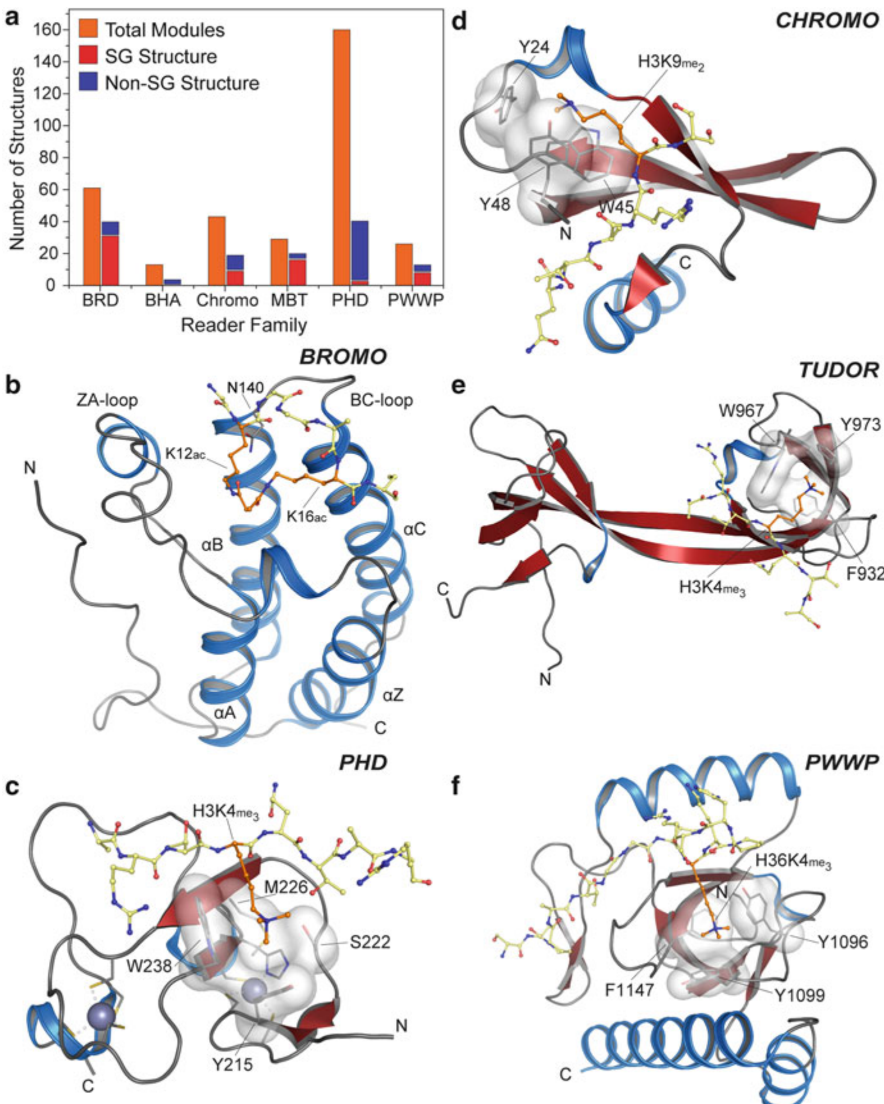



**Fig. 1** Structural coverage and representative folds of acetyl and methyl lysine specific binding modules. (**a**) Bar diagram showing available crystal structures of representative reader domains. Shown are all identified human targets (*orange bars*), structures determined by structural genomics laboratories (*red bars*) and other academic groups (*blue bars*). (**b**) The N-terminal bromodomain of BRD4 (1) (PDB-ID: 3UVX). (**c**) Representative structure of a PHD finger (PDB-ID: 2G6Q). $Zn^{2+}$ ions are shown as *spheres*. (**d**) Chromodomain of HP1 (PDB-ID: 1KNA). (**e**) Double tudor domain of JMJ2A (PDB-ID: 2GFA). (**f**) PWWP domain (PDB-ID: 2X4W). Peptide binding sites are indicated by showing interacting peptides in ball and stick representation, and binding pockets are shown as semi-transparent surfaces

sequences are structural highly diverse and contain mainly beta secondary structure. To date at least 300 methyl-lysine reader domains have been identified in the human proteome, but due to the high diversity of the sequence and structure of these readers, it is likely that the number of known methyl-lysine and methyl-arginine readers will increase in the future. This diverse family comprises members of the extended "Royal family" (KDM, MBT, Chromo), PWWP, WD40 repeat proteins, BAH (bromo adjacent homology) domains and the plant homeodomain (PHD).

PHD domains constitute the largest family of reader domains which has been originally identified as a conserved sequence motif in Arabidopsis. The substrate specificity of this small and diverse domain can vary from lysine trimethylation specific to domains recognizing sequences that contain unmodified lysine residues or even acetyl-lysine containing recognition sites [5–10]. The structure of a typical PHD finger comprises two atypical zinc fingers that coordinate one $Zn^{2+}$ ion each. The metal ions significantly stabilize the small, about 50 residues, PHD domain fold which mainly contains a beta sheet secondary structure. Small flexible domains such as the PHD pose a challenge to protein crystallography due to the difficulty of growing crystals. Most structural information on these small interaction modules has therefore been provided by nuclear magnetic resonance (NMR) studies. Interestingly, the PHD domain fold resembles the one described for E3 ligase of the RING finger family. However, PHD domains do not have E3 ligase activity due to the lack of the required interaction site for an E2 ligase. The methyl-lysine side chain is recognized by aromatic residues, so-called aromatic cages. Usually a central tryptophan that contributes one of the "walls" of the aromatic binding cage divides the PHD peptide recognition site into a primary methyl-lysine binding pocket and a secondary site that is recognized by flanking residues such as the unmodified arginine in the ING2 peptide complex shown in Fig. 1c.

The extended "Royal Family" consists of the methyl-lysine reader domains: Tudor, chromodomains, MBT (malignant brain tumour) and PWWP domains [11].

The Tudor, MBT and PWWP domains share a conserved architecture which is composed of five β-strands. The central three strands are also present in chromodomains which are C-terminally extended by an alpha helix. Members of the "Royal Family" that bind methylated sequence motifs coordinate the methyl ammonium group by a cluster of aromatic residues that form an aromatic cage-like structure.

More than 40 chromodomains are present in the human proteome. The characteristic features of this interaction domain are evident in the structure of the chromodomain of HP1 that binds methylated peptides in an extended β-strand conformation completing the β-sandwich architecture [12] (Fig. 1d). HP1 binds both di- and trimethyl lysine containing sequences, and structures of peptide complexes showed strong conservation of binding modes and peptide conformations. Interestingly, tudor domains are often found as double domains forming saddle-shaped arrangements (Fig. 1e). In the lysine demethylase JMJD2A the double tudor domain exchanges the third and fourth β strands with respect to the canonical tudor fold forming an extended cradle-like structure [13]. Interestingly, both tudor domains in JMJD2A contribute to the binding of one tri-methyl lysine site. The aromatic cage is mainly formed by residues in the tudor-2 motif whereas binding specificity is largely determined by side chain interactions with the tudor-1 motif [13].

PWWP domains are larger domains that comprise typically 110 residues and have been named after the central Pro-Trp-Trp-Pro motif. PWWP have been predicted to be protein interaction domains but have also been found to interact with DNA [14]. The first structure of the PWWP domain of the DNA methyltransferases Dnmt3a revealed a mixed beta sheet and alpha helical fold with the canonical N-terminal five-stranded beta barrel fold and a C-terminal half containing a five-helix bundle [14].

Comparison of seven representative diverse members of this protein interaction module family as well as complexes with interacting peptides revealed that PWWP domains harbour an insertion motif located between the second and third β-strands and a C-terminal α-helical bundle which together with the conserved beta-barrel recognizes methyl-lysine containing sequences in histones [2] (Fig. 1f).

BAH (bromo adjacent homology) domains have also been recently identified as methyl-lysine binding modules. The BAH domain of ORC1, a component of ORC (origin of replication complex) specifically recognizes histone H4 dimethylated at lysine 20 (H4K20$me_2$) [15]. In the case of this class of protein interaction modules, the dimethyl-ammonium moiety of H4K20$me_2$ containing peptides interacts with an aromatic cage by a π-interaction as well as by hydrogen bonding to a glutamate located in the binding pocket. However, aromatic cages are not found in all BAH domains suggesting that a large diversity of peptides may be recognized by these domains.

Members of the WD40 family have also been reported to recognize methylated lysine and arginine residues. WD40 is one of the largest families of protein interaction domains recognizing a vast diversity of different binding partners including DNA as well as unmodified, phosphorylated and methylated sequence motifs [16]. The WD40 domain architecture consists of a β-propeller fold comprising usually seven blades. The WD40 repeat protein EED (embryonic ectoderm development) is a regulator of the PRC2 (the polycomb repressive complex 2), and it is required for methyltransferase activity of the PRC2 histone lysine methyltransferase EZH2. Interestingly, the EED WD40 domain contains an aromatic cage which has been shown to specifically bind to H3K27$me_3$ [17]. Profiling against a larger set of methylated histone peptides and co-crystal structures showed that the EED WD40 domain also recognizes other $Kme_3$ histone sequences that contain the tri-methylated Ala-Arg-Lys-Ser motif present in H3K27, H3K9 and H1K26 marks, with similar binding affinities [18].

Sequences containing methylated arginine residues are recognized by the WD40 domain of WDR5, a subunit of the MLL (mixed lineage leukaemia) co-activator complex. Arginines can be dimethylated symmetrically (ω-NG,N′G-dimethylarginine) by type I protein-arginine methyltransferases (PRMTs) or asymmetrically (ω-NG,NG-dimethylarginine) by type II PRMTs. Interestingly, symmetric H3R2$me_2$ methylation enhances binding of WDR5 whereas asymmetry methylation of H3R2 inhibits binding [19].

The diversity of methyl-lysine binding domains makes it likely that many more protein interaction modules with methyl-lysine binding activity will be recognized in the future. An updated list of available crystal structures of reader domains as well as epigenetic enzymes and associated inhibitors is available online on the Structural Genomics Consortium Web site (http://apps.thesgc.org/resources/phylogenetic_trees/).

## Recognition of Histone Target Sequences

Protein interaction modules that recognize acetylated- or methylated-lysine and arginine residues are present in many nuclear proteins and often coexist in combinations with several interaction domains within the same protein. The multiple reader domain architecture suggests that recognition of epigenetic marks is a modular process involving interactions with a number of diverse marks that may be located in different histones, nuclear proteins or even in different nucleosomes. The complicated arrangement of reader domain in proteins and protein complexes poses a formidable challenge for our understanding of the epigenetic code.

For individual reader domains, the progress in peptide array technology has enabled systematic studies that elucidated sequence specificity of protein interaction modules [20]. A recent large-scale structural characterization of the family of human bromodomains highlighted the diversity of electrostatic properties that reader modules can exhibit, suggesting that despite their highly conserved tertiary structure, necessary for the specific recognition of ε-*N*-acetylated lysine residues, interactions with linear peptide motifs are mainly driven by the electrostatic nature of the surface area lining up the central recognition site (Fig. 2). In addition, the same study revealed histone peptide recognition motifs of most members of the bromodomain family [1]. This comprehensive study revealed that (a) many bromodomains do not target histones and are probably recruited to other nuclear or non-nuclear protein complexes; (b) many bromodomains recognize poly-acetylated rather than singly acetylated lysine sites; and (c) posttranslational modifications flanking the recognized acetylation site have a significant effect on the bromodomain peptide recognition process. This unexpected finding was first reported by Moriniere and co-workers by binding studies and co-crystallization on the murine bromodomain protein BRDT [21]. Interestingly, both acetyl-lysine side chains bind to the same bromodomain acetyl-lysine binding site, a binding mode that was also observed for the N-terminal bromodomain of human BRD4 [1]. In all crystal structures of BRD4 and BRDT histone H4 complexes with diacetylated peptides and the N-terminal bromodomains, both acetylated lysines bound simultaneously and with highly similar peptide conformations. In that binding mode the conserved bromodomain asparagine (N140 in BRD4(1)) forms a hydrogen bond with the carbonyl of the N-terminal histone peptide acetyl-lysine. The acetyl-lysine carbonyl forms an additional water mediated hydrogen bond with a conserved tyrosine residue (Y97 in BRD4(1)) (Fig. 3a). The binding mode of the N-terminal acetyl-lysine is reminiscent of the one observed in peptide complexes of bromodomains with singly acetylated histone peptides. The second acetyl-lysine does not form any direct polar interactions with the bromodomain; instead, it interacts mainly with the N-terminal acetyl-lysine by a water mediated hydrogen bond effectively occupying the bromodomain binding site.

In contrast to bromodomains that are specifically recognizing acetylated lysine containing sequences, most methyl-lysine reader families interact with a large variety of modifications including not only the preferential recognition of different

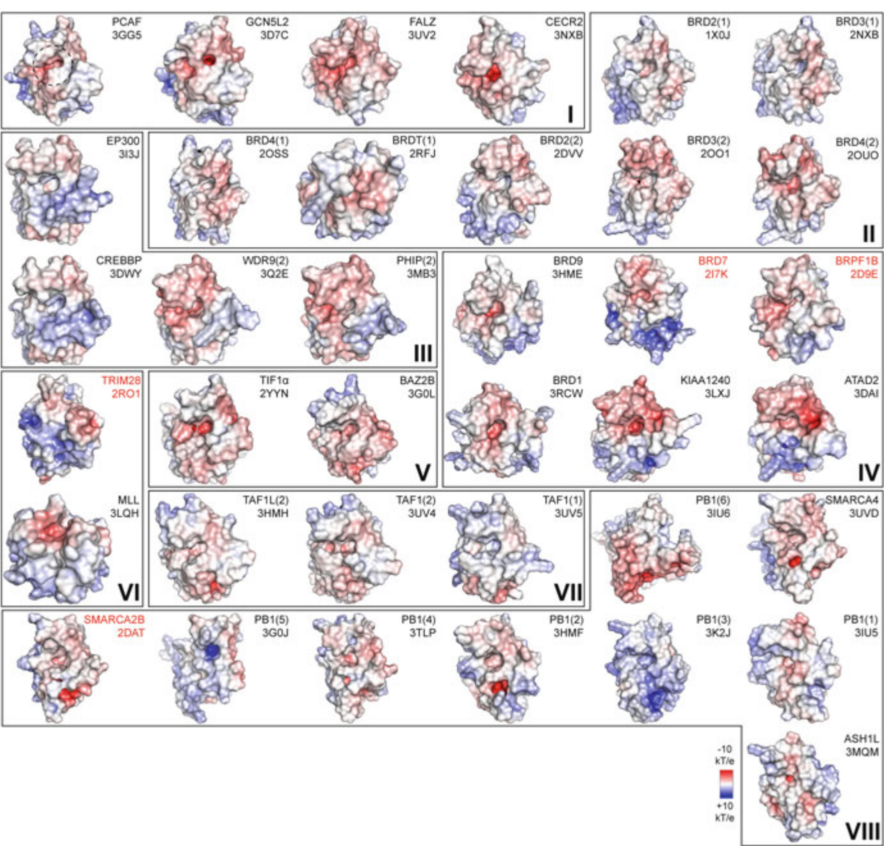



**Fig. 2** Surface charge properties of human Bromodomains. The domains are grouped into the eight BRD families (shown in *roman numerals*). Electrostatic surface potentials are shown between −10kT/e (*red*) and +10kT/e (*blue*). The BRD names and structures (PDB accession code in *black* for crystal structures and *red* for NMR models) are shown in the figure. All domains are shown in identical orientation with their acetyl-lysine binding site facing the reader and highlighted with a *dashed circle* on the *top-left* structure (PCAF). Reprinted from Cell, 149(1), Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, Felletar I, Volkmer R, Müller S, Pawson T, Gingras AC, Arrowsmith CH, Knapp S. Histone recognition and large-scale structural analysis of the human bromodomain family. 214–31, Copyright (2012), with permission from Elsevier

methylation states but also specific interaction with unmodified sequence motifs, as well as acetylated lysine residues. In methyl-lysine specific reader domains methyl-ammonium moieties are recognized by aromatic cages which stabilize binding by π-stacking interactions (Fig. 3b). Peptides bind usually in an extended conformation flanking the characteristic beta sheet structure of these reader domains and making multiple contacts with additional, usually acidic, binding residues on the reader domain surface. Reader domains that do not harbour aromatic cages usually recognize unmodified peptide sequences in a similar binding mode.

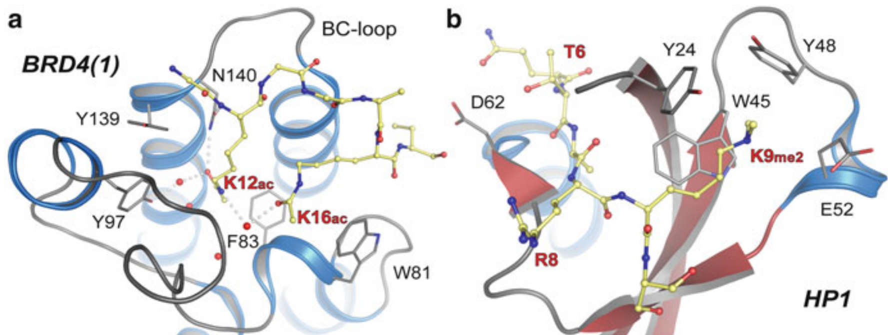



**Fig. 3** Examples of histone recognition by reader domains. (**a**) Complex of the N-terminal bromodomain of BRD4 with a histone H4 peptide containing acetylation sites at K12 and K16 (PDB-ID: 3UVX). Histone peptide carbon atoms are coloured in *yellow* and peptide residues are labelled in *red*. Carbon atoms in depicted bromodomain residues are shown in *grey*. Conserved binding site water molecules are highlighted as *spheres*. The main interacting residues of BRD4 are shown in ball and stick representation and are labelled. Key hydrogen bonds are highlighted by *dotted lines*. (**b**) Chromodomain complex of *Drosophila* HP1 with a histone H3 peptide harbouring K9me$_2$ (PDB-ID: 1KNA). Residues are labelled in a similar way as described in panel A

Structural evidence exists already suggesting that there is co-operativity between reader domains in recognizing multiple modifications on histones. For example, in order for histone H3 peptides to bind to TRIM33, implicated in the control of transcription elongation, the bromodomain module engages the histone tail via its K18ac mark (Fig. 4a) while the PHD finger engages to the K9me$_3$ mark (Fig. 4b). The peptide backbone of the histone substrate is recognized by the extended surface area of the tandem BRD/PHD module (Fig. 4c).

## Role of Reader Domains in Disease

Mutation and genetic translocations in reader domain containing proteins have linked many of these proteins to the development of diseases. Genetic rearrangements of bromodomains of the BET (bromo and extra-terminal) subfamily and CREBBP (cAMP response element binding binding protein) for instance have been found in a number of aggressive solid tumours and leukaemia [22–24]. Overexpression of a number of bromodomain-containing proteins has been reported in cancer, correlating with patient survival. ATAD2 for instance is overexpressed in >70 % of breast tumours and elevated protein levels correlate with tumour histologic grades, poor overall survival and disease recurrence [25]. Also, high expression levels of TRIM24 have been reported in breast cancer leading to poor prognosis and patient survival [26]. The role of bromodomain proteins in development of cancer and other diseases has recently been reviewed [27].

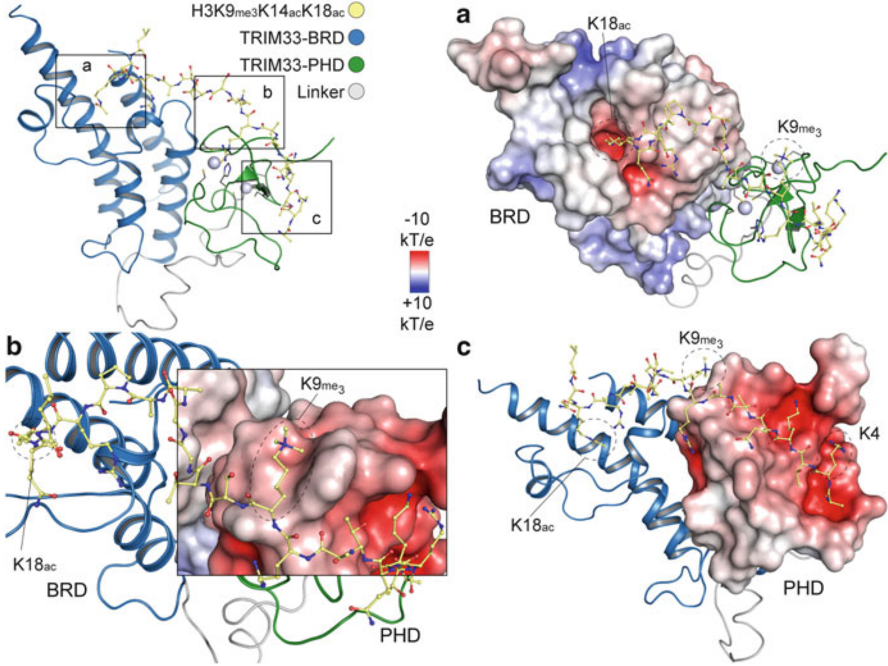



**Fig. 4** Structure of a TRIM33 BRD/PHD domain peptide complex. (**a**) Structural overview of the complex is shown in ribbon representation on the *left panel*. The peptide is shown in ball and stick representation, and $Zn^{2+}$ ions are indicated as *white spheres*. Three areas highlighted in the detailed views (**a**, **b** and **c**) are highlighted as *boxes*. In the highlighted area (**a**) interaction of K18ac with the bromodomain are shown. Panel (**b**) highlights interactions of $K9me_3$. Finally interactions of unmodified K4 are shown in panel (**c**). The figure has been generated using PDB entry 3U5O

A large number of mutations in PHD fingers have been associated with the development of diseases. For instance, chromosomal translocations and mutations have been detected in immune and neurological disorders as well as in cancer, recently reviewed by Baker et al. [28]. For example, immunodeficiency syndrome T-B-SCID (Severe Combined Immunodeficiency) and Omenn syndrome have been associated with mutations in the PHD domain of RAG2 recombinase, a key enzyme in V(D)J gene recombination [29]. The occurrence of disease causing mutations and genetic translocations in reader domains makes a compelling case for targeting these protein recognition modules for the development of novel therapeutic strategies.

## Bromodomain Inhibitors

The bromodomain acetyl-lysine binding pocket is particularly attractive as a targeting site for the development of protein interaction inhibitors. Acetylation of lysines neutralizes the charge of the ε-amino group resulting in largely hydrophobic and

aromatic binding pockets with good druggability [30]. The first bromodomain inhibitors were reported about 8 years ago by the Zhou laboratory [31]. The small molecule NP1 targets the bromodomain in the histone acetyl transferase PCAF (P300/CBP-associated factor) with an IC50 value of about 1.6 μM. The interaction of the inhibitor with the PCAF acetyl-lysine pocket was confirmed by 2D $^{15}N$ HSQC NMR spectroscopy.

Initial inhibitors have also been reported by the same laboratory that target CREBBP using azobenzene based templates such as 4-hydroxyphenylazo-benzenesulfonic acid (MS456) and Ischemin [32]. The most potent compound of this series exhibited an $IC_{50}$ value of 19.6 μM and modulated p53 function by interfering with the recruitment of CREBBP to this key regulator of cell cycle and DNA damage signalling. Surprisingly, the structure of the NP1 and MS7972 complexes with their target bromodomains PCAF and CREBBP showed that these inhibitors are not acetyl-lysine mimetics and do not therefore form a hydrogen bond with the conserved asparagine (N803 in PCAF or N1168 in CREBBP) potentially explaining the limited potency of these early inhibitors.

First highly potent inhibitors for bromodomains were reported in the patent literature. These inhibitors selectively target BET bromodomains (Example 1 from WO2009/84693 and Example 2 from WO/2011/054844) but were not rationally designed for these targets [33, 34]. Benzodiazepines and thienodiazepines were originally identified in phenotypic screens, monitoring the effect of gene expression of ApoA1 (apolipoprotein A1), a target for atherosclerosis and inflammation [35]. Interaction of these inhibitors with bromodomains of the BET family was subsequently identified by an elegant chemoproteomic analysis. The discovery of this interesting mode of action led to the development of the thienodiazepines JQ1 and MS417 as well as the benzodiazepine I-BET [36–38]. Co-crystal structures with these highly potent inhibitors revealed a conserved binding mode in which the triazolo-thieno-diazepine (JQ1) or triazolo-benzo-diazepine (IBET) ring system functions as an acetyl mimetic moiety (Fig. 5A (a)). Interestingly, the stereo centre in IBET and JQ1 led to highly active and analogous inactive stereoisomers. The introduction of a nitrogen atom in the place of the asymmetric carbon resulted in the discovery of highly selective benzotriazepines [39]. Substitution of the methyl-triazolo acetyl-lysine mimetic group with a methyl-isoxazole led to initial fragments of a different scaffold, Isoxazole 4d, which had an $IC_{50}$ value of 4.8 μM with excellent ligand efficiency, found to be selective for BET bromodomains despite its small fragment-like character [40]. The isoxazole acetyl-lysine scaffold was also used for the development of the highly potent and specific BET inhibitor I-BET151 which also showed improved pharmacokinetic properties [41]. The structure activity relationship (SAR) of bromodomain inhibitors has recently been reviewed [42].

A number of diverse other scaffolds binding to BET bromodomains have also been discovered. For instance, the fragment 3-methyl dehydroquinazoline-one which has been discovered as a fragment hit binding to CREBBP has been optimized to yield the highly potent and selective BET inhibitor PFI-1 (Fig. 5A (b)), the first chemical probe that has been co-developed by an academic group and an industrial partner [43]. Interestingly, the clinical inhibitor RVX-208 [44] which is

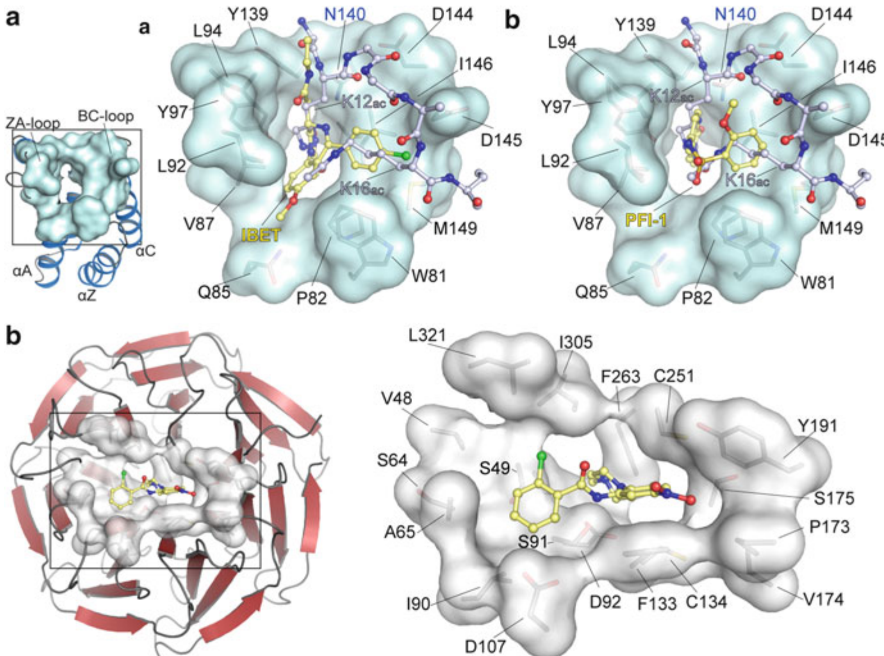



**Fig. 5** Inhibitor binding modes. (**a**) Superimposition of BRD4(1) peptide complexes with the inhibitors IBET (PDB ID: 3P5O) and PFI-1 (PDB ID: 4E96). Peptides and inhibitors are shown in ball and stick representation. Inhibitor carbon atoms are shown in *yellow* and peptide carbon atoms in *white*, respectively. A structural overview is shown in the *left panel*. (**b**) Binding of UNC1215 to WDR5 (PDB ID: 3SMR). Shown is a structural overview (*left panel*) as well as details of the interaction (*right panel*). The inhibitor is shown in ball and stick representation and main binding site residues (shown as a surface) are labelled

currently in phase II clinical trials for acute coronary syndromes, atherosclerosis and Alzheimer disease is reported to act as a BET bromodomain inhibitor (see company webpage for details http://www.resverlogix.com/).

## Methyl-Lysine Reader Inhibitors

Methyl-lysine reader domains have just recently emerged as targets for inhibitor development and only a small number of compounds have been reported so far that target the methyl-lysine binding site. Many methyl-lysine readers have also less attractive binding sites when compared to bromodomains, but a systematic analysis of all available methyl-reader structures and comparison with acetyl-lysine binding pockets identified a number of targets with good predicted druggability [45]. Most promising of these targets are members of the PWWP domains in addition to

**Fig. 6** Structures of current inhibitors that target reader domains. In the case of bromodomain inhibitors the acetyl-lysine mimetic groups that have been verified experimentally are highlighted in *red*

chromo-domains and MBT domains whereas Tudor and PHD fingers exhibited the lowest druggability scores.

First inhibitors with low μM binding affinity have been reported for the MBP domain L3MBTL1 (lethal (3) MBP protein 1) with most potent inhibitors having a $K_D$ of 5 μM [46, 47]. Inhibitors against this target have been recently developed further resulting in UNC1215, a highly potent (120 nM) cell active and selective inhibitor for L3MBTL3 [48].

A first inhibitor has also been reported for the WD40 protein WDR5. The WDR5 antagonist WDR5-0103, binds to the methyl-lysine binding pocket (Fig. 5B) with a $K_D$ of 450 nM and has been shown to inhibit the catalytic activity of the MLL core complex in vitro [49]. The good predicted druggability and the link of many methyl-lysine binding domains in the development make also these reader domains interesting targets for the development of specific chemical probe molecules. At the moment the chemical biology targeting reader domains has just emerged, and it will be interesting to see if any of the developed probe molecules will progress to new therapies in the clinic in the future. Structures of known bromodomain and methyl-lysine binder inhibitors are compiled in Fig. 6.

**Acknowledgements** The authors receive funding from the SGC, a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, Pfizer, Eli Lilly, Takeda, AbbVie, the Novartis Research Foundation, the Ontario Ministry of Research and Innovation and the Wellcome Trust. P.F. is supported by a Wellcome Trust Career Development Fellowship (095751/Z/11/Z). We apologise to the researchers that we were not able to cite as a result of space constraints.

## References

1. Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell. 2012;149(1):214–31.
2. Wu H, Zeng H, Lam R, Tempel W, Amaya MF, Xu C, et al. Structural and histone binding ability characterizations of human PWWP domains. PLoS One. 2011;6(6):e18919.
3. Filippakopoulos P, Knapp S. The bromodomain interaction module. FEBS Lett. 2012;586(17):2692–704.
4. Haynes SR, Dollard C, Winston F, Beck S, Trowsdale J, Dawid IB. The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. Nucleic Acids Res. 1992;20(10):2603.
5. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006; 442(7098):91–5.
6. Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature. 2006; 442(7098):100–3.
7. Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature. 2010; 466(7303):258–62.
8. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714–7.
9. Musselman CA, Kutateladze TG. Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res. 2011;39(21):9061–71.
10. Champagne KS, Kutateladze TG. Structural insight into histone recognition by the ING PHD fingers. Curr Drug Targets. 2009;10(5):432–41.
11. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci. 2003;28(2):69–74.

12. Jacobs SA, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science. 2002;295(5562):2080–3.
13. Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006;312(5774):748–51.
14. Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol. 2002; 9(3):217–24.
15. Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature. 2012;484(7392):115–9.
16. Xu C, Min J. Structure and function of WD40 domain proteins. Protein Cell. 2011; 2(3):202–14.
17. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury 3rd WJ, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009;461(7265):762–7.
18. Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci USA. 2010;107(45):19266–71.
19. Migliori V, Muller J, Phalke S, Low D, Bezzi M, Mok WC, et al. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. Nat Struct Mol Biol. 2012;19(2):136–44.
20. Nady N, Min J, Kareta MS, Chedin F, Arrowsmith CH. A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. Trends Biochem Sci. 2008; 33(7):305–13.
21. Moriniere J, Rousseaux S, Steuerwald U, Soler-Lopez M, Curtet S, Vitte AL, et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature. 2009; 461(7264):664–8.
22. French CA, Miyoshi I, Aster JC, Kubonishi I, Kroll TG, Dal Cin P, et al. BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t(15;19). Am J Pathol. 2001;159(6):1987–92.
23. Sobulo OM, Borrow J, Tomek R, Reshmi S, Harden A, Schlegelberger B, et al. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). Proc Natl Acad Sci USA. 1997;94(16):8732–7.
24. Panagopoulos I, Fioretos T, Isaksson M, Samuelsson U, Billstrom R, Strombeck B, et al. Fusion of the MORF and CBP genes in acute myeloid leukemia with the t(10;16)(q22;p13). Hum Mol Genet. 2001;10(4):395–404.
25. Ciro M, Prosperini E, Quarto M, Grazini U, Walfridsson J, McBlane F, et al. ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. Cancer Res. 2009;69(21):8491–8.
26. Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, et al. TRIM24 links a non-canonical histone signature to breast cancer. Nature. 2010;468(7326):927–32.
27. Muller S, Filippakopoulos P, Knapp S. Bromodomains as therapeutic targets. Expert Rev Mol Med. 2011;13:e29.
28. Baker LA, Allis CD, Wang GG. PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. Mutat Res. 2008;647(1–2):3–12.
29. Sobacchi C, Marrella V, Rucci F, Vezzoni P, Villa A. RAG-dependent primary immunodeficiencies. Hum Mutat. 2006;27(12):1174–84.
30. Vidler LR, Brown N, Knapp S, Hoelder S. Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. J Med Chem. 2012;55(17):7346–59.
31. Zeng L, Li J, Muller M, Yan S, Mujtaba S, Pan C, et al. Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. J Am Chem Soc. 2005;127(8):2376–7.
32. Borah JC, Mujtaba S, Karakikes I, Zeng L, Muller M, Patel J, et al. A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. Chem Biol. 2011;18(4):531–41.

33. Crowe M, Daugan AC-M, Gosmini RLM, Grimes RM, Mirguet O, Mordaunt JE. International patent application WO2011054844A1. 2011
34. Miyoshi S, Ooike S, Iwata K, Hikawa H, Sugahara K. International patent application WO2009084693A1. 2009.
35. Chung CW, Coste H, White JH, Mirguet O, Wilde J, Gosmini RL, et al. Discovery and characterization of small molecule inhibitors of the BET family bromodomains. J Med Chem. 2011;54(11):3827–38.
36. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, et al. Selective inhibition of BET bromodomains. Nature. 2010;468(7327):1067–73.
37. Nicodeme E, Jeffrey KL, Schaefer U, Beinke S, Dewell S, Chung CW, et al. Suppression of inflammation by a synthetic histone mimic. Nature. 2010;468(7327):1119–23.
38. Zhang G, Liu R, Zhong Y, Plotnikov AN, Zhang W, Zeng L, et al. Down-regulation of NF-kappaB transcriptional activity in HIV-associated kidney disease by BRD4 inhibition. J Biol Chem. 2012;287(34):28840–51.
39. Filippakopoulos P, Picaud S, Fedorov O, Keller M, Wrobel M, Morgenstern O, et al. Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. Bioorg Med Chem. 2012;20(6):1878–86.
40. Hewings DS, Wang M, Philpott M, Fedorov O, Uttarkar S, Filippakopoulos P, et al. 3,5-dimethylisoxazoles act as acetyl-lysine-mimetic bromodomain ligands. J Med Chem. 2011;54(19):6761–70.
41. Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan WI, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature. 2011;478(7370):529–33.
42. Hewings DS, Rooney TP, Jennings LE, Hay D, Schofield CJ, Brennan PE, et al. Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. J Med Chem. 2012;55(22):9393–413.
43. Fish PV, Filippakopoulos P, Bish G, Brennan PE, Bunnage ME, Cook AS, et al. Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. J Med Chem. 2012;55(22):9831–7.
44. Bailey D, Jahagirdar R, Gordon A, Hafiane A, Campbell S, Chatur S, et al. RVX-208: a small molecule that increases apolipoprotein A-I and high-density lipoprotein cholesterol in vitro and in vivo. J Am Coll Cardiol. 2010;55(23):2580–9.
45. Santiago C, Nguyen K, Schapira M. Druggability of methyl-lysine binding sites. J Comput Aided Mol Des. 2011;25(12):1171–8.
46. Herold JM, Wigle TJ, Norris JL, Lam R, Korboukh VK, Gao C, et al. Small-molecule ligands of methyl-lysine binding proteins. J Med Chem. 2011;54(7):2504–11.
47. Herold JM, Ingerman LA, Gao C, Frye SV. Drug discovery toward antagonists of methyl-lysine binding proteins. Curr Chem Genomics. 2011;5:51–61.
48. James LI, Barsyte-Lovejoy D, Zhong N, Krichevsky L, Korboukh VK, Herold JM, et al. Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. Nat Chem Biol. 2013;9(3):184–91.
49. Senisterra G, Wu H, Allali-Hassani A, Wasney GA, Barsyte-Lovejoy D, Dombrovski L, et al. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. Biochem J. 2013;449(1):151–9.

# Proteomic Analysis Reveals a Role for the GTPase RPAP4/GPN1 and the Cochaperone RPAP3 in Biogenesis of All Three Nuclear RNA Polymerases

**Diane Forget, Philippe Cloutier, Céline Domecq, and Benoit Coulombe**

**Abstract** Biogenesis of nuclear RNA polymerases (RNAP) is a poorly understood, yet central molecular process in eukaryotes. Recent analysis of interaction partners of RNAP II, the enzyme that synthesizes protein-coding mRNAs, in the soluble fraction of cell extracts identified a series of factors that play central roles in RNAP II biogenesis. The GPN loop GTPase RPAP4/GPN1 was shown to be required for nuclear import of RNAP II, and the HSP90 co-factor RPAP3 is essential for cytoplasmic assembly of this multisubunit enzyme. Examination of the list of interactors for RNAP II as well as RPAP4/GPN1 and RPAP3 reveals the presence of many specific subunits of RNAP I and III, which synthesize most of the cell's non-coding transcripts. This finding suggests that biogenesis of all three nuclear RNAPs may be coupled. Silencing of RPAP4/GPN1 and RPAP3 further indicates that both factors are essential for normal nuclear localization of the three polymerases. We present a model in which biogenesis of RNAP I, II and III is integrated through the action of assembly and nuclear import factors.

**Keywords** RNA polymerase • RNA polymerase II-associated proteins (RPAP) • Biogenesis • Nuclear import • Assembly

## Introduction

The nucleus of eukaryotic cells contains three types of RNA polymerases (RNAP). RNAP I is located in the nucleolus where it synthesises the 45S ribosomal RNA (rRNA) precursor that makes up the core of the ribosome. RNAP II and RNAP III

---

D. Forget • P. Cloutier • C. Domecq • B. Coulombe (✉)
Laboratory of Gene Transcription and Proteomics, Institut de recherches cliniques de Montréal, 110 avenue des Pins Ouest, Montréal, QC, Canada, H2W 1R7
e-mail: diane.forget@ircm.qc.ca; philippe.cloutier@ircm.qc.ca; celine.domecq@ircm.qc.ca; benoit.coulombe@ircm.qc.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_12,
© Springer Science+Business Media New York 2014

can both be found in the non-nucleolar nuclear space. RNAP II synthesizes mRNAs as well as a number of small nuclear RNAs (snRNA) and microRNAs (miRNA). RNAP III directs the synthesis of rRNA 5S, tRNAs and other small, non-coding RNAs. Each RNAP is assisted by a general transcription machinery which serves for promoter recognition, RNA chain elongation and response to transcriptional regulators. Work performed over the last three decades has revealed numerous mechanisms by which transcription is regulated in eukaryotes, many of the regulatory factors being involved in helping RNAPs to cope with the nucleosomal structure of chromatin [1–4; and references therein].

Much less is known about regulation of RNAP molecules prior to and following transcription. Biogenesis and recycling of these key enzymes have only been addressed recently. Two discoveries dealing with RNAP II, which has been the most intensively studied nuclear RNAP, have had a particularly important impact on our understanding of RNAP biogenesis. First, RNAP II molecules involved in transcription on chromatin are recycled at each cell cycle [5, 6]. Indeed, the bulk of RNAP II is released from chromatin during mitosis, more precisely at metaphase, and is re-imported to the nucleus after mitosis. Because the nuclear envelope reforms at late anaphase [7], this result implies that recycled RNAP II molecules have to re-enter the nucleus through nuclear import mechanisms. Recent results from our laboratory indicate that the same scenario is true for RNAP I and RNAP III (Forget et al., in preparation). Second, a group of factors that interact with RNAP II in the soluble cell fraction has been identified and at least some of them were shown to act as regulators of RNAP II biogenesis [8, 9]. The RNAP II-associated protein 4 (RPAP4; also termed GPN1) is member of a novel GTPase family characterized by the presence of a Glu-Pro-Asn (GPN) loop motif [10]. RPAP4/GPN1 shuttles between the cytoplasm and the nucleus in a CRM1-dependent manner [11, 12]. Silencing of RPAP4/GPN1 results in abnormal accumulation of RNAP II in the cell's cytoplasm, suggesting a role in nuclear import of this polymerase [11–13]. Substitutions in the RPAP4/GPN1 GPN loop or GTP binding motif provoke cytoplasmic retention of RNAP II, an indication that the GTPase activity is required for RNAP II nuclear import [12]. Treatment of the cells with benomyl, a compound that interferes with microtubule assembly/integrity, also interferes with RNAP II nuclear localization [12]. Notably, treatment of yeast strains having substitutions in RPAP4/GPN1 that produce slow growth phenotypes with sub-lethal concentrations of benomyl completely abolished growth. These results indicate that microtubule assembly is somehow involved in RNAP II nuclear import. The RNAP II-Associated Protein 3 (RPAP3) is a HSP90 co-factor that is part of a multisubunit complex consisting of RPAP3 itself, a TPR domain-containing factor that mediates direct interaction with HSP90, as well as other components that may have a chaperone function of their own, including prefoldin-like proteins and the RUVBL1 and RUVBL2 AAA+ATPases [14–16]. The canonical prefoldin complex is well known for its role in the chaperoning and polymerization of actin and tubulin [17, 18], while RUVBL1 and RUVBL2 are required for the assembly of other multimeric protein complexes, like snoRNPs [19]. Boulon et al. [20] have shown that HSP90 and RPAP3 are involved in assembly of RNAP II in the cell's cytoplasm prior to import

to the nucleus. Indeed, silencing of RPAP3, a mainly cytoplasmic protein, also causes abnormal accumulation of RNAP II in the cytoplasm. Because silencing of RPAP4/GPN1 and RPAP3 has similar effects on RNAP II localization, we propose that RNAP II assembly and nuclear import are tightly coupled. RNAP II molecules that are either recycled at mitosis or newly assembled in the cytoplasm through the action of HSP90 and RPAP3 are imported to the nucleus through the action of RPAP4/GPN1 in a process that requires microtubule assembly/integrity.

Biogenesis of RNAP I and III has not been characterized in much detail [21, 22]. For example, we do not know whether the same set of RNAP II-specific factors are involved or else, whether distinct machineries are at play [23]. What is established, however, is that all three nuclear RNAPs share some subunits [24, 25], suggesting that their biogenesis could somehow be interconnected. In this chapter, we report on results of affinity purification coupled with mass spectrometry (AP-MS) showing that RPAP4/GPN1 and RPAP3 are part of complexes containing subunits of all three nuclear RNAPs, including both shared and specific RNAP subunits. Moreover, silencing experiments reveal that both RPAP4/GPN1 and RPAP3 are necessary for normal nuclear import of all three nuclear RNAPs. These results strengthen the conclusion that biogenesis of RNAP I, II and III is tightly coupled, requiring some common factors including RPAP3 and RPAP4/GPN1.

## Materials and Methods

### *Protein Affinity Purification Coupled with Mass Spectrometry*

Generation of cell lines expressing tandem affinity peptide (TAP) or FLAG tagged RNAP or RPAP subunits and tandem affinity purification were performed as previously described [9, 12, 14, 26, 27]. After TCA precipitation, the eluates were digested with trypsin, and the resulting tryptic peptides were purified and identified by tandem mass spectrometry (LC-MS/MS) using a microcapillary reversed-phase high pressure liquid chromatography coupled LTQ-Orbitrap (ThermoElectron) quadrupole ion trap mass spectrometer with a nanospray interface, as we recently described [28]. Protein database searching was performed with Mascot 2.2 (Matrix Science) against the human NCBInr protein database. Mascot scores and spectral counts were used to select specific interactors for Fig. 1.

### *Antibodies*

The antibodies used in this study were obtained from various sources: POLR1A monoclonal antibody (Santa Cruz); anti-FLAG monoclonal antibody (Sigma); anti-RPAP4 antibody (CIM Antibody Core, Arizona State University, Tempe, Arizona);

| Interactors | Tagged proteins | | | | | |
|---|---|---|---|---|---|---|
| | POLR1A | POLR2A | POLR3A | RPAP4 | RPAP3 | |
| POLR2A | 135/23 | 8149/912 | | 6030/324 | 254/8 | |
| POLR2B | 159/6 | 5202/498 | | 4863/335 | 1566/72 | |
| POLR2C | | 1023/70 | | 1043/60 | | |
| POLR2D | | 972/95 | | 69/1 | | RNAP II subunits |
| POLR2G | | 651/36 | | 651/29 | | |
| POLR2I | | 506/30 | | 380/18 | | |
| POLR2J | | 269/22 | | 133/7 | | |
| POLR2E | 281/12 | 599/29 | 539/22 | 486/19 | 362/18 | |
| POLR2F | | 134/9 | 171/8 | 66/5 | | Shared subunits RNAP I, RNAP II and RNAP III |
| POLR2H | 383/11 | 743/122 | 522/20 | 538/13 | 118/4 | |
| POLR2K | | 55/4 | 82/7 | 60/3 | | |
| POLR2L | | 56/4 | 127/7 | 95/4 | | |
| POLR3A | 288/14 | 143/4 | 3727/177 | 2738/127 | 1918/88 | |
| POLR3B | | | 2213/105 | 2115/623 | 169/8 | |
| POLR3C | | | 1213/45 | | | |
| POLR3D | | | 469/21 | 466/21 | | |
| POLR3E | | | 1703/82 | 1598/70 | 92/4 | RNAP III subunits |
| POLR3F | | | 895/40 | | | |
| POLR3G | | | 476/22 | | | |
| POLR3H | | | 396/18 | 204/7 | | |
| POLR3K | | | 259/13 | 107/3 | | |
| CRCP | | | 494/23 | | | |
| POLR1C | 550/21 | 339/15 | 819/41 | 505/17 | 112/5 | Shared subunits RNAP I and RNAP III |
| POLR1D | | | 384/17 | 80/4 | | |
| POLR1A | 7552/711 | 1362/26 | | 92/3 | 226/7 | |
| POLR1B | 1083/57 | 443/11 | | 229/5 | 39/1 | RNAP I subunits |
| POLR1E | 255/11 | 268/10 | | | | |
| TWISTNB | 78/4 | 54/2 | | | | |
| CD3EAP | 112/5 | 366/19 | | | | |
| ZNRD1 | 110/2 | | | | | |
| RPAP4/GPN1 | | 427/33 | 70/2 | 1612/331 | 312/17 | RPAP4/GPN1 |
| RPAP3 | 1242/54 | 1303/60 | 794/25 | 1848/68 | 2882/175 | |
| PDRG1 | 51/3 | 227/11 | 142/4 | 144/12 | 186/10 | |
| PFDN2 | 199/6 | 245/8 | 185/7 | 250/7 | 209/6 | |
| PFDN6 | | 162/8 | 225/12 | 174/15 | 184/9 | |
| PIH1D1 | 489/14 | 563/15 | 196/07 | 634/25 | 281/9 | RPAP3/R2TP/PFDL complex |
| RUVBL1 | 1155/61 | 1414/72 | 583/18 | 1158/53 | 1755/122 | |
| RUVBL2 | 1281/60 | 1686/86 | 546/17 | 1554/68 | 1943/16 | |
| URI1 | 606/22 | 622/24 | 510/20 | 1095/45 | 1199/65 | |
| UXT | 52/2 | 267/5 | 86/4 | 74/10 | 204/6 | |
| WDR92 | 773/30 | 303/9 | 721/27 | 727/26 | 1038/52 | |
| HSP90AA1 | 1160/45 | 749/35 | 190/8 | 698/27 | 938/40 | |
| HSP90AB1 | 1204/51 | 987/45 | 232/9 | 1284/52 | 1246/52 | |
| HSPA8 | 1710/88 | 2868/260 | 1018/42 | 2107/99 | 1748/100 | |

**Fig. 1** Summary of affinity purification coupled with mass spectrometry data using POLR1A, POLR2A, POLR3A, RPAP4/GPN1 and RPAP3. Total mascot score/spectral counts are provided for each interaction. Interactors are grouped in various classes as shown on the *right*

anti-RPAP3 antibody (Abnova); horseradish peroxidase-conjugated secondary antibody (GE Healthcare); anti-B-tubulin monoclonal antibody (Sigma) and Alexa Fluor 488 (Invitrogen).

### *Transfection and siRNA Silencing*

Transfection experiments for generating stable HeLa cell lines expressing FLAG-tagged versions of POLR2A and POLR3A used lipofectamine, as described by the supplier (Invitrogen) [12]. RPAP4/GPN1 (ON-TARGETplus SMART pool), RPAP3 (ON-TARGETplus SMART pool), and control (siCONTROL Non-targeting pool) siRNAs (Dharmacon) were doubly transfected into HeLa cells using oligofectamine (Invitrogen) at a siRNA final concentration of 100 nM [12]. The efficiency of silencing was monitored for each experiment using western blotting.

### *Immunofluorescence and Imaging*

Immunofluorescence and imaging using HeLa cells were performed as previously described [12]. Immunofluorescence studies used an anti-FLAG antibody to localize exogenously expressed FLAG-POLR2A and FLAG-POR3A, whereas a monoclonal antibody raised against POLR1A was used to monitor localization of endogenous POLR1A. Indeed, we have been unable to generate a cell line expressing a FLAG tagged version of POLR1A that localizes normally to the nucleolus.

## Results

### *RPAP4/GPN1 and RPAP3 Interact with Subunits of RNA Polymerase I, II and III*

Affinity purification of RPAP4/GPN1 and RPAP3, followed by identification of binding partners by mass spectrometry, revealed that both factors interact with all three nuclear RNAPs, namely RNAP I, II and III subunits (Fig. 1). In addition to RNAP shared subunits (POLR1C, POLR1D, POLR2E, POLR2F, POLR2H, POLR2K and POLR2L) that copurified with RPAP4/GPN1 and/or RPAP3, specific RNAP I (POLR1A and POLR1B), RNAP II (POLR2A, POLR2B, POLR2C, POLR2D, POLR2G, POLR2I and POLR2J) and RNAP III (POLR3A, POLR3B, POLR3D, POLR3E, POLR3H and POLR3K) subunits were identified as well in these purifications. Interaction of the RPAPs with subunits of all three RNAPs suggests that they either interact independently with all three enzymes or with a

megacomplex containing the three polymerases. Interestingly, affinity purification of the largest RNAP subunits, POLR1A, POLR2A and POLR3A, resulted in purification of other large RNAP subunits (e.g. presence of POLR2A, POLR2B and POLR3A in the POLR1A purification, and of POLR1A and POLR3A in the POLR2A purification). These copurifications argue in favour of the existence of a megacomplex containing many subunits of all three RNAPs during nuclear RNAP biogenesis. However, the data in Fig. 1 also show that not all RNAP subunits copurify with all tagged RNAP subunits (for example, POLR3A was the only RNAP III subunit found to copurify with POLR2A in our experiments). Whether this finding reflects the formation of a megacomplex containing only a selection of RNAP subunits during the assembly process or reflects a lack of sensitivity of our AP-MS technology is not known at this time. Of note, POLR1A, POLR2A and POLR3A interact with the assembly chaperone HSP90 (HSP90AA1 and HSP90AB1). RPAP4/GPN1 also interacts reciprocally with the RPAP3-R2TP-PFDL complex. Together, these results indicate that RNAP I and III interact with the RPAPs involved in biogenesis of RNAP II.

### *Silencing of RPAP4/GPN1 and RPAP3 Results in Abnormal Accumulation of RNA Polymerase I, II and III Subunits in the Cytoplasm*

To address a putative function of RPAP4/GPN1 and RPAP3 in biogenesis of RNAP I and III, we used siRNA-directed silencing of both factors and monitored the intracellular localization of POLR1A and POLR3A, the largest subunit of RNAP I and III, respectively. As mentioned in the Introduction section, silencing of either RPAP4/GPN1 or RPAP3 results in cytoplasmic accumulation of RNAP II subunits. Figure 2 shows that independent silencing of RPAP4/GPN1 and RPAP3 has a similar effect on POLR2A, POLR3A and POLR1A, all three largest polymerase subunits showing an accumulation in the cytoplasm, as determined by immunofluorescence. Control siRNA did not alter RNAP localization. A western blot showing the efficiency of siRNA silencing is also included. These results indicate that similar mechanisms are at play to regulate biogenesis of nuclear RNAPs, and that they involve at least some of the same regulatory factors, including RPAP4/GPN1 and RPAP3.

## Discussion

Our results indicate that biogenesis of all three eukaryotic nuclear RNAPs, RNAP I, RNAP II and RNAP III, uses a common set of factors. Indeed, silencing of the GTPase RPAP4/GPN1 and the HSP90 cochaperone RPAP3 are essential to

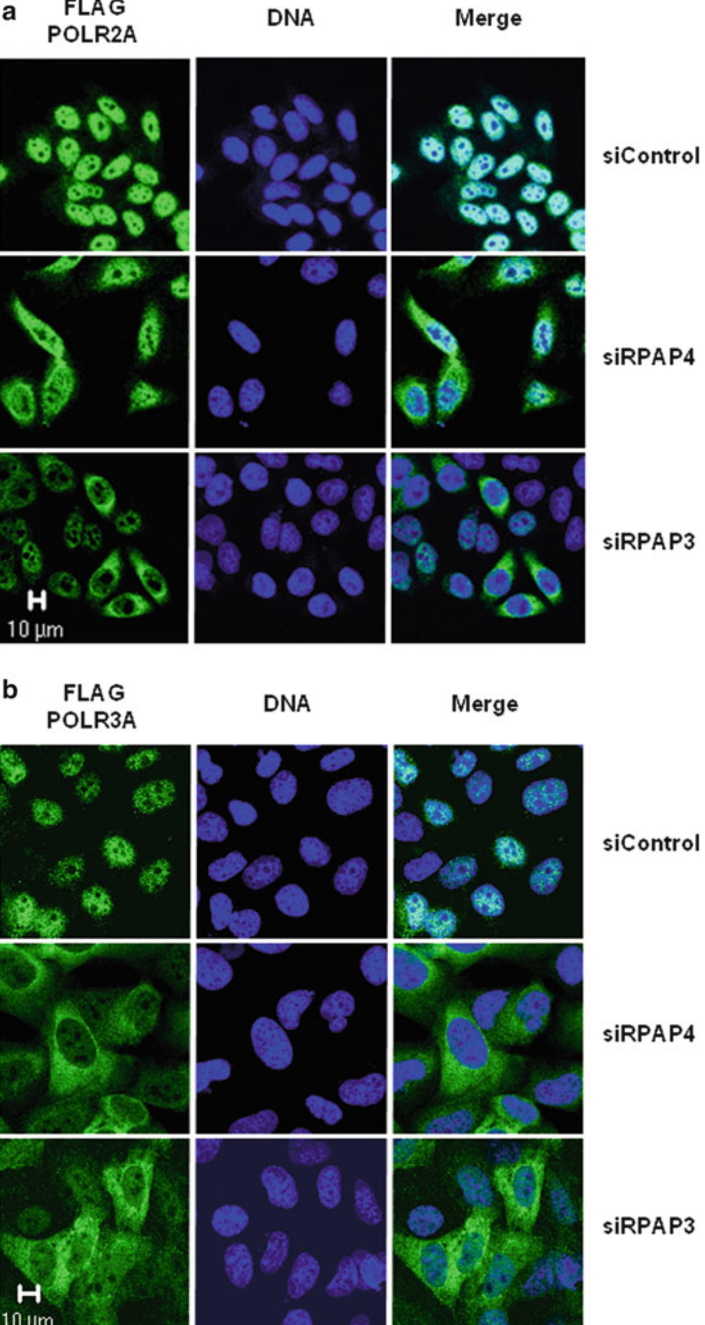



**Fig. 2** Immunofluorescence experiments showing the intracellular localization of POLR2A (**a**), POLR3A (**b**) and POLR1A (**c**) following RPAP4/GPN1 and RPAP3 silencing. In each case, a control experiment is shown for comparison. DNA staining with TO-PRO-3 iodide served to visualize nuclei. Silencing efficiencies have been monitored by western blotting (**d**)

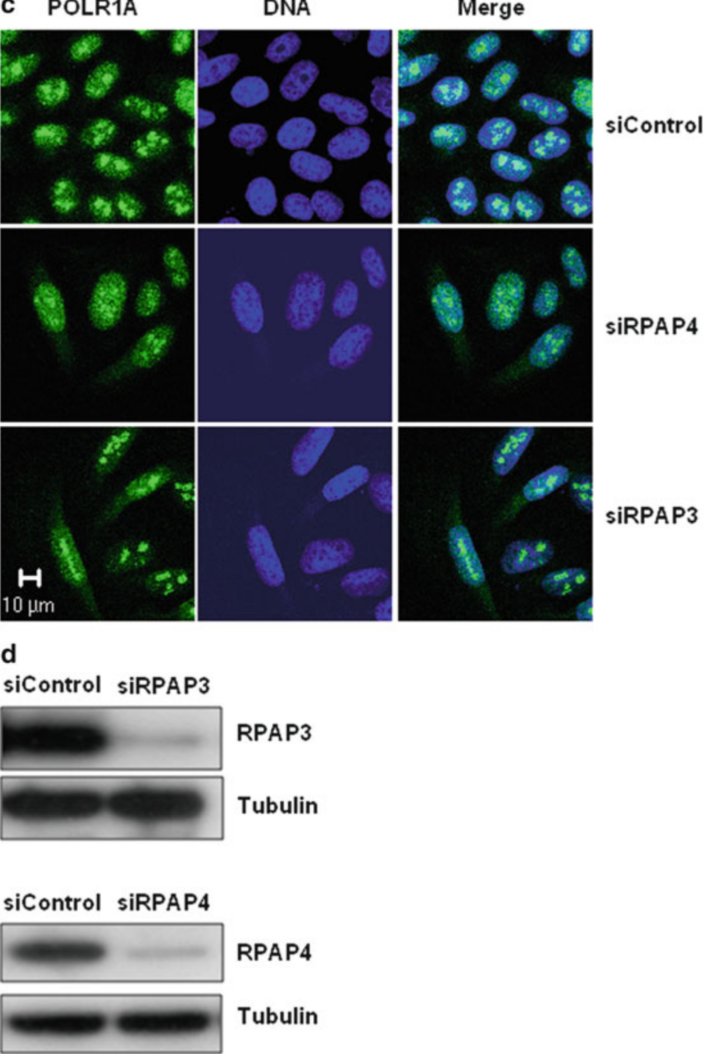



**Fig. 2** (continued)

maintain normal nuclear localization of all three enzymes. Silencing of either factor resulted in the cytoplasmic accumulation of all three RNAPs. These results further suggest that nuclear RNAPs are assembled and imported to the nucleus in a tightly coordinated manner.

Contrary to RNAP II and III, RNAP I molecules have to be targeted to the nucleolus after accessing the nuclear space. It is interesting to note that RPAP4/GPN1 and RPAP3 silencing both lead to accumulation of POLR1A in the cytoplasm, although the effect not being as striking as in the case of POLR2A and POLR3A.

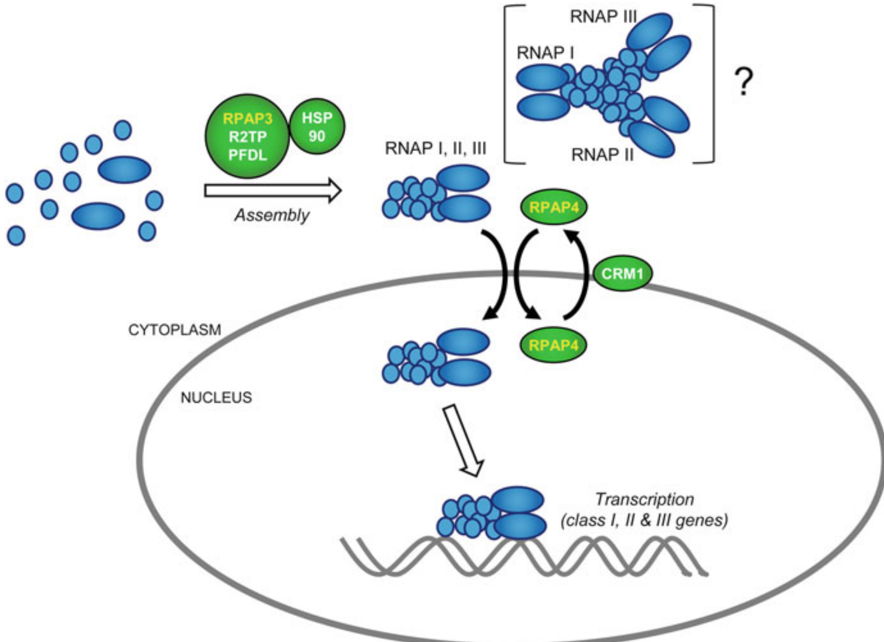



**Fig. 3** Model depicting that RNAP I, RNAP II and RNAP III biogenesis is coordinated through the action of common factors. RPAP3 and RPAP4/GPN1 are essential for biogenesis of all three nuclear RNAPs as their silencing results in abnormal cytoplasmic accumulation of the largest subunit of all three polymerases. The model takes into account previous results showing that RPAP3 is a HSP90 cochaperone involved in RNAP II assembly, and RPAP4/GPN1 is a GTPase involved in RNAP II nuclear import. The existence of a putative multi-RNAP megacomplex as an intermediate in RNAP assembly is shown, but remains mainly speculative at this point

Figure 3 presents a model in which RNAP I, RNAP II and RNAP III biogenesis proceeds through a common pathway involving the same set of regulatory factors. In this model, RPAP3 is involved in assembly of all three nuclear enzymes, most likely through the action of HSP90, and RPAP4/GPN1 participates in nuclear import of the three polymerases. Whether these factors interact independently with each RNAP or else with a megacomplex composed of RNAP I, II and III subunits is not known, although our proteomic data argues in favour of the existence of such a megacomplex as an intermediate in nuclear RNAP assembly. We expect that one or more not yet identified additional factors might be required to target RNAP I to the nucleolus.

Identification of factors required for biogenesis of nuclear RNAPs has been mainly the result of targeted proteomics studies. Our own group published a number of AP-MS datasets which differ by the use of an always increasing number of affinity purified tagged components [9, 12, 14, 27]. Figure 4 shows an interaction network defined by our laboratory using classical AP-MS from soluble whole cell extracts [12]. In these experiments, chromatin is discarded prior to protein

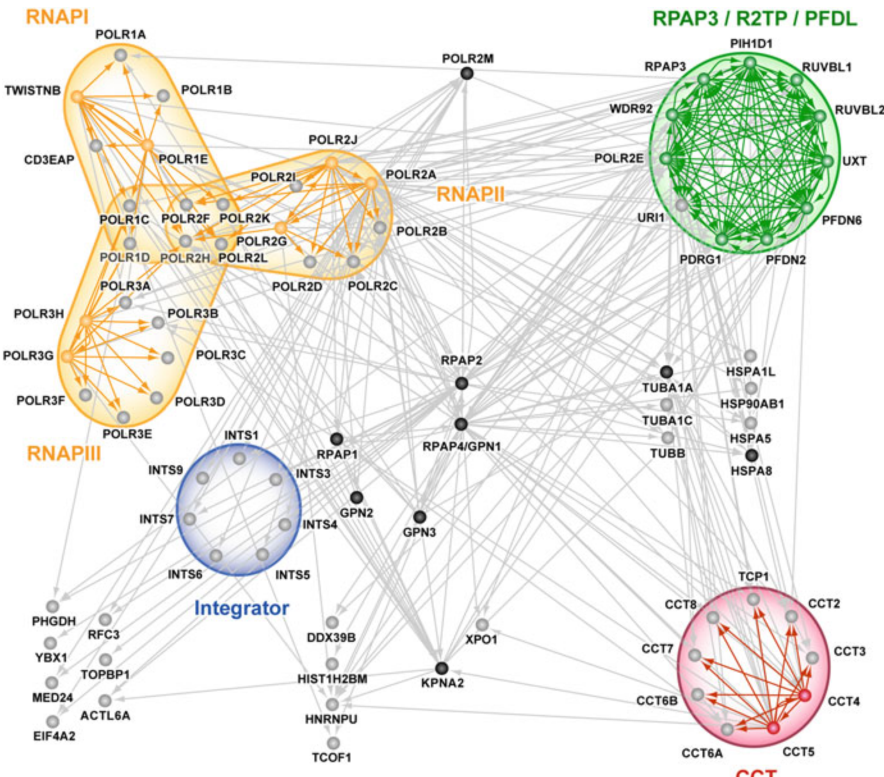



**Fig. 4** Network of interactions formed by nuclear RNAP and RPAP subunits in the soluble cell fraction. Components of previously characterized multisubunit complexes are grouped. In this diagram affinity tagged proteins used in AP-MS experiments are coloured and their copurified interactors are represented by an edge

extraction. As a consequence the resulting network is largely enriched in soluble factors (i.e. factors that interact with RNAP during transcription on chromatin are mostly absent). Other methods designed specifically to characterize protein complexes on chromatin are more suited to characterize transcription relevant complexes [28]. This procedural aspect explains why the network presented in Fig. 4 mainly contains factors involved in RNAP biogenesis. The network in Fig. 4 integrates high confidence AP-MS data obtained with 28 tagged proteins (coloured nodes, as opposed to grey nodes). Only interactions that obtained high interaction reliability (IR) scores are included, as we described previously [12].

Examination of this network not only reveals interactions made by subunits of all three nuclear RNAP, with shared subunits identified through a Venn diagram-like presentation, but also interactions connecting RNAP to other proteins, including the RPAP3/R2TP/PFDL, the chaperonin/CCT and the Integrator complexes. Not surprisingly, the RPAP3/R2TP/PFDL complex is itself connected to HSP90 as these proteins were shown to act in concert in RNAP II assembly [20]. Presence of the

chaperonin/CCT complex, which has previously been shown to play a central role in microtubule assembly, may be explained by our finding that microtubule assembly/integrity is required for RNAP II nuclear import [12]. The Integrator complex has been shown to interact with RNAP II and regulate snRNA processing [29]. Other RPAPs occupy a central position in this network.

In conclusion, a large network of factors associates with RNAP in the soluble cell fraction. Some of these factors, namely RPAP4/GPN1 and RPAP3, play a role in biogenesis of all three nuclear RNAPs. Additional work is required to define putative roles of other network components in assembly or nuclear import of these important molecular machines.

**Acknowledgments** We wish to thank members of our laboratory for helpful discussions. This work was supported by grants from the *Fonds de la recherche en santé du Québec* (FRSQ) and the Canadian Institutes for Health Research (CIHR).

## References

1. Conaway JW, Shilatifard A, Dvir A, Conaway RC. Control of elongation by RNA polymerase II. Trends Biochem Sci. 2000;25(8):375–80.
2. Coulombe B, Burton ZF. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. Microbiol Mol Biol Rev. 1999;63(2):457–78.
3. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev. 1998;62(2):465–503.
4. Kornberg RD. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci USA. 2007;104(32):12955–61.
5. Dey A, Nishiyama A, Karpova T, McNally J, Ozato K. Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. Mol Biol Cell. 2009;20(23):4899–909.
6. Parsons GG, Spencer CA. Mitotic repression of RNA polymerase II transcription is accompanied by release of transcription elongation complexes. Mol Cell Biol. 1997;17(10):5791–802.
7. Tseng LC, Chen RH. Temporal control of nuclear envelope assembly by phosphorylation of lamin B receptor. Mol Biol Cell. 2011;22(18):3306–17.
8. Czeko E, Seizl M, Augsberger C, Mielke T, Cramer P. Iwr1 directs RNA polymerase II nuclear import. Mol Cell. 2011;42(2):261–6.
9. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell. 2007;27(2):262–74.
10. Gras S, Chaumont V, Fernandez B, Carpentier P, Charrier-Savournin F, Schmitt S, et al. Structural insights into a new homodimeric self-activated GTPase family. EMBO Rep. 2007;8(6):569–75.
11. Carre C, Shiekhattar R. Human GTPases associate with RNA polymerase II to mediate its nuclear import. Mol Cell Biol. 2011;31(19):3953–62.
12. Forget D, Lacombe AA, Cloutier P, Al Khoury R, Bouchard A, Lavallee-Adam M, et al. The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. Mol Cell Proteomics. 2010;9(12):2827–39.
13. Reyes-Pardo H, Barbosa-Camacho AA, Perez-Mejia AE, Lara-Chacon B, Salas-Estrada LA, Robledo-Rivera AY, et al. A nuclear export sequence in GPN-loop GTPase 1, an essential

protein for nuclear targeting of RNA polymerase II, is necessary and sufficient for nuclear export. Biochim Biophys Acta. 2012;1823(10):1756–66.
14. Cloutier P, Al Khoury R, Lavallee-Adam M, Faubert D, Jiang H, Poitras C, et al. High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes. Methods. 2009;48(4):381–6.
15. Gstaiger M, Luke B, Hess D, Oakeley EJ, Wirbelauer C, Blondel M, et al. Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. Science. 2003;302(5648): 1208–12.
16. Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc Natl Acad Sci USA. 2008;105(5):1454–9.
17. Martin-Benito J, Gomez-Reino J, Stirling PC, Lundin VF, Gomez-Puertas P, Boskovic J, et al. Divergent substrate-binding mechanisms reveal an evolutionary specialization of eukaryotic prefoldin compared to its archaeal counterpart. Structure. 2007;15(1):101–10.
18. Simons CT, Staes A, Rommelaere H, Ampe C, Lewis SA, Cowan NJ. Selective contribution of eukaryotic prefoldin subunits to actin and tubulin binding. J Biol Chem. 2004;279(6): 4193–203.
19. McKeegan KS, Debieux CM, Watkins NJ. Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. Mol Cell Biol. 2009;29(18):4971–81.
20. Boulon S, Pradet-Balade B, Verheggen C, Molle D, Boireau S, Georgieva M, et al. HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. Mol Cell. 2010;39(6):912–24.
21. Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, et al. A kinetic framework for a mammalian RNA polymerase in vivo. Science. 2002;298(5598):1623–6.
22. Hardeland U, Hurt E. Coordinated nuclear import of RNA polymerase III subunits. Traffic. 2006;7(4):465–73.
23. Wild T, Cramer P. Biogenesis of multisubunit RNA polymerases. Trends Biochem Sci. 2012;37(3):99–105.
24. Archambault J, Friesen JD. Genetics of eukaryotic RNA polymerases I, II, and III. Microbiol Rev. 1993;57(3):703–24.
25. Young RA. RNA polymerase II. Annu Rev Biochem. 1991;60:689–715.
26. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol. 2007;8(8):645–54.
27. Jeronimo C, Langelier MF, Zeghouf M, Cojocaru M, Bergeron D, Baali D, et al. RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. Mol Cell Biol. 2004;24(16):7043–58.
28. Lavallee-Adam M, Rousseau J, Domecq C, Bouchard A, Forget D, Faubert D, et al. Discovery of cell compartment specific protein-protein interactions using affinity purification combined with tandem mass spectrometry. J Proteome Res. 2013;12(1):272–81.
29. Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R. Integrator, a multi-protein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 2005;123(2):265–76.

# Targeting Chromatin Modifying Enzymes in Anticancer Drug Discovery

**Victoria M. Richon, Mikel P. Moyer, and Robert A. Copeland**

**Abstract** Over the past decade the sequencing of human cancer genomes has provided a wealth of information on recurrent genetic alterations in specific subsets of cancers. The understanding of the underlying genetic alterations responsible for oncogenesis in these cancers has led to the successful development of new therapies specifically targeting the genetic alterations. The successes so far have come mainly from targeting genetic alterations in kinases. For example, the identification of mutant V600E RAF in melanoma led to the development of the vemurafenib (a BRAF inhibitor) for the subset of melanoma patients containing this mutation (Bollag et al Nat Rev Drug Discov 11(11):873–76, 2012). Likewise, the identification of the EML4-ALK translocation in a subset of non-small cell lung cancer patients led to approval of crizotinib (an ALK inhibitor) in this patient subset (Ou et al Oncologist 17(11):1351–75, 2012). In addition to the identification of kinase driver mutations, these genomic analyses have also identified chromatin modifying enzymes, specifically, enzymes involving protein methylation, as some of the most frequently observed somatic alterations in cancer. This chapter will focus on the protein methyltransferase class of chromatin modifying enzymes, providing the basis for their emergence as high priority targets for cancer drug discovery and the progress made in the development of inhibitors against this class of targets.

**Keywords** Epigenetics • Chromatin • Histone • Methylation • Methyltransferase • Cancer • Genetic alterations • Inhibitor

---

V.M. Richon, Ph.D.
Oncology Discovery and Preclinical Sciences, Global Oncology Division,
Sanofi, Cambridge, MA, USA

M.P. Moyer, Ph.D. • R.A. Copeland, Ph.D. (✉)
R&D Department, Epizyme, Inc., 400 Technology Square, 4th Floor,
Cambridge, MA 02139, USA
e-mail: rcopeland@epizyme.com

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_13,
© Springer Science+Business Media New York 2014

## Introduction

Over the past decade the sequencing of human cancer genomes has provided a wealth of information on recurrent genetic alterations in specific subsets of cancers. The understanding of the underlying genetic alterations responsible for oncogenesis in these cancers has led to the successful development of new therapies specifically targeting the genetic alterations. The successes so far have come mainly from targeting genetic alterations in kinases. For example, the identification of mutant V600E RAF in melanoma led to the development of the vemurafenib (a BRAF inhibitor) for the subset of melanoma patients containing this mutation [1]. Likewise, the identification of the EML4-ALK translocation in a subset of non-small cell lung cancer patients led to approval of crizotinib (an ALK inhibitor) in this patient subset [2]. In addition to the identification of kinase driver mutations, these genomic analyses have also identified chromatin modifying enzymes, specifically, enzymes involving protein methylation, as some of the most frequently observed somatic alterations in cancer. This chapter will focus on the protein methyltransferase (PMT) class of chromatin modifying enzymes, providing the basis for their emergence as high priority targets for cancer drug discovery and the progress made in the development of inhibitors against this class of targets.

Histones, nucleosomes, and the chromatin structures assembled from them function as integral components of the transcriptional machinery that determines cellular identity and fate. The nucleosome is the fundamental repeating chromatin structure within the cell and consists of approximately 146 base pairs of DNA wrapped around a protein octamer containing two molecules each of the core histones (H2A, H2B, H3, and H4). Many studies have identified a myriad of posttranslational modifications to the core histone amino terminal tails that play important regulatory roles in transcription. These modifications include lysine acetylation, methylation, ubiquitinylation and crotinylation, arginine methylation, and serine/threonine phosphorylation [3–5]. The precise location of the modification and combination of modifications are associated with specific states of the chromatin. For example, histone H3 lysine 4 trimethylation (H3K4me3) is commonly associated with open, transcriptionally active euchromatin [6, 7], while trimethylation of histone H3 lysine 9 (H3K9me3) is associated with transcriptionally silent areas of chromatin, referred to as heterochromatin [8]. In the last decade many researchers have worked to identify the pattern of histone modifications, correlate these modifications to specific states of the chromatin, and identify the enzymes that catalyze the addition or removal of the modification as well as the specific proteins that recognize these modifications (Fig. 1).

This chapter will focus on one family of histone modifying enzymes, namely the PMTs and their potential as a new class of therapeutic targets [9, 10]. While the demonstration of lysine and arginine methylation on histones occurred several decades ago, the identification of the enzymes that catalyze these reactions occurred only within the past decade. Today, over 60 enzymes have been described that have been shown to possess either lysine or arginine methyltransferase

**Fig. 1** Histone methylation sites

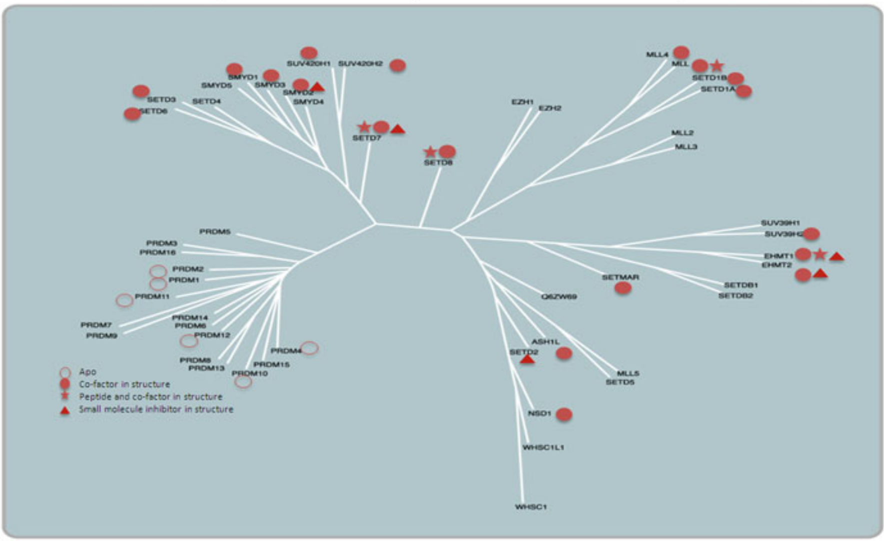



**Fig. 2** Protein lysine methyltransferases (PKMTs)

activity [10, 11]; these enzymes can be divided into two structurally related families based roughly on their substrate preference. One family is the SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain containing lysine methyltransferases that consists of 51 proteins that cluster into four major branches (Fig. 2) [11]. Within each of these branches there is at least one protein that has been demonstrated to possess lysine methyltransferase activity. DOT1L is the only lysine methyltransferase that does not share the canonical SET domain active site. Rather, the catalytic domain of DOT1L is more closely related to that of the arginine methyltransferases. Hence, the arginine methyltransferase family contains DOT1L along with 41 additional proteins. To date 11 proteins have been described to possess arginine methyltransferase activity [12] and an additional 30 enzymes that bear structural homology with the catalytic domain of the arginine

**Protein arginine methyltransferases (PRMTs)**

**Fig. 3** Protein arginine methyltransferases (PRMTs)

methyltransferases and DOT1L (Fig. 3) [11]. These additional 30 enzymes fall into two major groups annotated as the METTL and NSUN proteins. Rather than arginine methyltransferases these enzymes are annotated as RNA methyltransferases [13] and N-terminal protein methyltransferases [14].

Originally, histone lysine and arginine methylation was believed to be highly stable with turnover only occurring through new histone synthesis and histone exchange or incorporation during DNA replication. One rationale for this hypothesis was the inability to identify enzymes that catalyzed histone demethylation. Researchers have recently identified two major classes of enzymes that demethylate lysine residues, namely the amine oxidases, KDM1A and KDM1B, and the jumonjiC domain containing proteins which constitute a family of approximately 27 proteins [10, 15]. These findings indicate that at least lysine methylation is a reversible modification and is specifically regulated. No arginine demethylases have been reported although enzymes (arginine deiminases) that convert methylated arginines to citrulline have been described [16].

## Genetic Alterations in Protein Methylation Pathways

Sequencing of primary human tumors has identified mutations, translocations, and amplifications in protein methyltransferases, demethylases, as well as the histones themselves. The mutations found in specific protein methyltransferases take the form of both gain of function and loss of function and appear to be cell type specific. These findings indicate that the protein methyltransferases, like the kinases, constitute new therapeutic targets for specific cancer indications that bear the genetic

alteration. There are several direct and indirect mechanisms by which mutations within particular proteins can confer a unique dependency on PMT activity to specific cancer cells. For example, many studies have now shown that the gain of function alterations directly in protein methyltransferases create dependencies on the catalytic activity of that protein methyltransferase. Alternatively, in other cancer subtypes genetic alterations in one protein methyltransferase creates unique dependencies on a distinct protein methyltransferase. Finally, mutations arising in other members of chromatin modifying enzymes have also been shown to create dependency on a specific protein methyltransferase. These results provide support for the development of inhibitors as new cancer therapies and are described in the following sections.

## Genetic Alterations in Protein Methyltransferases

### *EZH2 and Histone H3K27 Methylation*

EZH2 is the catalytic subunit of the multiprotein polycomb repressive complex (PRC2) that catalyzes the mono-, di-, and tri-methylation of lysine 27 of histone H3 (H3K27) [17, 18]. H3K27me3 is associated with transcriptionally silent chromatin and is also associated with poised genes when found in combination with trimethylation of lysine 4 of histone H3 (H3K4me3) [19]. PRC2 plays important roles in cell fate, pluripotency, differentiation, and development [18]. EZH2 overexpression has been observed in many cancer types [20] as a result of either amplification [21] or overexpression [22, 23]. The overexpression of EZH2 has been associated with poor prognosis in breast [23] and prostate cancer [22] and melanoma [24]. Additionally, increased H3K27 trimethylation activity has been observed in non-Hodgkin lymphomas due to change of function mutations in the catalytic domain [25–27]. Diffuse large B cell lymphoma cell lines bearing EZH2 Y641 or A677 mutations are highly sensitive to the antiproliferative effects of small molecule inhibitors of EZH2 both in vitro [28, 29] and in vivo [29], indicating a dependency on the catalytic activity of EZH2.

In addition to gain of function alterations in EZH2, additional alterations in enzymes regulating H3K27 methylation have been described. For example, inactivating mutations of the H3K27me3 demethylase, KDM6A (also known as UTX), are found in multiple cancer types including 10 % of myelomas, 8 % of esophageal squamous cell carcinomas, and 1 % renal cell carcinomas [30]. Surprisingly, in this study no correlation was found with H3K27me3 levels and KDM6A mutation status. KDM6A mutations have also been described in subgroups 3 and 4 of medulloblastoma [31]. Both human subgroups 3 and 4 tumors and a mouse model reflecting human subgroup 3 medulloblastoma demonstrate high levels of H3K27me3. This high level of H3K27me3 due to either loss of KDM6A or high level expression of EZH2 in these tumors may indicate a dependency on EZH2 activity. Hence, EZH2 inhibitors may provide a new therapeutic potential for these poor prognosis subtypes of medulloblastoma.

Inactivating mutations in the ATP-dependent chromatin modifying complex component SMARCB1 (also known as SNF) have been described in a variety of human tumors including malignant rhabdoid tumors [32, 33]. SMARCB1 is a component of the SWI/SNF complex and this complex antagonizes PRC2-mediated gene silencing [34]. Mouse models of SMARCB1 loss result in highly penetrant tumor formation and elevated levels of EZH2 [35]. The elevated levels of EZH2 are functionally important in that the tumors that arise due to SMARCB1 loss require EZH2. A conditional knockout of EZH2 in the SMARCB1 mouse model blocks tumor formation. These results indicate that SMARCB1 deficient tumors may be responsive to therapeutic inhibition of EZH2 activity and highlight the complex interdependencies between chromatin modifying and remodeling activities in cancer.

Loss of function mutations of EZH2 leading to low levels of H3K27me3 have been observed in approximately 10 % of patients with myelodysplastic syndrome and myeloproliferative neoplasm, and these mutations are associated with a poor prognosis [36–38]. Loss of function mutations have also been observed in other components of the PRC2 complex (e.g., SUZ12) suggesting an important role of this complex and H3K27 methylation in the development of these disorders [39].

Recently, mutations in histone H3 at K27 have been identified in pediatric gliomas [40, 41]. The mutations result in a change in lysine 27 to methionine of histone H3. The H3K27 mutation is limited to two histone H3 variants, *H3F3A* (histone H3.3) [40, 41] and *HIST1H3B* (histone H3.1) [40]. These mutations are found in approximately 30 % of pediatric glioblastoma [41] and 60 % of diffuse intrinsic pontine glioma [40] but not in other non-brainstem glioblastomas or other types of gliomas. Mutations in glycine 34 (G34V) of histone H3.3 were also described in pediatric glioblastomas [41]. This is the first description of mutations of histones in cancer, and the functional significance of these mutations is yet to be determined.

## *DOT1L and MLL*

DOT1L is a histone H3K79 mono-, di-, and tri-methyltransferase [42]. It is the only PMT identified to catalyze methylation of this site, and no histone demethylase for this modification has been identified. Methylation of H3K79 is associated with transcriptionally active chromatin. Studies undertaken to identify DOT1L interacting proteins revealed that DOT1L is associated with AF10 [43]. This observation defined a connection between DOT1L and leukemogenesis, as AF10 is a translocation partner of the MLL histone H3K4 methyltransferase in MLL-rearranged leukemia. The 11q23 MLL translocation occurs in approximately 5 % of acute lymphocytic leukemia and 5–10 % of acute myeloid leukemia. Over 80 translocation partners with MLL have been described and over 80 % of the partners are found in a complex with DOT1L [44, 45]. The MLL fusion protein retains the N-terminal region required for correct gene targeting but lacks the methyltransferase domain in the MLL protein. DOT1L is thus now recruited by the MLL fusion protein to gene

locations under the regulation of MLL, leading to aberrant H3K79 methylation and resultant expression of a subset of MLL target genes [46–48]. These studies support a role for DOT1L in MLL-rearranged leukemia and provide a new therapeutic target for this poor prognosis subset of AML and ALL. To this end, potent small molecule DOT1L inhibitors have been described that selectively inhibit MLL-rearranged leukemia in cells in culture and a human xenograft mouse model of MLL-rearranged leukemia [49, 50].

In addition to the translocations in leukemias, mutations in related MLL proteins have been described in various other cancers, including small cell lung cancer [51], multiple myeloma [52], diffuse large B cell lymphoma [53, 54], and follicular lymphomas [54]. Additional studies are required to determine the functional significance of these mutations in the development of these cancers.

## *The NSD Branch*

WHSC1 (also known as NSD2) is a protein methyltransferase that specifically catalyzes the methylation of histone H3K36, and this methylation is associated with active transcription [55]. WHSC1 is overexpressed in many human cancers [56], most notably in multiple myeloma (MM) [57]. A hallmark of B cell malignancies, including myeloma, is increased expression of oncogenes by translocation to an immunoglobulin locus [58]. The most common translocation is t(4;14)(p16;q32) occurring in approximately 15 % of MM patients [59]. The t(4;14) remains one of the worst prognostic subgroups in MM and these patients are in need of new therapies. The t(4;14) translocation results in overexpression of both WHSC1 and fibroblast growth factor receptor 3 (FGRF3) [57]. FGFR3 expression is lost in approximately 30 % of t(4;14) patients, while WHSC1 expression is maintained in 100 % of patients.

Several groups have performed experiments that support a pathogenic role for WHSC1 in t(4;14) myeloma patients [60, 61]. Genetic knockdown of WHSC1 or disruption of the translocated allele in t(4;14) myeloma cells results in inhibition of cell proliferation and tumorigenicity. Enforced expression of WHSC1 to a t(4;14) cell line from which the overexpressed gene was genetically disrupted increases cell growth [62]. Additionally, the high level of WHSC1 expression detected in t(4;14) myeloma cells is associated with an increase in global levels of histone H3K36 dimethylation [62, 63]. These results provide the rationale for developing selective WHSC1 inhibitors as a specific therapeutic for this poor prognosis subtype defined by the t(4;14) translocation.

In addition to WHSC1, two additional related methyltransferases, NSD1 and WHSC1L1(NSD3), catalyze methylation of histone H3K36, and these enzymes are also genetically altered in defined cancer subtypes. NSD1 is translocated in a subset of acute myelogenous leukemias [64] and the expression of the translocation product, NUP98-NSD1, is sufficient to promote leukemogenesis [65]. WHSC1L1 is amplified and overexpressed in breast [66] and lung [67] cancers, and rare NUP98

translocations with WHSC1L1 have been described in acute myeloid leukemia patients [68].

## Arginine Methyltransferases

Compared to the protein lysine methyltransferases, very few examples of genetic alterations in protein arginine methyltransferases have been identified in cancer. Several arginine methyltransferases (RMTs) have been described to be overexpressed in different cancer subtypes and to be required for the transformed phenotypes. Also, a number of RMTs methylate substrates beyond histones that may play significant roles in cancer development. For example, PRMT5 is overexpressed in mantle cell lymphoma, and this overexpression is associated with high levels of

**Table 1** Selected examples of PMT inhibitors

| Compound name | Compound structure | Primary target potency | Comments |
|---|---|---|---|
| SAH | | Product of reaction for all PMTs. $IC_{50}$s range from 0.1 to 20 μM | Nonselective; Competitive with SAM |
| Sinefungin | | Natural product analog of SAM/SAH. $IC_{50}$s range from 0.1 to 20 μM | Nonselective; Competitive with SAM |
| DZNep | | | Inhibits SAH hydrolase—consequently nonselective |
| Tokyo Medical and Dental University compound 1c | | SET7/9 inhibitor of $IC_{50}$ 10 μM | Presumed to be competitive with SAM |
| EPZ004777 | | DOT1L inhibitor of $K_i$ 0.3 nM | >1,200-fold selectivity against other PMTs; competitive with SAM |
| EPZ005687 | | EZH2 inhibitor of $K_i$ 24 nM | Competitive with SAM; >500× selective versus a panel of 13 other PMTs and 50× selective versus the closely related EZH1 |

(continued)

**Table 1** (continued)

| Compound name | Compound structure | Primary target potency | Comments |
|---|---|---|---|
| GSK126 | | EZH2 inhibitor of $K_i^{app} = 0.5–3$ nM | Competitive with SAM; >1,000× selective versus a panel of 20 other PMTs and 150× selective versus EZH1 |
| BIX-01294 | | Inhibitor of G9a (IC50 = 0.2–1.7 μM) and GLP (IC50 = 0.03–38 μM) | Uncompetitive with SAM |
| UNC-0638 | | $K_i$ 2.5 nM; G9a inhibitor ($IC_{50}$ <15 nM) | Noncompetitive with SAM |
| AZ505 | | SMYD2 inhibitor of $IC_{50}$ 0.12 μM | Uncompetitive with SAM |
| Methylgene compound | | CARM1 inhibitor of $IC_{50}$ 60 nM | Noncompetitive with SAM |
| BMS compound | | CARM1 inhibitor of $IC_{50}$ 40 nM | Noncompetitive with SAM |
| Compound 5 | | PRMT1 inhibitor of $IC_{50}$ 55 μM | Presumed to be a bisubstrate inhibitor |

methylation of its histone substrate, H3R8 [69]. The tumor suppressor p53 has also been described as a substrate for PRMT5; arginine methylation of p53 alters the p53 response to DNA damage [70]. Other arginine methyltransferases that are associated with the development of transformed phenotype include PRMT1 and CARM1.

## Direct Inhibitors of PMTs

The growing body of data suggesting that genetic alterations in PMTs drive tumorigenesis in a number of human cancers has generated interest in identifying selective inhibitors of these enzymes [9, 71]. The identification of such selective agents

against these driver alterations that confer to a cancer cell a unique dependence on the enzymatic activity of the PMT may provide a basis for clinical benefit with a reasonable therapeutic index.

Significant progress has been made toward the identification of small molecule, drug-like, and selective inhibitors of PMTs. The earliest reports of small molecule inhibitors of PMTs were not particularly promising. These compounds were either nonspecific SAM analogues, such as S-adenosylhomocysteine (SAH), the universal product of SAM hydrolysis or sinefungin (Table 1), or indirect inhibitors, such as 3-deazaneplanocin (DZNep) which inhibits the enzyme SAH hydrolase and thereby increases cellular levels of SAH resulting in non-selective inhibition of PMT (and other SAM-utilizing enzymes) activity [72].

Since these inauspicious beginnings, significant progress toward diverse, selective, and drug-like inhibitors has been made. Table 1 contains representative inhibitors across a number of PMTs with their reported potency against their primary PMT targets as well as information on how these inhibitors bind the enzyme. These compounds generally either compete with SAM for its binding pocket or compete with the substrate peptide at the methyl-accepting amino acid binding pocket.

Aminonucleoside analogs of SAM have been reported [49, 73, 74] that are competitive with SAM and display a range of target affinity and selectivity. The compound EPZ004777, for example, is a 300 pM, SAM-competitive inhibitor of DOT1L that demonstrates >1,200-fold selectivity for this enzyme over all other tested PMTs [49]. The high affinity of EPZ004777 and related analogues derives from a conformational adaptation mechanism in which the enzyme changes conformation to close down the ligand binding pocket around the inhibitor [75]. This mechanism drives compound affinity largely by reducing the rate of dissociation for the enzyme–inhibitor binary complex, thus resulting in very long drug-target residence time [71, 76].

EPZ004777 was found to selectively kill *MLL*-rearranged leukemia cells in vitro with little antiproliferative effect on non-rearranged cells, a finding consistent with the notion that the chromosomal translocation confers a unique dependence on DOT1L enzymatic activity in *MLL*-rearranged leukemia. Downstream effects of DOT1L inhibition were also observed—namely, a decrease in cellular histone H3K79 methylation followed by a decrease in MLL fusion target gene expression and induction of apoptosis in *MLL*-rearranged leukemia cell lines. The induction of apoptosis is preceded by accumulation of cells in the G1 phase of the cell cycle and gene expression changes consistent with hematopoietic differentiation [49]. Additionally, EPZ004777 is the first PMT inhibitor to demonstrate anticancer effects and survival benefit in an animal model of *MLL*-rearranged leukemia [49] (vide infra).

The aminonucleoside inhibitors of DOT1L and SET7/9 act in a competitive manner with respect to SAM, a fact that is not surprising given their clear structural resemblance to SAM and SAH. Interestingly, other SAM-competitive inhibitors of PMTs have been recently reported that bear no chemical resemblance to the nucleoside substrate. For example, the indazole EPZ005687 is a potent and selective SAM-competitive inhibitor of EZH2-containing PRC2 that has been shown to inhibit wild-type and NHL-associated mutant EZH2 with nanomolar affinity [77, 78]. Although this compound inhibits both wild-type and mutant EZH2, one would

expect the mutant-bearing NHL cells would have a unique dependence on EZH2 activity for proliferation, such that compounds like EPZ005687 would demonstrate selective killing of these cells. Indeed, EPZ005687 has been shown to selectively kill NHL cells that are heterozygous for EZH2 mutations at either Y641 or A677G, with limited impact on the growth of homozygous wild-type EZH2 containing NHL cells [77]. Compounds of similar structure, potency, and EZH2 selectivity have also been reported in a set of publications and patent applications from the group at GlaxoSmithKline [29, 79–82] (one of these compounds, with reported in vivo activity, is shown in Table 1).

In contrast to the SAM-competitive mode of inhibition, a large number of PMT inhibitors (for both PKMT and PRMT targets) act by binding not to the SAM pocket but instead to the protein-substrate binding site and engaging recognition elements within the amino acid channel. For example, the compound BIX-01294 was among the first PMT inhibitors to be reported and is a potent and selective inhibitor of the SET-domain PKMT EHMT2 (also known as G9a) [83]. Structural analogues of this compound were subsequently designed as dual EHMT1 and EHMT2 inhibitors with much greater target affinity and cell permeability. The compound UNC-0638, for example, was shown to reduce levels of H3K9 methylation in MDA-MB231 cells with an $IC_{50}$ of 81 nM [84]. Likewise, the compound AZ505 has been shown crystallographically to bind to the SET-domain PKMT SMYD2 within the lysine binding channel [85]. Similarly, the structurally related compounds from the Methylgene and Bristol-Myers Squibb groups are nanomolar inhibitors of the PRMT CARM1 (also known as PRMT4) that have been shown crystallographically to bind within the arginine channel of this enzyme [86–88].

A final modality of enzyme interaction is exemplified by the PRMT1 inhibitor Compound 5, which is presumed to act as a bisubstrate inhibitor, engaging recognition elements both within the SAM and arginine binding pockets. This compound is a relatively modest inhibitor of PRMT1, but demonstrates cellular activity in reducing methylation in HepG2 cells [89]. Hence, this may be an interesting starting point for further compound optimization.

## Structural Studies of PMTs

Considerable research has been conducted on the structure and function of protein methyltransferase enzymes that has been summarized in a number of publications [90–94]. The current status of these efforts is illustrated in Figs. 2 and 3. This research has produced an understanding of the critical structural features necessary for methyl transfer [95–97] and of the determinants of the multiplicity of methylation [95]. Importantly for drug discovery, a number of protein structures with bound inhibitors have been solved, opening the potential for structure-based drug design. Conformational flexibility of PMTs is an important consideration in the design of potent, selective inhibitors—and can be manifested in unexpected ways. Figure 4 shows an example of a potent DOT1L inhibitor, EPZ004777, bound to the DOT1L

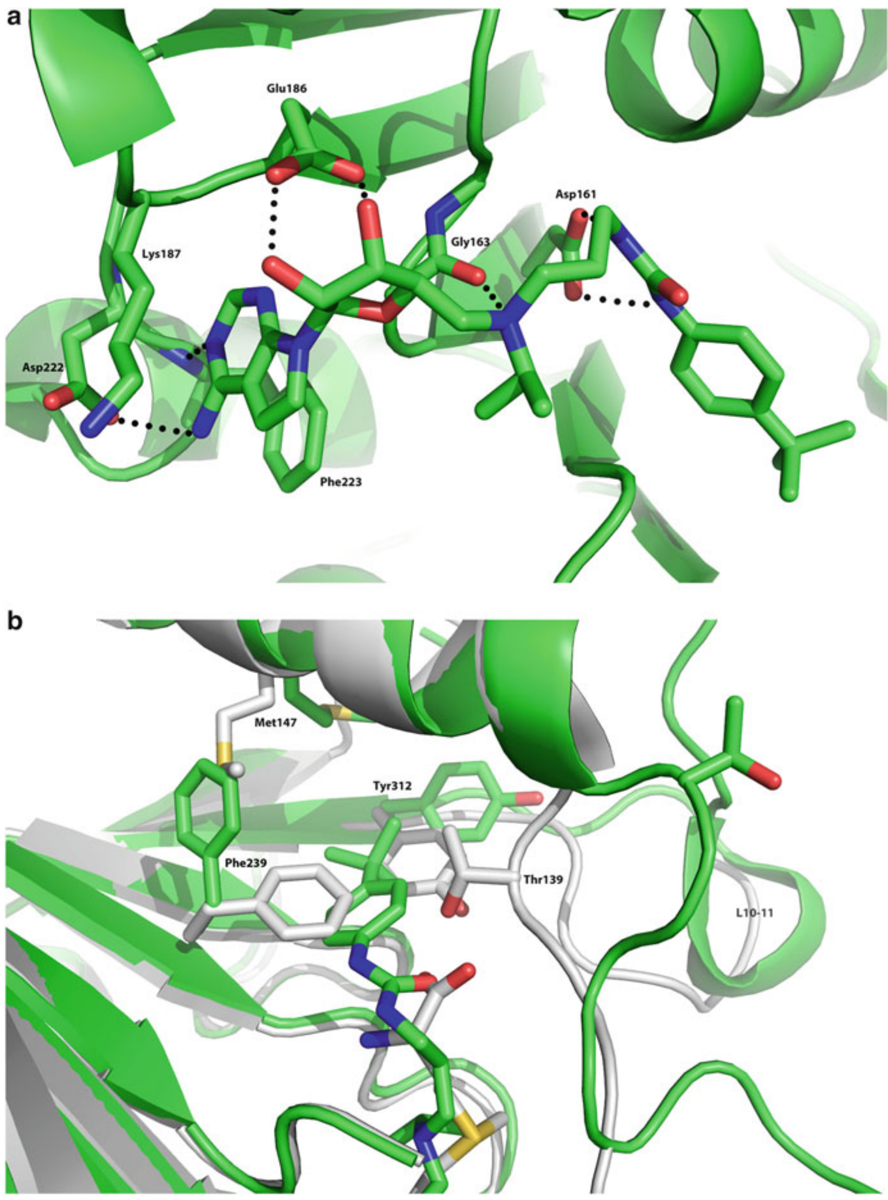



**Fig. 4** (**a**) Key interactions between the Dot1L protein and EPZ004777 are shown in the crystal structure of the complex (*green*; PDB 4EKI). Hydrogen bonds are indicated with *dotted lines*. (**b**) EPZ004777 binding opens a hydrophobic pocket in the Dot1L protein (*green*) not seen in the Dot1L-SAM complex (*grey*, PDB 3QOW). Key residues that move to accommodate compound binding are labeled. (**c**) Ribbon diagram of the extensive structural rearrangement seen in the Dot1L-EPZ004777 protein structure (*green*) around the binding site when compared to the Dot1L structure with SAM (*grey*)

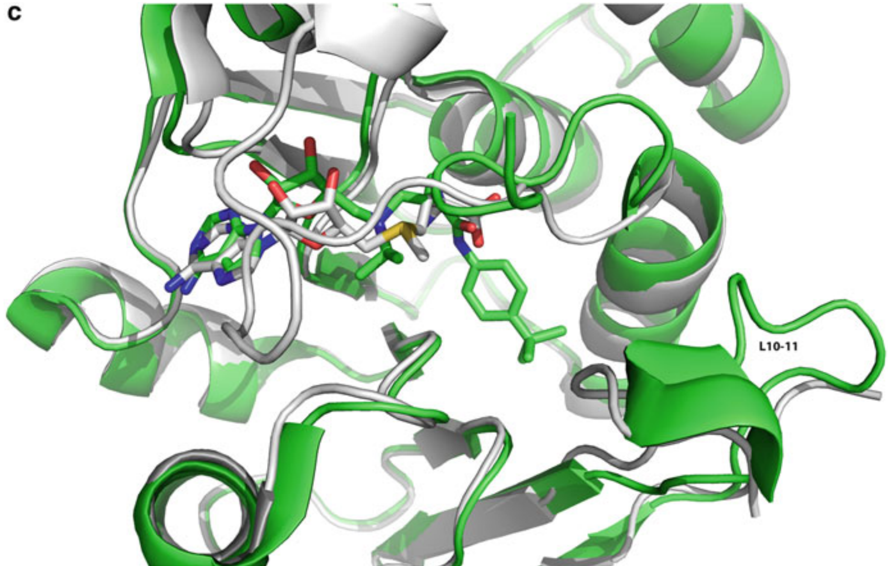



**Fig. 4** (continued)

enzyme [75]. Instead of the extended tether and terminal hydrophobic group reaching into the proximal lysine binding channel as expected, a novel binding pocket is exposed. The tert-butyl phenyl group opens up the novel hydrophobic pocket by changing the side chain conformation of Phe239, Tyr312, Met147, and Thr139. This also induces additional changes in two loops of the protein as shown in Fig. 4c. Notwithstanding the challenges posed by protein flexibility, the potential impact of the structural information of PMTs on the discovery of novel therapeutics is quite large and will ultimately depend upon the degree of prospective design that is possible with this class of targets.

## Summary

Genomic analysis has identified enzymes involving protein methylation as some of the most frequently observed somatic alterations in cancer. The protein methyltransferase class of chromatin modifying enzymes has emerged as high priority targets because of a growing body of data suggesting that these alterations drive tumorigenesis in a number of human cancers. Progress has been made in discovering potent and selective inhibitors of a number of these enzymes—aided by a wealth of structural biology data. The most advanced inhibitors have recently begun human clinical trials, signaling the beginning of a very exciting time when the promise of these enzyme targets in the treatment of cancer is put to the test.

## References


1. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat Rev Drug Discov. 2012;11(11):873–86.
2. Ou SH, Bartlett CH, Mino-Kenudson M, Cui J, Iafrate AJ. Crizotinib for the treatment of ALK-rearranged non-small cell lung cancer: a success story to usher in the second decade of molecular targeted therapy in oncology. Oncologist. 2012;17(11):1351–75.
3. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell. 2011;146(6): 1016–28.
4. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA. 1964;51:786–94.
5. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293(5532):1074–80.
6. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, et al. Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419(6905):407–11.
7. Sims 3rd RJ, Nishioka K, Reinberg D. Histone lysine methylation: a signature for chromatin function. Trends Genet. 2003;19(11):629–39.
8. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001;410(6824):116–20.
9. Copeland RA, Solomon ME, Richon VM. Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov. 2009;8(9):724–32.
10. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012;11(5):384–400.
11. Richon VM, Johnston D, Sneeringer CJ, Jin L, Majer CR, Elliston K, et al. Chemogenetic analysis of human protein methyltransferases. Chem Biol Drug Des. 2011;78(2):199–210.
12. Di Lorenzo A, Bedford MT. Histone arginine methylation. FEBS Lett. 2011;585(13): 2024–31.
13. Frye M, Watt FM. The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. Curr Biol. 2006;16(10):971–81.
14. Webb KJ, Lipson RS, Al-Hadid Q, Whitelegge JP, Clarke SG. Identification of protein N-terminal methyltransferases in yeast and humans. Biochemistry. 2010;49(25):5225–35.
15. Shi Y, Whetstine JR. Dynamic regulation of histone lysine methylation by demethylases. Mol Cell. 2007;25(1):1–14.
16. Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. Science. 2004;306(5694):279–83.
17. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science. 2002;298(5595):1039–43.
18. Cao R, Zhang Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr Opin Genet Dev. 2004;14(2):155–64.
19. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315–26.
20. Simon JA, Lange CA. Roles of the EZH2 histone methyltransferase in cancer epigenetics. Mutat Res. 2008;647(1–2):21–9.
21. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J. 2003;22(20): 5323–35.
22. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature. 2002;419(6907):624–9.
23. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci USA. 2003;100(20):11606–11.

24. Bachmann IM, Halvorsen OJ, Collett K, Stefansson IM, Straume O, Haukaas SA, et al. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. J Clin Oncol. 2006;24(2):268–73.
25. Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, et al. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci USA. 2010;107(49):20980–5.
26. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010;42(2):181–5.
27. Yap DB, Chu J, Berg T, Schapira M, Cheng SW, Moradian A, et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood. 2011;117(8):2451–9.
28. Knutson SK, Wigle TJ, Warholic NM, Sneeringer CJ, Allain CJ, Klaus CR, et al. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nat Chem Biol. 2012;8(11):890–6.
29. McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature. 2012;492(7427):108–12.
30. van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009;41(5):521–3.
31. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, et al. Novel mutations target distinct subgroups of medulloblastoma. Nature. 2012;488(7409):43–8.
32. Sevenet N, Sheridan E, Amram D, Schneider P, Handgretinger R, Delattre O. Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers. Am J Hum Genet. 1999;65(5):1342–8.
33. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011;11(7):481–92.
34. Kia SK, Gorski MM, Giannakopoulos S, Verrijzer CP. SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. Mol Cell Biol. 2008; 28(10):3457–64.
35. Wilson BG, Wang X, Shen X, McKenna ES, Lemieux ME, Cho YJ, et al. Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. Cancer Cell. 2010;18(4):316–28.
36. Nikoloski G, Langemeijer SM, Kuiper RP, Knops R, Massop M, Tonnissen ER, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat Genet. 2010;42(8):665–7.
37. Makishima H, Jankowska AM, Tiu RV, Szpurka H, Sugimoto Y, Hu Z, et al. Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. Leukemia. 2010;24(10):1799–804.
38. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet. 2010;42(8):722–6.
39. Score J, Hidalgo-Curtis C, Jones AV, Winkelmann N, Skinner A, Ward D, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood. 2012;119(5):1208–13.
40. Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet. 2012;44(3):251–3.
41. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482(7384):226–31.
42. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, et al. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol. 2002;12(12):1052–8.

43. Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, et al. hDOT1L links histone methylation to leukemogenesis. Cell. 2005;121(2):167–78.
44. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer. 2007;7(11):823–33.
45. Slany RK. The molecular biology of mixed lineage leukemia. Haematologica. 2009;94(7): 984–93.
46. Krivtsov AV, Feng Z, Lemieux ME, Faber J, Vempati S, Sinha AU, et al. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell. 2008;14(5):355–68.
47. Bernt KM, Zhu N, Sinha AU, Vempati S, Faber J, Krivtsov AV, et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell. 2011;20(1): 66–78.
48. Monroe SC, Jo SY, Sanders DS, Basrur V, Elenitoba-Johnson KS, Slany RK, et al. MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia. Exp Hematol. 2011;39(1):77–86.
49. Daigle SR, Olhava EJ, Therkelsen CA, Majer CR, Sneeringer CJ, Song J, et al. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. Cancer Cell. 2011;20(1):53–65.
50. Basavapathruni A, Jin L, Daigle SR, Majer CR, Therkelsen CA, Wigle TJ, et al. Conformational adaptation drives potent, selective and durable inhibition of the human protein methyltransferase DOT1L. Chem Biol Drug Des. 2012;80(6):971–80.
51. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet. 2012;44(10):1104–10.
52. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471(7339):467–72.
53. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci USA. 2012;109(10):3879–84.
54. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476(7360): 298–303.
55. Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury 3rd WJ, et al. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J Biol Chem. 2009;284(49):34283–95.
56. Hudlebusch HR, Skotte J, Santoni-Rugiu E, Zimling ZG, Lees MJ, Simon R, et al. MMSET is highly expressed and associated with aggressiveness in neuroblastoma. Cancer Res. 2011; 71(12):4226–35.
57. Chesi M, Nardini E, Lim RS, Smith KD, Kuehl WM, Bergsagel PL. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood. 1998;92(9):3025–34.
58. Chng WJ, Glebov O, Bergsagel PL, Kuehl WM. Genetic events in the pathogenesis of multiple myeloma. Best Pract Res Clin Haematol. 2007;20(4):571–96.
59. Keats JJ, Maxwell CA, Taylor BJ, Hendzel MJ, Chesi M, Bergsagel PL, et al. Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients. Blood. 2005;105(10):4060–9.
60. Lauring J, Abukhdeir AM, Konishi H, Garay JP, Gustin JP, Wang Q, et al. The multiple myeloma associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. Blood. 2008;111(2):856–64.
61. Marango J, Shimoyama M, Nishio H, Meyer JA, Min DJ, Sirulnik A, et al. The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. Blood. 2008;111(6):3145–54.
62. Martinez-Garcia E, Popovic R, Min DJ, Sweet SM, Thomas PM, Zamdborg L, et al. The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. Blood. 2011;117(1):211–20.

63. Kuo AJ, Cheung P, Chen K, Zee BM, Kioi M, Lauring J, et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol Cell. 2011;44(4):609–20.
64. Hollink IH, van den Heuvel-Eibrink MM, Arentsen-Peters ST, Pratcorona M, Abbas S, Kuipers JE, et al. NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. Blood. 2011;118(13):3645–56.
65. Wang GG, Cai L, Pasillas MP, Kamps MP. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat Cell Biol. 2007;9(7):804–12.
66. Angrand PO, Apiou F, Stewart AF, Dutrillaux B, Losson R, Chambon P. NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. Genomics. 2001;74(1):79–88.
67. Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y, et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci USA. 2005;102(27):9625–30.
68. Rosati R, La Starza R, Veronese A, Aventin A, Schwienbacher C, Vallespi T, et al. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). Blood. 2002;99(10):3857–60.
69. Pal S, Baiocchi RA, Byrd JC, Grever MR, Jacob ST, Sif S. Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. EMBO J. 2007; 26(15):3558–69.
70. Jansson M, Durant ST, Cho EC, Sheahan S, Edelmann M, Kessler B, et al. Arginine methylation regulates the p53 response. Nat Cell Biol. 2008;10(12):1431–9.
71. Copeland RA. Conformational adaptation in drug-target interactions and residence time. Future Med Chem. 2011;3(12):1491–501.
72. Chiang PK. Biological effects of inhibitors of S-adenosylhomocysteine hydrolase. Pharmacol Ther. 1998;77(2):115–34.
73. Mori S, Iwase K, Iwanami N, Tanaka Y, Kagechika H, Hirano T. Development of novel bisubstrate-type inhibitors of histone methyltransferase SET7/9. Bioorg Med Chem. 2010;18(23):8158–66.
74. Yao Y, Chen P, Diao J, Cheng G, Deng L, Anglin JL, et al. Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies. J Am Chem Soc. 2011;133(42):16746–9.
75. Basavapathruni A et al. Conformational adaption drives potent, selective and durable inhibition of the human protein methyltransferase DOT1L. Chem Biol Drug Des. 2012;80: 971–80.
76. Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. Nat Rev Drug Discov. 2006;5(9):730–9.
77. Knutson SK et al. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nat Chem Biol. 2012;8(11):890–6.
78. Copeland RA, et al. Patent Application WO2012034132. 2012.
79. Brackley J. Inventor WO20111403242011.
80. Duquenne C, et al. Inventor WO20111403252011.
81. Burgess J, et al. Inventor WO20120058052012.
82. Knight SD, et al. Inventor WO20120750802012.
83. Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, Teodoro ML, et al. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. Mol Cell. 2007;25(3):473–81.
84. Vedadi M, Barsyte-Lovejoy D, Liu F, Rival-Gervier S, Allali-Hassani A, Labrie V, et al. A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. Nat Chem Biol. 2011;7(8):566–74. doi:10.1038/nchembio.599.
85. Ferguson ALN, Howard T, Pollard H, Green I, Grande C, Cheung T, et al. Structural basis of substrate methylation and inhibition of SMYD2. Structure. 2011;19(9):1262–73.
86. Purandare AV, Chen Z, Huynh T, Pang S, Geng J, Vaccaro W, et al. Pyrazole inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. 2008;18(15):4438–41.

87. Allan M, Manku S, Therrien E, Nguyen N, Styhler S, Robert MF, et al. N-Benzyl-1-heteroaryl-3-(trifluoromethyl)-1H-pyrazole-5-carboxamides as inhibitors of co-activator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. 2009;19(4):1218–23.
88. Huynh T, Chen Z, Pang S, Geng J, Bandiera T, Bindi S, et al. Optimization of pyrazole inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). Bioorg Med Chem Lett. 2009;19(11):2924–7.
89. Spannhoff A, Machmur R, Heinke R, Trojer P, Bauer I, Brosch G, et al. A novel arginine methyltransferase inhibitor with cellular activity. Bioorg Med Chem Lett. 2007;17(15):4150–3.
90. Campagna-Slater V, Mok MW, Nguyen KT, Feher M, Najmanovich R, Schapira M. Structural chemistry of the histone methyltransferases cofactor binding site. J Chem Inf Model. 2011;51(3):612–23.
91. Volkel P, Angrand PO. The control of histone lysine methylation in epigenetic regulation. Biochimie. 2007;89(1):1–20.
92. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol. 2005;6(8):227.
93. Qian C, Wang X, Manzur K, Sachchidanand, Farooq A, Zeng L, et al. Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. J Mol Biol. 2006;359(1):86–96.
94. Schapira M. Structural chemistry of human SET domain protein methyltransferases. Curr Chem Genomics. 2011;5 Suppl 1:85–94.
95. Zhang X, Cheng X. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. Structure. 2003;11(5):509–20.
96. Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, et al. Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature. 2003;421(6923):652–6.
97. Trievel RC, Flynn EM, Houtz RL, Hurley JH. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. Nat Struct Biol. 2003;10(7):545–52.

# Genome Organization in Cancer Cells

**Harry Yu Zhou and Jennifer A. Mitchell**

**Abstract** Tumor cells display dramatic changes in gene expression compared to the normal cells in the surrounding tissue. These changes are often due to chromosomal rearrangements or somatic mutations which result in the altered expression or function of proteins that regulate transcription. Indeed, recurrent chromosomal translocations are hallmarks of human cancers and vary between different types of cancer. Recent studies have implicated genome organization and the frequency of DNA double strand breaks as important factors contributing to the formation of specific chromosomal translocations. It has also become increasingly clear that nonrandom organization of the genome regulates gene expression in a cell type-specific manner. Single nucleotide changes in intergenic regions of the genome have been shown to affect gene regulation through the formation of chromatin loops. Therefore, understanding the interplay between genome organization and transcription in normal cell types and its perturbations in cancer will provide important insights into the causes of tumorigenesis.

**Keywords** Chromatin • Translocation • Transcription factor • Regulatory element • Chromatin loop • Fluorescence in situ hybridization • Chromosome conformation capture

## List of Abbreviations

| | |
|---|---|
| AID | Activation-induced cytidine deaminase |
| DSBs | DNA double-strand breaks |
| FISH | Fluorescence in situ hybridization |

H.Y. Zhou, B.Sc. • J.A. Mitchell, Ph.D. (✉)
Department of Cell and Systems Biology, University of Toronto, 25 Harbord Street, Toronto, ON, Canada, M5S 3G5
e-mail: harry.zhou@mail.utoronto.ca; ja.mitchell@utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_14,
© Springer Science+Business Media New York 2014

| | |
|---|---|
| GWAS | Genome-wide association studies |
| H3K4me3 | Histone H3 lysine 4 trimethylation |
| H3K9ac | Histone H3 lysine 9 acetylation |
| H3K9me2 | Histone H3 lysine 9 dimethylation |
| H3K9me3 | Histone H3 lysine 9 trimethylation |
| H4K20me3 | Histone H4 lysine 20 trimethylation |
| Ig | Immunoglobulin locus |
| ISH | In situ hybridisation |
| RNAPI | RNA polymerase I |
| RNAPII | RNA polymerase II |
| RNAPIII | RNA polymerase III |
| rRNA | Ribosomal RNA |
| SNPs | Single nucleotide polymorphisms |
| TCR | T-cell receptor locus |

## Introduction

Although cancers are diverse, they all involve dramatic changes in gene expression in the cancer cell compared to the normal cell type from which it arose. These dramatic changes in gene expression are often due to changes in expression or function of proteins that regulate gene expression. Indeed, the most frequently mutated gene in human cancer cells is the tumor suppressor gene *TP53*, which encodes the p53 protein, a transcription factor usually expressed at low levels in normal cells [1]. The p53 protein binds specific DNA sequences throughout the genome and regulates transcription of numerous target genes [2]. When p53 is induced in normal cells, it can inhibit cell cycle progression, promote cellular senescence, or induce apoptosis [1]. Conversely, p53 mutations in cancer cells can promote cell proliferation and survival as well as disruption of normal tissue architecture and metastasis, depending on the specific mutation and the cellular context. Cancer-causing mutations in p53 are thought to both abrogate its wild-type protective functions and impart gain of function phenotypes [1]. While mutations in p53 are found in more than half of all tumors, many other transcription factors appear to have a role in cancer cells where they are either mutated or aberrantly expressed. Changes in transcription factor function can be caused by mutations within the gene itself or chromosomal rearrangements that fuse two genes creating a novel fusion protein. Similarly, changes in transcription factor expression can be caused by single nucleotide changes in intergenic regulatory regions or translocation events that expose the gene to a different set of regulatory elements. To understand how changes in the genome sequence lead to changes in gene expression observed in cancer cells, we must understand transcriptional regulation of gene expression in the context of the nuclear space. Probing nuclear architecture has changed the way we view genome organization in the nucleus, the processes of transcription and replication, and contributed greatly to our understanding of genome function.

## Principles of Nuclear and Genome Organization

The nucleus is specialized to store, replicate, and transcribe the genome, as well as allow for response to external signals. It is a highly organized organelle that contains numerous protein compartments that segregate specific enzymatic activities from each other. The nuclear envelope encapsulates the nuclear contents with the nuclear lamina lining the inside of the inner nuclear membrane. Nuclear pores span the double membrane of the nuclear envelope and control the entrance and exit of proteins containing nuclear import or export signals while allowing for smaller macromolecules such as mRNAs to exit to the cytoplasm for translation.

Transcription is spatially organized in the nucleus with separate locations for transcription carried out by the three nuclear polymerase complexes, RNA polymerase I (RNAPI), RNA polymerase II (RNAPII), and RNA polymerase III (RNAPIII) [3–5]. The nucleolus, the largest nuclear compartment, is the location of RNAPI activity where ribosomal RNA (rRNA) genes, located in arrays on several chromosomes, are transcribed. In the nucleolus rRNA is also processed and assembled with ribosomal proteins [6]. This spatial organization of transcriptional activity is not limited to RNAPI as RNAPII and RNAPIII are also segregated in the nucleus into smaller nucleoplasmic foci. While cells have between 1 and about 25 nucleoli, hundreds to thousands of smaller foci containing RNAPII proteins have been detected in mammalian nuclei [3, 4, 7–10]. These nuclear foci rich in RNAPII complex proteins and transcribing RNAPII-dependent genes are termed transcription factories [3, 4, 11]. RNAPII complexes transcribe protein-coding genes and noncoding transcripts of various lengths including long noncoding RNAs and microRNA precursors. Similar to the nucleolus, transcription factories contain multiple active polymerase complexes capable of simultaneously transcribing genes located on different chromosomes as well as components of the splicing and RNA processing machinery [3, 4, 7, 12, 13].

In the nucleus DNA is wound around octamers of the core histones (H2A, H2B, H3, and H4) and associated with other chromatin-related proteins which regulate the accessibility of the underlying DNA and ultimately affect genome function. More densely packed heterochromatin is generally observed at both the nucleolar and nuclear periphery while the nucleoplasm contains less densely packed euchromatin [14, 15]. The amount of condensed chromatin varies between different cell types with undifferentiated pluripotent embryonic stem cells containing mainly dispersed “open” chromatin, whereas more differentiated cell types contain more densely packed chromatin at the nuclear periphery [16]. The most dramatic example of changes in chromatin organization with cellular phenotype has been observed in rodent rod nuclei where dense heterochromatin is observed at the center of the nucleus surrounded by decondensed chromatin; this inverted organization is hypothesized to aid night vision in nocturnal mammals [17].

Individual chromosomes occupy contiguous globular territories in the nucleus [18]. The relative position of specific chromosome territories in the nucleus varies with the more gene-dense chromosomes tending to occupy a more central position

and gene-poor chromosomes tending to occupy more peripheral positions adjacent to the nuclear lamina [19–21]. Interestingly, chromosome positional preferences differ between cell types. For example, mouse chromosome 15 localizes to the nuclear interior in liver cells but to the nuclear periphery in lung cells [22]. Moreover, preferential chromosome pairing differs depending on the cell type. For example, chromosome 12, 14, and 15 form clusters at significantly higher frequencies in mouse lymphocytes while chromosome 5 and 6 more frequently associate in mouse hepatocytes [22]. In addition, these pairing preferences change as cells differentiate, for example, as cells progress from preadipocyte to adipocyte, human chromosomes 12 and 16 become more closely associated in the nucleus [23]. Chromosome territories have been observed to intermingle with each other along their boundaries by both confocal and electron microscopy suggesting they are able to make contact at the molecular level [24]. While these cell type-specific chromosome pairing preferences have been observed as trends in cell populations, they are not a requirement for cell identity and in a given population of cells many different chromosome arrangements can be observed.

## What Role Does Genome Organization Play in Cancer?

Cancer arises due to both inherited genetic variations that predispose individuals to the disease and in addition to an accumulation of somatic mutations and epigenetic changes that convert normal cells into cancer cells [25]. Genome-wide association studies (GWAS) identify single nucleotide polymorphisms (SNPs) linked to specific human diseases or phenotypic traits and can potentially identify genetic variations that predispose individuals to diseases. These studies have revealed that greater than 80 % of disease-linked SNPs are located in noncoding regions of the genome where functional significance is difficult to determine [26]. Although sequencing cancer genomes identifies mutations present in established tumors, most cancer cells develop a genome instability phenotype which further contributes to the accumulation of mutations [27]. Mutations that accumulate in tumor cells complicate the identification of causal mutations in cancer initiation and can be heterogeneous even within individual tumors [25]. The number of observed single nucleotide mutations in human cancer cells is influenced by chromatin organization and accessibility [28]. Specifically, increased mutation rates are found in regions of the genome bearing histone modifications associated with heterochromatin (H3K9me3, H3K9me2, H4K20me3) and conversely decreased in regions bearing histone modifications associated with euchromatin (H3K4me3, H3K9ac). Some cancers are thought to arise due to translocation events which alter chromosome structure forming chimeric chromosomes. Cancer causing translocations disrupt normal cellular function by either producing a hybrid protein product due to the fusion of two open reading frames or fusing the regulatory elements for one gene in the vicinity of another gene leading to altered gene expression. Translocation frequencies observed

in malignancies are correlated with relative chromosome position in the nucleus of normal cells, suggesting that nuclear organization predisposes certain cell types to specific translocation events [24, 29]. Furthermore, translocations cause changes in nuclear position of the fused chromosome segments and widespread changes in gene expression, some of which may be due to the altered nuclear position [21, 30].

## Chromosome Translocation and Cancer

Recurrent balanced rearrangements, also known as translocation, are hallmarks of neoplasia [31]. Translocations occur when DNA double-strand breaks (DSBs) on two nonhomologous chromosomes fuse to form two chimeric chromosomes. In the case of translocations associated with cancer, the fusion frequently results in the activation or deregulation of an oncogene or the creation of fusion genes that contribute to the development of cancer. The first recurrent translocation identified in human cancer was the t(9;22)(q34;q11), widely known as the Philadelphia chromosome, that causes the fusion of *BCR* gene on chromosome 9 and *ABL1* gene on chromosome 22. This translocation was first identified in chronic myeloid leukemia by Nowell and Hungerford in 1960 [32]. To date, 61,846 translocation events involving 975 fusion genes have been identified in hematopoietic malignancies and solid tumors [31]. Despite the prevalence of translocations in cancer, the mechanisms that contribute to translocation have only started to become clear in the past decade.

Translocation requires that a DNA DSB occurs in at least two sites in the genome. DSBs represent a form of DNA damage generated by ionizing radiation, free radical oxidative damage, or spontaneous hydrolysis [33, 34]. They can also be generated by normal cellular processes. For example, early B or T cells undergo a process called V(D)J recombination to assemble immunoglobulin (*Ig*) or T-cell receptor (*TCR*) genes using one V, one D, and one J gene segment, which are selected from three sets of segments [34]. The RAG endonuclease is responsible for introducing DSBs into the *Ig* or *TCR* locus, allowing recombination to occur [34]. Following V(D)J recombination, mature B cells undergo class switch recombination that changes the type of *Ig* by introducing DSBs into the *Igh* locus by the activation-induced cytidine deaminase (AID) protein [35]. DSB can be repaired by two main pathways: homologous recombination repair and nonhomologous end-joining [36]. The homologous recombination repair pathway uses the homologous region on another chromosome as template to repair the DSB. If the homologous region to be used was located on the sister chromosome, no translocation would occur; however, if the homologous region was located on a nonhomologous chromosome, translocation would occur. Nonhomologous end-joining, on the other hand, joins the ends of two DSBs with little or no homology and as a result can cause translocations if two DSBs are located in spatial proximity to each other.

## Techniques to Investigate Organization of the Cell Nucleus

Advances in understanding nuclear and genome organization were initially dominated by microscopy techniques which have also become diagnostic tools to identify chromosomal rearrangements in patient samples. In situ hybridization (ISH) detects the location of specific DNA or RNA sequences in the nucleus. The use of fluorescence in situ hybridization (FISH) provides the best resolution in signal location using conventional light microscopy and can be combined with immunofluorescence for specific proteins to simultaneously detect nuclear protein compartments. Electron microscopy allows for the greatest resolution of nuclear ultrastructure and in combination with electron spectroscopic imaging individual nucleosomes and chromatin fibers can be visualized [16, 37]. RNA FISH detects the location of RNA molecules; with the use of probes that bind specifically to gene introns, primary transcripts can be observed at the site of transcription as introns are rapidly co-transcriptionally spliced [38, 39]. Primary transcript RNA FISH for protein coding genes, combined with immunofluorescence detection of one of the proteins in the RNAPII complex, reveals that the vast majority (>90 %) of primary transcripts co-localize with RNAPII factories [7, 40]. DNA FISH detects the location of specific DNA sequences in the nucleus irrespective of transcriptional status for a particular gene. DNA FISH of metaphase chromosomes is extensively used to identify translocations in patient samples by using probes both 5′ and 3′ of common breakpoints in specific cancers. For example, fusions of the *MYC* and *IGH* loci are the most common translocations observed in Burkitt's lymphoma, and DNA FISH for these two loci can therefore be used in disease diagnosis [41]. As mentioned above, entire chromosomes occupy distinct territories in the interphase nucleus; the location of specific chromosome territories can be detected by DNA FISH using a labeled probe generated from degenerate oligonucleotide-primed-PCR amplification of flow sorted chromosomes [42]. While these techniques investigate nuclear position of specific sequences in individual cells, the resolution is limited and specific interactions at the molecular level are difficult to investigate using microscopy-based techniques.

Imaging labeled chromatin in live cells has revealed that the genome does not exist as a static structure in the nucleus; instead chromatin is mobile in the nuclear space and appears to undergo nearly constant motion [43–45]. After mitosis, as the chromosomes decondense, chromosome territories are quickly established in early G1; when individual labeled chromosomes were observed in real time, the overall position of the territory in the nucleus did not change dramatically after decondensation and only small changes in total volume were observed [46]. However, individual tagged loci show more dynamic behaviors, including rapid ATP-dependent movements and diffusional mobility which can be restricted by association with nuclear compartments [43, 44]. The process of transcription is also dynamic with individual genes undergoing transcriptional pulses of varying magnitude and frequency [7, 47, 48]. While chromatin has been observed to undergo both diffusional and directed movements, fluorescently tagged DSB ends, induced by

site-specific cutting, remain relatively stationary in the nuclear space supporting a model in which chromosomal translocations are predominantly formed between spatially proximal DSBs [49].

## Spatial Organization of the Genome Influences Translocation Frequency

The spatial positioning of gene loci involved in chromosomal translocation has been extensively studied in the context of hematopoietic malignancies such as chronic myeloid leukemia and acute myeloid leukemia. As mentioned above, chronic myeloid leukemia frequently involves the t(9;22)(q34;q11) translocation that generates the BCR-ABL fusion gene [31]. DNA FISH of the *ABL* and *BCR* loci in normal bone marrow cells revealed that the *ABL* and *BCR* loci are more frequently juxtaposed with each other compared to their homologues [50]. Furthermore, this close proximity of *ABL* and *BCR* is maintained throughout the cell cycle in hematopoietic progenitor cells, myeloid cell types, and lymphoid cells [51]. These studies suggest that spatial proximity between *ABL* and *BCR* is a common phenomenon in hematopoietic cell types. This is consistent with the frequency of *BCR-ABL* translocations in diverse hematopoietic lineages of chronic myeloid leukemia patients [52]. Other loci pairs frequently involved in hematopoietic cancer causing translocations have been observed in close proximity in the nucleus, including *PML* and *RARα*, *MYC* and the *IGH*, *IGK*, and *IGL* loci, *IGH* and the *BCL1*, *BCL2*, and *BCL6* loci [13, 51, 53]. Furthermore, the translocation frequencies observed in cancer cells positively correlate with closer locus proximity in normal cell nuclei [13, 53]. Interestingly, mouse *Igh* and *Myc* loci, frequent translocation partners in B cell cancers of humans and mice, are located in closer spatial proximity in B cells compared to kidney cells [13]. These two genes are frequently found in the same transcription factory when both genes are transcribed in activated B cells [13]. As transcriptionally active loci have a higher incidence of recombination events, co-localization in a shared transcription factory may contribute to translocation potential [54]. These studies suggest that the spatial proximity of translocation-prone gene loci in normal hematopoietic cells can explain the high frequencies of translocations between these loci in hematopoietic malignancies.

In addition to hematopoietic malignancies, thyroid tumors also provide evidence for the link between spatial proximity in the nucleus and translocation frequency [55–57]. Thyroid cancers frequently involve intrachromosomal inversions on chromosome 10 causing the fusion of the *RET* gene with the *H4* gene, 18 Mb telomeric of *RET* [58–60]. Juxtaposition of the *RET* and *H4* genes in the nucleus occurs at a significantly higher frequency than that of *RET* and *D10S539*, despite the fact that *D10S539* is located only 12 Mb telomeric of *RET* [55]. This juxtaposition in the nuclear space does not occur in all cell types, as it was observed in thyroid cells and in peripheral blood lymphocytes but not in normal mammary epithelial cells.

Similarly, inversions on chromosome 1 causing the fusion of *NTRK1* and *TPR* gene loci have been observed in thyroid tumors, and *NTRK1* and *TPR* gene loci are more frequently juxtaposed in normal thyroid cell nuclei compared to peripheral blood lymphocytes [56]. All together, these studies suggest that gene loci frequently observed to participate in rearrangements found in specific cancers tend to be located in close proximity in the nucleus of the normal cells from which the cancer cell originated. Furthermore, this cell type-specific genome organization may predispose different cell types to different genome rearrangements.

In addition to the single loci studies, experiments observing chromosome territory positions have revealed their proximity also correlates with translocation frequency [22, 24, 29, 61, 62]. As mentioned above, chromosome territories are organized radially in the nuclear space such that the gene-dense chromosomes preferentially localize to the nuclear interior whereas gene-poor chromosomes preferentially localize to the nuclear periphery [19–21]. In general, translocation frequency is higher between chromosomes with similar gene densities than between those with different gene densities reflecting their proximity in the nucleus [61]. Human chromosome 17, 19, and 22 are among the most gene-rich chromosomes, having densities of 15, 23, and 10 genes per Mb, and they localize to the nuclear interior; chromosome 18 on the other hand is gene-poor, having a density of 6 genes per Mb and is observed more often at the nuclear periphery. Analysis of translocations observed in routine cytogenetic analysis demonstrated that chromosome 17 translocates more frequently to chromosome 19 or 22 than to chromosome 18 [61]. Furthermore, when DNA DSBs are introduced randomly in the genome through low linear energy transfer, the translocations occur more frequently between chromosomes with higher gene density that are more centrally positioned in the nucleus [62]. Chromosomes located in close spatial proximity in the nucleus have been observed to intermingle by DNA FISH of cryosectioned nuclei, with the degree of intermingling between pairs of chromosome territories showing a strong correlation with the frequency of translocations between the same pairs [24].

## Techniques to Investigate Chromatin–Chromatin Interactions

To investigate genome organization principles with finer resolution than achieved in microscopy studies, a related group of techniques has been developed which identifies chromatin–chromatin proximity at the molecular level. The chromosome conformation capture (3C) technique was developed in 2002 to investigate the folding of a yeast chromosome [63]. Since that time, several techniques based on the 3C assay have been developed including the 4C and Hi-C techniques (Fig. 1). The basic 3C technique uses formaldehyde cross-linking to fix regions of the genome interacting in vivo. A restriction enzyme is used to break the genome up into separate chromatin complexes which are ligated under dilute conditions such that ligation events occur between DNA contained within the same multiprotein/DNA complex. Cross-links are then reversed, DNA is purified, and ligation events can be identified by PCR using primers directed against two different locations of the genome.

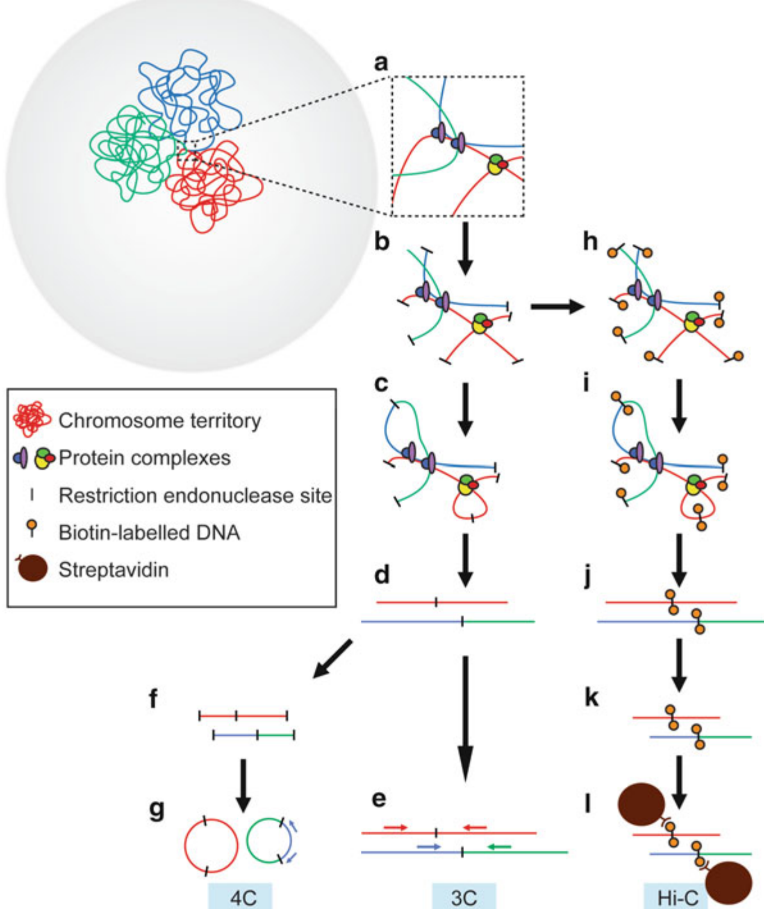



**Fig. 1** Chromosome conformation capture techniques. In the original chromosome conformation capture (3C) protocol, formaldehyde is introduced into the culture medium which introduces chemical cross-links and preserves chromatin–chromatin interactions (**a**). The nuclei are lysed and the chromatin is fragmented using a sequence-specific restriction endonuclease enzyme (**b**). The resulting chromatin complexes are diluted to minimize interaction between different complexes and ligated under these dilute conditions which allows restriction digested DNA ends contained in the same chromatin complex to fuse (**c**). Next the formaldehyde cross-links are reversed; the proteins are digested and the DNA is purified (**d**). Interactions in the resulting 3C DNA library are then investigated using primers specific for different regions of interest (**e**). In 4C, circular chromosome conformation capture; the 3C protocol is followed to step (**d**) at which point the resulting DNA is further fragmented using another restriction enzyme (**f**). The smaller fragments are then circularized by ligation and primers specific to a bait region of interest (*blue*) are used in inverse PCR to amplify the regions interacting with the bait fragment (**g**). The amplified 4C library can be analyzed using custom microarrays or subjected to deep sequencing. This technique identifies genome-wide interactions with a specific bait region of interest. The Hi-C technique is similar to the 3C technique, however, after digestion (**b**) the restriction enzyme digested sticky ends are extended to a blunt end in the presence of a biotinylated nucleotide to label the chromatin ends (**h**). Dilute ligation is carried out in a similar manner to the 3C protocol (**i**). Next the formaldehyde cross-links are reversed, the proteins are digested, the DNA is purified, and biotin nucleotides are removed from the unligated ends (**j**). The purified DNA is sonicated to reduce the average fragment size (**k**), and streptavidin-linked beads are used to purify biotinylated fragments (**l**). The purified material is then subjected to deep sequencing to identify genome-wide chromatin–chromatin interactions

The 3C technique was quickly adapted from the original yeast assay to identify chromatin loops formed between the β-globin gene promoter and distal enhancers of the locus control region several kb away [64, 65]. The next adaptations of this technique, termed 4C, were designed for unbiased identification of sequences interacting with a specific region of interest (Fig. 1). After generating the 3C material, 4C relies on circularization in a subsequent ligation step. A specific region of interest, the bait region, and all fragments ligated to that bait are then amplified by inverse PCR using primers specific for the bait region [66–68]. To investigate chromatin–chromatin interactions and folding of the entire genome, Hi-C was developed [69]. This technique relies on massively parallel sequencing of ligation fragments thereby capturing interactions genome wide (Fig. 1). The Hi-C technique is similar to the 3C technique in that formaldehyde cross-linking and restriction enzyme digestion are used to capture interactions and fragment the genome, respectively. After digestion with a restriction enzyme that generates 5′ overhangs, DNA polymerase is used to extend from the 3′ end "filling in" the overhang using a nucleotide mixture containing one biotinylated nucleotide which therefore marks the now blunt cut sites with biotin. The dilute ligation is then carried out to capture interacting complexes by forming novel ligation junctions labeled with biotin. After further fragmenting the genome by sonication, the biotinylated junctions can be purified with streptavidin conjugated beads and prepared for massively parallel sequencing to identify genome-wide interaction frequencies. While this technique has the potential to generate high resolution genome-wide interaction frequency profiles, the resolution is limited by the depth of sequencing with over 100 million sequences providing roughly 20 kb resolution across the human genome [70].

## Genome-Wide Identification of Translocations

Several studies have used high-throughput sequencing-based technologies to quantitatively measure genome-wide translocation frequency in B cells [71–74]. In addition, assessing translocation frequency and genome organization, with 4C and Hi-C, in the same cell population allows for a genome-wide investigation of the role of nuclear proximity in translocation events. The techniques to identify translocation frequency are similar to the 4C technique in that they identify genome-wide chromosomal rearrangements with a specific targeted "bait" region of interest [71, 72]. A DSB is introduced in the bait region by incorporating the 18 bp I-SceI meganuclease target sequence into the chosen gene locus (Fig. 2). The cells are then infected with a retrovirus expressing I-SceI. To detect the rearrangements, chromatin is fragmented by either restriction digestion [71] or sonication [72] and ligated to asymmetric adaptors. The wild-type bait loci can be eliminated by a second digestion with I-SceI [72]. Rearrangements are enriched by performing a nested PCR using a bait-specific primer and the adaptor primer and quantified using paired-end deep sequencing [71, 72]. These studies were conducted in B cells which undergo V(D)J recombination (DSBs mediated by RAG) or class switch recombination (DSBs mediated by AID) at different stages of maturation [75].

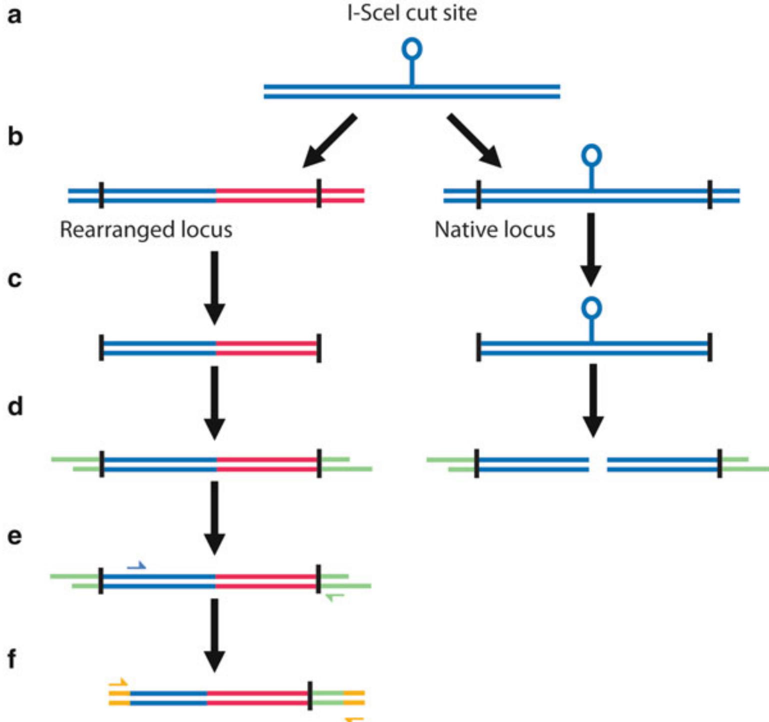



**Fig. 2** High-throughput detection of genome-wide chromosomal rearrangements. A DNA double-stranded break can be introduced at a bait gene locus (*blue*) by incorporating the I-SceI recognition site into the locus and infecting cells with a retrovirus encoding I-SceI (**a**). Double-strand break repair pathways repair the break to the native locus arrangement or generate a rearranged locus (**b**). To identify the target loci (*red*) that rearrange with the bait locus (*blue*), chromatin is fragmented by restriction digest or sonication (**c**) and ligated to asymmetric adaptors (*green*, **d**). The native loci are eliminated by a second digest with I-SceI (**d**). Rearrangements are amplified by PCR using bait-specific and adaptor primers (**e**). The resulting material is ligated to adaptors for paired-end deep sequencing (*orange*), amplified with sequencing adaptor-specific primers, and sequenced to identify chromosomal rearrangements as described in Klein et al. 2011 and Chiarle et al. 2011 (**f**)

When specific I-SceI-dependent DSBs were introduced at either the *Myc* or *Igh* loci in mouse B cells undergoing AID-dependent *Igh* class switching, the majority of junctions formed were between sequences located on the same chromosome within 10 kb of the introduced DSB. This observation supports the idea that nonhomologous end-joining preferentially joins DSBs intrachromosomally [71, 72, 76–78]. This is also consistent with the observation that most intrachromosomal rearrangements in human breast cancers involve DSBs that are less than 2 Mb apart [79]. In addition to the large proportion of intrachromosomal rearrangements, translocations with *Myc* or *Igh* were detected throughout the genome, including those between *Myc* and *Igh* which are frequent translocation partners in lymphomas [80]. Interestingly, most of these interchromosomal translocation hotspots were found in close proximity to transcription start sites of active genes and are dependent on AID

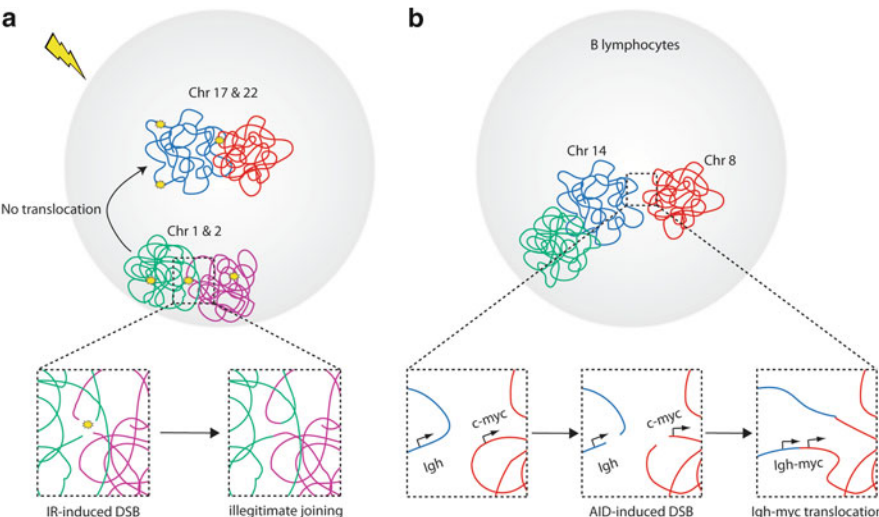



**Fig. 3** The role of chromosome position in translocation events. (**a**) When DNA double strand breaks (DSB) are introduced randomly in the genome, for example, by ionizing radiation (IR, *yellow*), high throughput translocation identification revealed that translocation events occur more frequently between chromosomes that are frequently adjacent to each other in the nucleus of the cell type investigated. (**b**) In B lymphocytes where activation-induced cytidine deaminase (AID) induces DSB as part of the class switch recombination mechanism, the location of translocation events correlates with the location of AID-induced DSB throughout the genome, for example, DSB introduced at both the *Igh* and *Myc* loci lead to frequent translocations which have been identified in Burkitt's Lymphoma patients

activity, as translocation hotspots were reduced or altered by AID deletion [71, 72]. Further investigation identified AID-mediated recruitment of replication protein A at translocation hotspots in B cells [73].

Combining high throughput translocation identification and chromosome conformation capture techniques allows for a genome-wide investigation of the role of nuclear spatial proximity in translocation events [73, 74]. G1-arrested pro-B cells with I-SceI induced DBSs on chromosome 2, 7, and 18 were used to investigate this relationship, as growth arrest activates the RAG recombinase which induces DSBs at select other gene loci (*Igκ, Tcrγ, Tcrα*). The I-SceI-dependent DSBs were most frequently translocated to RAG-induced DSBs, suggesting the frequency of DSBs is the driving force in translocation frequency [74]. Conversely, when ionizing radiation (IR) was used to introduce random DSBs across the genome in the same cells, translocations mainly occurred within the chromosome arm containing the I-SceI-induced DSBs. In this case the translocation frequency is highly correlated with the Hi-C interaction frequency genome wide, suggesting that spatial proximity has an important effect on translocation frequency when DSBs occur at similar frequencies across the genome. Taken together it appears that both the frequency of DSBs and spatial proximity of two loci are important factors that influence translocation frequency (Fig. 3). Furthermore, while the genomic

distribution of sporadic translocations is highly correlated with nuclear architecture, the location and incidence of recurrent translocations involved in B cell malignancies are correlated with site-specific DNA damage [71–74].

## How Do Translocations Affect Genome Organization?

Translocations create two hybrid chromosomes and can in turn cause changes in the nuclear position of the hybrid chromosomes. This occurs when there is discordance between the nuclear positions of the original chromosomes. For example, human chromosome 18 is preferentially found at the nuclear periphery while chromosome 19 occupies a more central location in the nucleus [21]. In asymptomatic individuals with a balanced translocation between chromosomes 18 and 19, the entire chromosome territories were not significantly altered; however, the translocated portions of the chromosomes retained their original positional preferences altering the orientation of the chromosomes in the nucleus [21]. These differences can also be seen with individual genes: in Ewing sarcoma cells fusion genes formed between the *EWSR1* and *FLI1* loci are found positioned midway between the preferred locations of the wild-type *EWSR1* and *FLI1* genes [81]. Another example of altered nuclear position was observed in phenotypically normal carriers of the t(11;22)(q23;q11) reciprocal translocation between chromosomes 11 and 22. Gene expression analysis revealed that specific genes located several Mb away from the break point on chromosome 11 were upregulated. Interestingly, the derivative chromosome 11 territory is shifted significantly toward the center of the nucleus as compared with the normal chromosome 11 territory which occupies a more peripheral position suggesting that changes in nuclear position may contribute to the observed changes in gene expression [30]. Changes in gene expression were also observed throughout the genome in individuals with this translocation which may be due to altered nuclear organization or altered expression of transcriptional regulatory proteins.

## How Do Single Nucleotide Changes in Intergenic Regions Affect Gene Expression?

GWAS have identified SNPs associated with numerous human diseases and phenotypic traits, many of which are located in uncharacterized noncoding regions of the human genome [26]. In addition, GWAS studies have been performed using SNP arrays containing less than 7 % of the 30 million known SNPs in the human genome [82]. As a result, the identified disease-linked SNPs may not be the functionally significant ones. Functional significance may be associated with other SNPs in the same linkage disequilibrium block, which tends to segregate with the disease-linked SNP. Functional genomics data on transcription factor binding, histone modification, chromatin accessibility, and associated chromatin regulatory proteins aid in

attributing functional significance to these regions. The picture is further complicated as distal regulatory elements have been observed to contact genes located at Mb distances or even on different chromosomes, and few appear to contact the closest gene in the linear genome [83–85]. As a result further investigation is required to understand the functional significance of disease-linked SNPs found in intergenic regions of the genome. While most intergenic disease-linked SNPs remain uncharacterized, a few have been investigated further [86–88].

One example is a cancer-associated SNP which lies within a 1.5 Mb intergenic region. This SNP overlaps a region which contacts the *MYC* gene promoter located 335 kb away [86]. While the chromatin loop is present at alleles that do not contain the SNP, the SNP allele preferentially binds the T-cell factor (*TCF4*) transcription factor and enhances transcription of the linked *MYC* allele [86]. Another example is a prostate cancer-linked SNP found in a 2 Mb gene dessert which lies in a 130 kb linkage disequilibrium block [87]. By comparison with histone modification data and DNase I sensitivity data, measuring chromatin accessibility, prostate cancer-specific putative enhancer regions were identified within the 130 kb linkage disequilibrium block. One enhancer overlaps the disease-linked SNP and was found to contact the *SOX9* gene, a transcription factor and oncogene that is over-expressed in prostate tumors, located 1 Mb away [87]. DNase I hypersensitive sites within the enhancer region overlap five known SNPs, two of which when mutated to the variant allele increased expression of a reporter gene and recruited androgen receptor or activator protein-1 transcription factors [87]. These studies provide a mechanism through which intergenic disease-linked SNPs may function in altering the expression of transcription factors which can in turn cause widespread changes in gene expression.

How does an aberrantly expressed transcription factor function outside of its normal cellular context? While many studies have investigated genome-wide binding of transcription factors by ChIP-Seq or ChIP-chip, few have investigated the differences in transcription factor binding between cancer cells and normal cells. One study that investigated genome-wide binding of the transcription factor TAL1 in normal erythroid cells, where the transcription factor is expressed, and in leukemic T cells where TAL1 is aberrantly upregulated, revealed striking differences in TAL1 binding throughout the genome [89]. TAL1 (also known as SCL) is a sequence-specific DNA binding transcription factor that acts as a master regulator of haematopoiesis. Normal expression of *TAL1* in the erythroid cell lineage promotes differentiation and in this context TAL1 is a regulator of numerous genes related to erythroid cellular functions [90]. Wild-type TAL1 is aberrantly upregulated in greater than 60 % of T-cell acute lymphoblastic leukemia cases and is considered a major factor in initiating the cellular transformation associated with leukemia [91]. This upregulation of TAL1, normally expressed in the erythroid lineage, is associated with a dramatic perturbation of the normal T-cell transcriptional regulatory network. Investigating genome-wide binding of TAL1 revealed only 15 % overlap between the sites bound by TAL1 in leukemic T cells compared to normal erythroid cells [89]. This indicates that TAL1 binds to different regions of the genome depending on the cellular context and the array of different transcription factors and cofactors expressed. In stark contrast to the role of TAL1 in promoting differentiation in the

erythroid cell lineage, TAL1 represses tumor suppressor genes, blocks apoptosis, and inhibits T-cell differentiation in leukemic T cells [89, 90]. In both cell types, a significant proportion of the TAL1 bound regions are located in intergenic regions rather than in the immediate promoter of expressed genes suggesting TAL1 has a significant role in regulating gene expression through chromatin looping.

As suggested by the widespread binding of TAL1 in leukemic T cells and associated changes in gene expression, widespread changes in epigenetic signatures have been identified in cancer cells. This has been investigated in colorectal cancer cells where thousands of regions bearing an epigenetic signature of enhancers (H3K4me1, H3k27ac, DNaseI sensitivity) were either acquired or lost when compared to normal epithelial cells [92]. Epigenetic modifications found at distal enhancer regions are generally cell type specific and comparison to enhancer features found in nine noncolon cell types revealed that many of the acquired enhancer signatures are present in noncolon cells [92, 93]. Conversely, enhancer signatures lost in colorectal cancer cells were generally specific to colon cells and not found in noncolon cell types. These variable enhancer regions appear to be functional as genes associated with acquired enhancer signatures were more highly expressed in colorectal cancer cells compared to normal epithelial cells, while genes associated with lost enhancer signatures were expressed at lower levels in colorectal cancer cells compared to normal epithelial cells. Thus, it appears that in colorectal cancer cells chromatin features are altered such that enhancer signatures of noncolon cells are acquired while the normal enhancer signatures of epithelial cells are lost. This is similar to the situation in T cells where TAL1 expression blocks T-cell differentiation interfering with the regulatory networks in normal T cells.

## Conclusion

Cell type variable genome organization predisposes normal cells to cancer-associated translocations between alleles that are more frequently found in close proximity in the nuclei of normal cells. However, genome-wide studies have revealed that increased rates of DSBs at specific loci (due to recruitment of enzymes that recombine the immunoglobulin loci) are the driving force for recurrent translocations observed in lymphomas. Translocation events or single nucleotide changes can cause changes in both transcription factor expression and function which in turn have genome-wide effects on transcriptional programs. Transcription factor binding in the genome and histone modifications to chromatin are dramatically altered in cancer cells reflecting the dramatic changes in cancer cell transcriptomes compared to the normal cells from which the cancer cell arose. Genome organization, chromatin accessibility, and epigenetic modifications influence the distribution of acquired mutations in cancer cells which can further contribute to disease progression. Recent efforts to identify regulatory elements in the human genome and the genes they regulate in normal cells will provide an important framework to understand altered transcriptional programs and epigenetic states in cancer cells.

**Acknowledgment** This work was supported by Canada Foundation for Innovation, Canadian Institutes of Health Research, and the Ontario Ministry of Economic Development and Innovation. We would like to thank members of the Mitchell lab for helpful discussions and critical review.

## References

1. Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. Genes Dev. 2012;26(12): 1268–86.
2. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006;124(1):207–19.
3. Jackson DA, Hassan AB, Errington RJ, Cook PR. Visualization of focal sites of transcription within human nuclei. EMBO J. 1993;12(3):1059–65.
4. Wansink DG, Schul W, van der Kraan I, van Steensel B, van Driel R, de Jong L. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. J Cell Biol. 1993;122(2):283–93.
5. Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. EMBO J. 1999;18(8):2241–53.
6. Nemeth A, Langst G. Genome organization in and around the nucleolus. Trends Genet. 2011;27(4):149–56.
7. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet. 2004;36(10): 1065–71.
8. Faro-Trindade I, Cook PR. A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value. Mol Biol Cell. 2006;17(7):2910–20.
9. Eskiw CH, Fraser P. Ultrastructural study of transcription factories in mouse erythroblasts. J Cell Sci. 2011;124(Pt 21):3676–83.
10. Ragoczy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. Genes Dev. 2006;20(11):1447–57.
11. Cook PR. A model for all genomes: the role of transcription factories. J Mol Biol. 2010; 395(1):1–10.
12. Iborra FJ, Pombo A, Jackson DA, Cook PR. Active RNA polymerases are localized within discrete transcription “factories” in human nuclei. J Cell Sci. 1996;109(Pt 6):1427–36.
13. Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol. 2007;5(8):e192.
14. Belmont AS, Braunfeld MB, Sedat JW, Agard DA. Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. Chromosoma. 1989;98(2): 129–43.
15. Kireev I, Lakonishok M, Liu W, Joshi VN, Powell R, Belmont AS. In vivo immunogold labeling confirms large-scale chromatin folding motifs. Nat Methods. 2008;5(4):311–3.
16. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One. 2010;5(5):e10531.
17. Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell. 2009;137(2):356–68.
18. Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, et al. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. Crit Rev Eukaryot Gene Expr. 2000;10(2):179–212.
19. Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet. 2001;10(3):211–9.

20. Cremer M, Kupper K, Wagler B, Wizelman L, von Hase J, Weiland Y, et al. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. J Cell Biol. 2003;162(5):809–20.
21. Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol. 1999;145(6):1119–31.
22. Parada LA, McQueen PG, Misteli T. Tissue-specific spatial organization of genomes. Genome Biol. 2004;5(7):R44.
23. Kuroda M, Tanabe H, Yoshida K, Oikawa K, Saito A, Kiyuna T, et al. Alteration of chromosome positioning during adipocyte differentiation. J Cell Sci. 2004;117(Pt 24):5897–903.
24. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 2006;4(5):e138.
25. Urbach D, Lupien M, Karagas MR, Moore JH. Cancer heterogeneity: origins and implications for genetic association studies. Trends Genet. 2012;28(11):538–43.
26. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363(2):166–76.
27. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability—an evolving hallmark of cancer. Nat Rev Mol Cell Biol. 2010;11(3):220–8.
28. Schuster-Bockler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012;488(7412):504–7.
29. Parada LA, McQueen PG, Munson PJ, Misteli T. Conservation of relative chromosome positioning in normal and cancer cells. Curr Biol. 2002;12(19):1692–7.
30. Harewood L, Schutz F, Boyle S, Perry P, Delorenzi M, Bickmore WA, et al. The effect of translocation-induced nuclear reorganization on gene expression. Genome Res. 2010;20(5):554–64.
31. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer. 2007;7(4):233–45.
32. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. J Natl Cancer Inst. 1960;25:85–109.
33. Tsai AG, Lieber MR. Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics. 2010;11 Suppl 1:S1.
34. Jung D, Giallourakis C, Mostoslavsky R, Alt F. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006;24:541–70.
35. Dudley D, Chaudhuri J, Bassing C, Alt F. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. Adv Immunol. 2005;86:43–112.
36. Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat Rev Mol Cell Biol. 2009;10(4):243–54.
37. Bazett-Jones DP, Ottensmeyer FP. Phosphorus distribution in the nucleosome. Science. 1981;211(4478):169–70.
38. van de Corput MP, Grosveld FG. Fluorescence in situ hybridization analysis of transcript dynamics in cells. Methods. 2001;25(1):111–8.
39. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. 2005;17(3):251–6.
40. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet. 2010;42(1):53–61.
41. Ferry JA. Burkitt's lymphoma: clinicopathologic features and differential diagnosis. Oncologist. 2006;11(4):375–83.
42. Telenius H, Pelmear AH, Tunnacliffe A, Carter NP, Behmel A, Ferguson-Smith MA, et al. Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow-sorted chromosomes. Genes Chromosomes Cancer. 1992;4(3):257–63.
43. Chubb JR, Boyle S, Perry P, Bickmore WA. Chromatin motion is constrained by association with nuclear compartments in human cells. Curr Biol. 2002;12(6):439–45.
44. Levi V, Ruan Q, Plutz M, Belmont AS, Gratton E. Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. Biophys J. 2005;89(6):4275–85.

45. Bornfleth H, Edelmann P, Zink D, Cremer T, Cremer C. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. Biophys J. 1999;77(5):2871–86.
46. Muller I, Boyle S, Singer RH, Bickmore WA, Chubb JR. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. PLoS One. 2010;5(7):e11560.
47. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol. 2006;16(10):1018–25.
48. Muramoto T, Cannon D, Gierlinski M, Corrigan A, Barton GJ, Chubb JR. Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. Proc Natl Acad Sci USA. 2012;109(19):7350–5.
49. Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, et al. Positional stability of single double-strand breaks in mammalian cells. Nat Cell Biol. 2007;9(6):675–82.
50. Lukásová E, Kozubek S, Kozubek M, Kjeronská J, Rýznar L, Horáková J, et al. Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. Hum Genet. 1997;100(5–6):525–35.
51. Neves H, Ramos C, da Silva M, Parreira A, Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood. 1999;93(4):1197–207.
52. Melo J. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. Blood. 1996;88(7):2375–84.
53. Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet. 2003;34(3):287–91.
54. Pratt-Hyatt MJ, Kapadia KM, Wilson TE, Engelke DR. Increased recombination between active tRNA genes. DNA Cell Biol. 2006;25(6):359–64.
55. Nikiforova M, Stringer J, Blough R, Medvedovic M, Fagin J, Nikiforov Y. Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. Science. 2000;290(5489):138–41.
56. Roccato E, Bressan P, Sabatella G, Rumio C, Vizzotto L, Pierotti M, et al. Proximity of TPR and NTRK1 rearranging loci in human thyrocytes. Cancer Res. 2005;65(7):2572–6.
57. Gandhi M, Medvedovic M, Stringer J, Nikiforov Y. Interphase chromosome folding determines spatial proximity of genes participating in carcinogenic RET/PTC rearrangements. Oncogene. 2006;25(16):2360–6.
58. Grieco M, Santoro M, Berlingieri MT, Melillo RM, Donghi R, Bongarzone I, et al. PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. Cell. 1990;60(4):557–63.
59. Bongarzone I, Butti MG, Coronelli S, Borrello MG, Santoro M, Mondellini P, et al. Frequent activation of ret protooncogene by fusion with a new activating gene in papillary thyroid carcinomas. Cancer Res. 1994;54(11):2979–85.
60. Santoro M, Dathan NA, Berlingieri MT, Bongarzone I, Paulin C, Grieco M, et al. Molecular characterization of RET/PTC3; a novel rearranged version of the RETproto-oncogene in a human thyroid papillary carcinoma. Oncogene. 1994;9(2):509–16.
61. Bickmore WA, Teague P. Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population. Chromosome Res. 2002;10(8):707–15.
62. Arsuaga J, Greulich-Bode K, Vazquez M, Bruckner M, Hahnfeldt P, Brenner D, et al. Chromosome spatial clustering inferred from radiogenic aberrations. Int J Radiat Biol. 2004;80(7):507–15.
63. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306–11.
64. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell. 2002;10(6):1453–65.
65. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet. 2003;35(2):190–4.

66. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006;38(11):1341–7.
67. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006;38(11):1348–54.
68. Wurtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended chromosome conformation capture methodology. Chromosome Res. 2006;14(5):477–95.
69. Lieberman-Aiden E, van Berkum N, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93.
70. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376–80.
71. Chiarle R, Zhang Y, Frock R, Lewis S, Molinie B, Ho Y-J, et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell. 2011;147(1):107–19.
72. Klein I, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, et al. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell. 2011;147(1):95–106.
73. Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon K-R, Jankovic M, et al. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature. 2012;484(7392):69–74.
74. Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, et al. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell. 2012;148(5):908–21.
75. Soulas-Sprauel P, Rivera-Munoz P, Malivert L, Le Guyader G, Abramowski V, Revy P, et al. V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. Oncogene. 2007;26(56):7780–91.
76. Ferguson DO, Sekiguchi JM, Chang S, Frank KM, Gao Y, DePinho RA, et al. The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. Proc Natl Acad Sci USA. 2000;97(12):6630–3.
77. Yan CT, Boboila C, Souza EK, Franco S, Hickernell TR, Murphy M, et al. IgH class switching and translocations use a robust non-classical end-joining pathway. Nature. 2007;449(7161):478–82.
78. Mahowald GK, Baron JM, Mahowald MA, Kulkarni S, Bredemeyer AL, Bassing CH, et al. Aberrantly resolved RAG-mediated DNA breaks in Atm-deficient lymphocytes target chromosomal breakpoints in cis. Proc Natl Acad Sci USA. 2009;106(43):18339–44.
79. Stephens P, McBride D, Lin M-L, Varela I, Pleasance E, Simpson J, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009;462(7276):1005–10.
80. Kuppers R, Dalla-Favera R. Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene. 2001;20(40):5580–94.
81. Taslerova R, Kozubek S, Lukasova E, Jirsova P, Bartova E, Kozubek M. Arrangement of chromosome 11 and 22 territories, EWSR1 and FLI1 genes, and other genetic elements of these chromosomes in human lymphocytes and Ewing sarcoma cells. Hum Genet. 2003; 112(2):143–55.
82. Grant SF, Hakonarson H. Microarray technology and applications in the arena of genome-wide association. Clin Chem. 2008;54(7):1116–24.
83. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. Cell. 2006;126(2):403–13.
84. Ferrai C, Pombo A. 3D chromatin regulation of Sonic hedgehog in the limb buds. Dev Cell. 2009;16(1):9–11.
85. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489(7414):109–13.

86. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol Cell Biol. 2010;30(6):1411–20.
87. Zhang X, Cowper-Sal Lari R, Bailey SD, Moore JH, Lupien M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. Genome Res. 2012;22(8):1437–46.
88. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010;42(3):255–9.
89. Palii CG, Perez-Iratxeta C, Yao Z, Cao Y, Dai F, Davison J, et al. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. EMBO J. 2011; 30(3):494–509.
90. Kassouf MT, Hughes JR, Taylor S, McGowan SJ, Soneji S, Green AL, et al. Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. Genome Res. 2010;20(8):1064–83.
91. Aifantis I, Raetz E, Buonamici S. Molecular pathogenesis of T-cell leukaemia and lymphoma. Nat Rev Immunol. 2008;8(5):380–90.
92. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, et al. Epigenomic enhancer profiling defines a signature of colon cancer. Science. 2012; 336(6082):736–9.
93. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. [see comment]. Nat Genet. 2007;39(3):311–8.

# Histone Chaperones, Epigenetics, and Cancer

**Mina Rafiei and Rod Bremner**

**Abstract** Histone chaperones such as CAF1, ASF1, HIRA, DEK, DAXX, and several others play central roles in the transport, modification, replication, and replacement of nucleosomes. These diverse roles affect many processes including epigenetic memory, genome stability, transcription, Polycomb function, and others. Here, we review these functions and their relevance to heterochromatin, epigenetic inheritance, and cancer.

**Keywords** Histone chaperone • Epigenetics • Nucleosome • CAF1 • ASF1 • HIRA • DEK • DAXX • Polycomb

## Histone Chaperones, Epigenetics, and Cancer

The first histone chaperone was isolated as a nuclear protein that prevents histone precipitation and facilitates nucleosome formation [1]. Chaperones are typically acidic, ideal for binding to basic, positively charged histones [2]. They bind histones in the cytoplasm, carry them into the nucleus, and facilitate nucleosome formation (Fig. 1) [3, 4]. Chaperones were originally thought of as simple carriers that prevent

M. Rafiei, Ph.D.
Laboratory of Medicine and Pathobiology, Genetics and Development Division,
University of Toronto, Toronto, ON, Canada
e-mail: mrafiei@uhnres.utoronto.ca

R. Bremner, Ph.D. (✉)
Department of Ophthalmology and Vision Sciences, Toronto Western Hospital,
Toronto, ON, Canada

Toronto Western Research Institute, Mc6-424, 399 Bathurst St., Toronto, ON, Canada,
M5T2S8
e-mail: rbremner@uhnres.utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_15,
© Springer Science+Business Media New York 2014

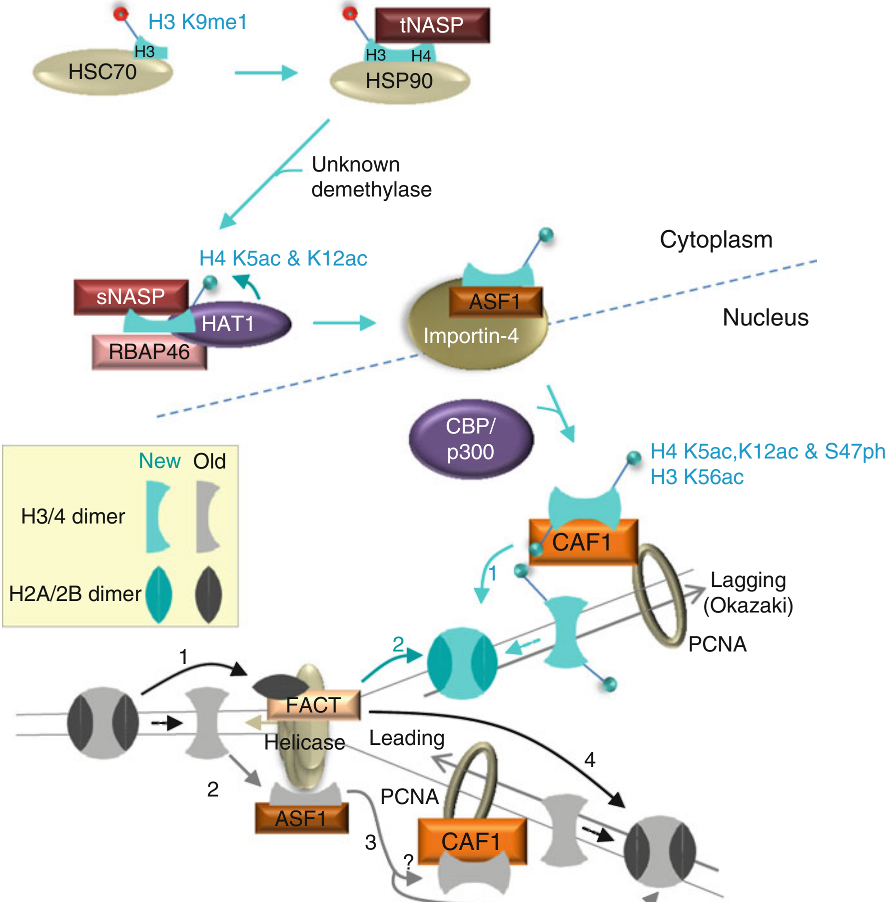



**Fig. 1** Histone chaperone functions during DNA replication. In the cytoplasm (*top*) HSC70 chaperones newly translated H3, which is methylated by an unknown enzyme. Next, HSP90 and tNASP chaperone H3K9me1, and histone H4 joins the complex to generate H3-H4 dimers. H3-H4 is then transferred to a complex containing the chaperones sNASP and RBAP46 and the enzyme HAT1. H3 is demethylated at this stage by an unknown enzyme, and HAT1 acetylates H4 on K5 and K12. Next, the acetylated H3-H4 dimer is passed to ASF/importin for nuclear import, following which CBP/p300 acetylates H3 on K56. The latter facilitates transfer to the final chaperone in the chain, CAF1, which assembles H3-H4 dimers into $(H3\text{-}H4)_2$ tetramers. CAF1 is tethered to PCNA at the replication fork and deposits the $(H3\text{-}H4)_2$ tetramer on DNA, following which H2A-H2B dimers are deposited, for example, by the dimeric chaperone complex FACT, which is tethered at the fork through an interaction with MCM4 in the helicase. At the fork the helicase unwinds DNA (*towards the left*), and behind it (*to the right*) polymerases (not shown for simplicity) synthesize the new leading or lagging strands of DNA. Ahead of the helicase, parental H2A-H2B dimers are displaced by FACT, and ASF1 then separates the remaining H3-H4 tetramer into dimers. The resultant H3-H4-Asf1 complex is tethered to the fork through an interaction of the histones with the helicase. H3-H4 is then redeposited on DNA, but it is unclear whether this occurs through another chaperone intermediate (e.g., CAF1 etc.), or whether tetramers form directly on DNA after release from ASF1 (see text). The octamer is then completed by addition of two H2A-H2B dimers. Old and new histone H3-H4 tetramers do not mix (as shown), but old and new H2-H2B dimers can mix (here, for simplicity, no mixing is indicated)

nonspecific DNA–histone interactions, but are now known to regulate histone disassembly and reassembly, covalent histone and DNA modifications, and the maintenance of epigenetic and chromatin states during replication, transcription, and DNA repair.

Classically, epigenetics is "the inheritance of variation (−genetics) beyond (epi-) changes in the DNA sequence [5]," but the term is now used commonly to describe any modification of a nucleosome, even in post-mitotic cells where "inheritance" is by definition a misnomer. When DNA is replicated or repaired, cells "remember" which genes were silent or active and nucleosomal positioning and covalent histone modifications are crucial for this process. Chaperones redistribute parental histones and deposit new histones on replicated DNA and thus they also have a profound role in epigenetic memory. Understanding this process is important because epigenetic memory is perturbed in multiple human diseases, including cancer as we review here for histone chaperones.

## Nucleosome Replication: Conservative Versus Semiconservative

Nucleosomes consist of a core $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers attached on either side, with 146 bp of DNA wrapped around the entire octamer. Histones as well as DNA are replicated, and must be incorporated into new nucleosomes and covalently modified to maintain the chromatin state. To understand how histone chaperones operate it is important to first understand nucleosome replication. For simplicity, we will limit our discussion on replication of the $(H3\text{-}H4)_2$ tetramer to the three non-centromeric mammalian somatic H3 variants, H3.1, H3.2, and H3.3, that combine with an invariant H4 [6]. H3.1 and 3.2 show similar properties and we will refer predominantly to the more highly expressed H3.1 isoform.

Nucleosomes could, in theory, follow conservative or semi-conservative modes of replication. In the former model, old (parental) histones are transferred intact to one of the two replicated DNA strands, and new histones are used to generate a second nucleosome at the matching position. In semi-conservative replication, the parental histone octamers are split, and they mix with new histones to create two hybrid nucleosomes. An elegant study suggests that in HeLa cells H2A-H2B dimers are always replicated semi-conservatively (new mixed with old), but H3.1-H4 tetramers undergo conservative replication [7]. Inducible Flag-tagged Histone H3.1 plus growth in normal lysine (K0) media was used to label old nucleosomes, then Flag-H3.1 expression was extinguished, cells were arrested at G2/M, released and grown in heavy lysine isotope (K8) to label new histones through the next S-phase. Mononucleosomes with old Flag-tagged H3.1 were purified, and Mass Spectrometry (MS) used to analyze gel purified histones. Subtracting background label, only 2 % of H3.1 or H4 in Flag-tagged (i.e. old) nucleosomes contained K8 (i.e. new) label, and this did not increase much even after a second S-phases. Thus, there is almost

no splitting of $(H3.1\text{-}H4)_2$ during S phase, indicating conservative replication. In contrast, H2A and H2B in Flag-tagged nucleosomes were ~50 % labeled with K8 after one S-phase, indicating semi-conservative replication. The latter is consistent with the fact that H2A-H2B dimers are not in contact in the nucleosome, but reside either side of the tightly bound pair of H3-H4 dimers [8].

The replication/deposition of (**H3.3**-H4)$_2$ tetramers is different from that of (**H3.1**-H4)$_2$ tetramers. Human H3.1 or H3.2 differ at only five or four amino acid positions from H3.3, respectively, but whereas their expression is restricted to S-phase, H3.3 is expressed throughout the cell cycle [9]. Also, while **Asf1** (anti silencing function 1) and **CAF1** are the chaperones that deposit new H3.1-H4 dimers, **HIRA** (histone regulator A), **DAXX** (death associated protein six), or **DEK** [10] chaperones deposit new H3.3-H4 dimers [11–15]. Consistent with separate deposition pathways, H3.1/3.2-H4 dimers are not found with H3.3-H4 dimers in nucleosomes [12]. Using the labeling system discussed above, Xu et al. found that ~6 % and ~20 % of nucleosomes with old Flag-tagged/K0 H3.3 contained new K8-labeled H3.3 after one or two S-phases, respectively [7]. This $(H3.3\text{-}H4)_2$ tetramer splitting occurs during DNA synthesis but not during replication-independent deposition, as it is greatly reduced when S-phase is blocked.

In summary, H3.1-H4 tetramers do not split but undergo conservative replication; some H3.3-H4 tetramers split and undergo semi-conservative replication, and new H3.3-H4 complexes added outside S-phase are deposited as tetramers and do not mix with H3.1-H4 dimers.

## Chaperones Regulate Transport, Modification, and Deposition of New Histones

Newly generated histones must avoid aggregation, move into the nucleus, accumulate posttranslational modifications, and form nucleosomes on DNA. Chaperones play crucial roles in all these processes. A recent study combined protein purification, biochemical reconstitution assays, and RNAi/genetics to deduce a comprehensive view of the early stages of histone synthesis and transport into the mammalian or yeast nucleus (Fig. 1) [16]. Here, we highlight the mammalian process.

As H3.1 emerges from the ribosome its folding is assisted by HSC70 [16]. H3.1 is monomethylated at this early stage on lysine 9 (K9me1), although the mechanism is unclear [17]. The newly synthesized non-replicative histone H3.3 is typically dimethylated (K9me2) or exhibits acetylation (K9ac and K14ac) [17].

From **HSC70**, H3 is passed to **HSP90** which cooperates with the chaperone **tNASP1** ("testicular Nuclear Autoantigenic Sperm Protein") to assemble H3.1-H4 dimers (Fig. 1) [16]. Through splicing the NASP gene also encodes "somatic" NASP (sNASP) [18]. H3-H4 dimers in the tNASP-HSP90 complex are passed onto **sNASP**, which also recruits the HAT1 (histone acetyl transferase 1) holoenzyme, made up of HAT1 and the chaperone **Rbap46** (RBBP7) [16]. Rbap46 and its fly homolog p55 bind histone H4 [19] and sNASP preferentially binds histone H3 [20],

providing a logical explanation as to how histone dimers are held by the sNASP-Rbap46 dual chaperone complex. When H3-H4 dimers are passed to the sNASP-Rbap46 complex, a dramatic reduction in H3K9me occurs by an unknown mechanism, and HAT1 catalyzes H4 acetylation on K5 and K12 (Fig. 1). sNASP can dimerize [21], and complexes have been observed containing one or two H3-H4 dimers, but not tetramers [16] which, as described later, form in the nucleus (Fig. 1). In budding yeast (*Saccharomyces cerevisiae*) the equivalent of sNASP is Hif1, and the Hat1p/Hat2 holoenzyme performs acetylation, much like the mammalian HAT1-Rbap46 complex. HAT1 or Hat1p driven K5ac and K12ac modifications appear to enhance association with nuclear transporters and thus nuclear uptake [22, 23].

NASP is essential for murine development [18] and plays a key role in adjusting the soluble reservoir of H3-H4. Reducing NASP leads to autophagy-mediated depletion of soluble H3-H4, whereas reducing HAT1 (and thus histone acetylation) or Asf1 does not lower cytosolic H3-H4 [24]. Thus, NASP provides a tunable cytoplasmic source of H3-H4.

After passing from the HSC70 to the tNASP complex and then to the sNASP/Rbap46/HAT1 complex, the final cytosolic stage involves the transfer of H3-H4 dimers to a complex containing the conserved chaperone ASF1 and Importin-4. There are two human ASF1 genes, A and B, but only the B protein is found at this stage, although A likely takes over when B is artificially removed [16]. Importin-4 is a karyopherin family member that mediates transport through the nuclear pore. ASF1 acts as a sink for H3-H4 dimers since most non-nucleosomal nuclear histone H3-H4 is found associated with this chaperone [25].

Once in the nucleus, H3.1-H4 dimers pass from ASF1 to the **CAF1** (chromatin assembly factor 1) complex. H3.3-H4 dimers are handled by HIRA for deposition on genic regions [11, 13], or **DEK** for an as yet uncharacterized target [10], and **DAXX** for deposition on heterochromatin [14, 15, 26]. For simplicity, we will focus on the H3.1-H4/CAF1 interaction below, but we will also discuss DEK and DAXX in the section on chaperones and cancer.

CAF1 was originally isolated as a human complex that promotes in vitro nucleosome assembly on replicating SV40 viral DNA templates [27]. It is conserved from yeast to humans and consists of three subunits termed Cac1-3 in yeast [28], p180, p105, and p55 (NURF) in *Drosophila* [29, 30], or p150 (CHAF1A), p60 (CHAF1B), and p48 in humans [27]. The latter is also called Rbap48 (RBBP4), a close relative of Rbap46 (RBBP7) mentioned above (see sNasp/Rbap46/HAT1 complex, Fig. 1)). Asf1 can bind directly to the mid-sized subunit of CAF1 (p60/CHAF1B) [31, 32], while the small subunit (Rbap48) binds both H3 and H4 [33].

For many years it was unclear whether newly synthesized histones H3 and H4 form tetramers prior to deposition by CAF1, or only after deposition on DNA. Three recent studies all suggest the former. The first study found that mutations affecting neuronal fate in *Caenorhabditis elegans* mapped to a C-terminal region of histone H3 required for H3-H4 tetramerization [34]. *C. elegans* has 24 histone H3 genes, but mutation of only one acted as a dominant negative to block nucleosome formation. Depleting CAF1 or PCNA (required to recruit CAF1—see below) caused the same neuronal phenotype, whereas depleting Asf1 did not alter fate. These data suggest

that the phenotype results from an inability of CAF1 to assemble H3-H4 tetramers. The second piece of evidence came from cross-linking studies which found that a single molecule of yeast CAF1 binds H3-H4 tetramers [35]. And the third involved thermodynamic studies showing that once yeast ASF1-H3-H4 binds to CAF1, ASF1 is dislodged and a second H3-H4 dimer is introduced to form the H3-H4 tetramer (Fig. 1) [36].

The latter study also found that the affinity of CAF1 for unmodified H3-H4 dimers is ~2-fold lower than that of ASF1, raising the question of how the transfer could be thermodynamically favorable. Notably, however, posttranslational histone marks increase the affinity of histones for downstream chaperones [37]. In the nucleus the yeast or mammalian acetyl transferases Rtt109 or CBP/p300 catalyze H3K56 acetylation, respectively, which is stimulated considerably by Asf1 or the yeast histone chaperone **Vps75** [38–40], and this modification promotes association of H3 with CAF1 or the yeast histone chaperone **Rtt106** [41]. Yeast lacking Asf1 or H3K56 acetylation have a reduced life span [42]. Another modification that influences binding is PAK2-mediated phosphorylation of H4 Ser 47, which diverts H3.3-H4 dimers away from CAF1 to HIRA [43]. Thus, chaperones facilitate histone modification and/or exploit these events to guarantee the passage of histones along the chaperone chain in the right direction and with the correct partners. Several other marks have been described on newly synthesized histones, several of which affect chromatin assembly and sensitivity to DNA damaging agents, but the underlying mechanisms are largely unclear [44].

The final step in the journey of a new H3-H4 tetramer is deposition onto its highest affinity partner, DNA. For the CAF1-$(H3\text{-}4)_2$ complex, this is facilitated by interaction with PCNA (proliferating cell nuclear antigen). PCNA forms a trimeric clamp that completely encircles replicating DNA strands [45]. Its interaction with DNA involves water molecules, allowing it to glide along the template and improve DNA polymerase processivity (the number of nucleotides replicated without polymerase dissociation). The large subunit of CAF1 (p150/CHAF1A in humans) binds directly to PCNA [46]. As discussed below, CAF-1 is critical for the inheritance of heterochromatin, and mutations in yeast PCNA that affect CAF1 recruitment disrupt heterochromatin-mediated silencing [47], and this interaction is also critical for chromatin assembly after DNA damage [48]. Moreover, dominant negative CAF1 mutants that do not bind PCNA or p60/CHAF1B disrupt chromatin deposition and lead to activation of the DNA damage checkpoint [49].

In summary, chaperones guide H3-H4 dimers from their synthesis in the cytoplasm to their destination on DNA through a thermodynamically favorable chain of binding reactions (Fig. 1).

## Chaperones Recycle Old Histones During Replication

Chaperones also disassemble nucleosomes as the replication fork passes and then reassembles them on both strands of replicated DNA. The first step in the disassembly of parental nucleosomes is removal of H2A/H2B. **FACT** (facilitates chromatin

transcription) is an H2A/H2B chaperone and while most of the work on its function focuses on transcription, it is also involved in altering chromatin structure during DNA replication [50, 51]. FACT is recruited to the replication fork through an interaction with MCM4, one of six proteins (MCM2-7) that make up the helicase that unwinds DNA [52]. Another chaperone that could be involved in removing H2A/H2B from DNA is **Nap1** (nucleosome assembly protein 1) which operates in collaboration with the ATP-dependent chromatin remodeling factor RSC in vitro [53]. Nap1 is thought to be particularly important for removal of H2A/H2B during transcription [54], reviewed in [55].

Once H2A/H2B is removed, the more stable H3/H4 can be displaced. H3/H4 dimers are then assembled onto nascent DNA quickly, whereas histones H2A and H2B are added 2–10 min after fork passage [56]. Asf1 plays a key role in coordinating the removal of H3/H4 dimers with their redeposition behind the fork [57]. Asf1 splits $(H3\text{-}H4)_2$ histone tetramers into dimers, and an Asf1-H3/H4-MCM sandwich briefly tethers them to the helicase (Fig. 1). Knockdown of both Asf1 genes (Asf1a & b) prevents nucleosome removal, the helicase fails to unwind DNA (reflected in reduced levels of single stranded (ss) DNA levels at the fork), and cells arrest in S-phase. Inhibiting DNA polymerase causes an accumulation of parental histones on Asf1, identifiable by covalent marks absent on newly synthesized histones (H3K9me3/H4K16ac) [57]. Asf1 is also required to bring new H3-H4 dimers to DNA (see above [58]), thus when H3-H4 dimers are over-expressed, Asf1 becomes limiting, parental histones are not dislodged, the helicase stalls, ssDNA levels drop, and cells arrest in S-phase [57]. This S-phase block can be rescued by elevating Asf1 levels [57]. Therefore, Asf1 coordinates both recycling of parental and introduction of new histone H3-H4 dimers.

The above data raise an interesting conceptual problem. As noted earlier, $(H3.1\text{-}H4)_2$ tetramers exhibit conservative replication [7]. But Asf1, which plays a key role in tethering old H3-H4 dimers to the replication fork [57], splits tetramers [2, 59]. Once old H3.1-H4 tetramers are split into dimers, how do they re-associate and remain separate from new H3.1-H4 dimers? Re-association of old dimers might be the only option given that new tetramers are preassembled on CAF1 (see above). Whether H3.1-H4 dimers from an old nucleosome remain closely associated to a pair of Asf1 molecules at the fork is unclear. Perhaps Asf1 traffics old histones through CAF1 at the fork, although in vitro assays show that CAF-I cannot assemble histones H3 and H4 purified from cellular chromatin onto DNA [27, 60]. Whether this is the case at the replication fork in vivo, however, is unclear. Nap1 or its close relative Vps75 form homodimers that adopt an earmuff structure and directly bind the two H3 proteins in an intact $(H3\text{-}H4)_2$ tetramer [61], but whether this feature is exploited during conservative replication of the core nucleosome is unknown. Alternatively, covalent modifications and/or associated proteins (chaperones?) on old and/or new dimers preclude mixing during deposition.

There are two ASF genes in mammalian cells and while knockdown of both ASF1A and B is required for acute arrest of cultured cells, removing ASF1B, but not A, blocks growth in colony formation assays [62]. The reason for this difference is not fully resolved, but ASF1B has other qualities that distinguish it from 1A, such

as its downregulation in quiescent or senescent cells, and ASF1B deficiency causes unique effects on the transcriptome, and the appearance of mitotic defects such as micronuclei and DNA bridges [62].

## Chaperones and Epigenetic Memory: The Example of CAF1

Maintenance of nucleosomes during replication or repair is, by definition, a key aspect of epigenetic memory. However, in addition to passing on old and depositing new histones, chaperones play additional epigenetic roles. The full extent to which chaperones regulate this process and the mechanisms therein are largely obscure. Most work has been performed on CAF1 and ASF1, particularly on maintenance of heterochromatin. Below, we summarize the data for CAF1 [63].

Heterochromatin is typically rich in repetitive DNA such as centromeric satellite sequences and telomeric repeats. These dense domains are gene poor, less accessible, coated with repressive hypoacetylated histones and histone H3 trimethylated on lysine 9 (H3K9me3), and replicated late in S-phase. H3K9me3 tethers heterochromatin protein 1 (HP1) which self-associates to promote condensation [64]. Moreover, HP1 recruits the H3K9 methyl transferases SUV39H1 and SUV39H2, which propagate this chromatin mark during replication (Fig. 2) [65, 66]. The DNA in these dense regions is easily visible upon staining with intercalating fluorescent dyes such as 4′,6-diamidino-2-phenylindole (DAPI). DAPI-intense heterochromatin spots remain even during their replication, so how is such densely packed DNA and chromatin duplicated? As discussed below, CAF1 plays a major role in this process. Indeed, CAF1 is crucial for heterochromatin maintenance in yeast, plant, fly, frog, mouse, and human cells [67–71].

Links between CAF1 and heterochromatin arose from work in budding yeast (*S. cerevisae*) where, although dispensable for survival, its subunits are essential for efficient silencing of marker genes proximal to telomeres [28, 72, 73]. Initially, it was thought that silencing at mating type loci did not require CAF1, but more sensitive assays revealed that CAF1 is essential to maintain silencing, but not for reestablishment (e.g., following Sir protein disruption) [74].

Heterochromatin in *S. cerevisae* relies on proteins like Rap2 and the Sir family of histone deacetylases, but different factors are utilized in higher eukaryotes such as the HP1 family. Remarkably, however, subsequent work provided that a direct link also exists between heterochromatin maintenance and CAF1 in higher eukaryotes.

This connection came from the discovery that the N-terminus of murine Chaf1A/p150 binds directly to HP1 [73]. In human or murine cell lines, CAF1 colocalizes with HP1 in late S-phase at sites of pericentric heterochromatin [75]. Labeling with thymidine analogues, such as BrdU, and 3D imaging showed that heterochromatin is replicated at the surface of DAPI-dense spots and is then buried inside the domain after replication [76]. In addition to trimethylation by Suv39h1/2, HP1 binding to the core of DAPI-rich heterochromatin also requires an RNA component, but in contrast Chaf1A/p150 tethers an RNase and an *Suv39h1/h2*-null resistant HP1

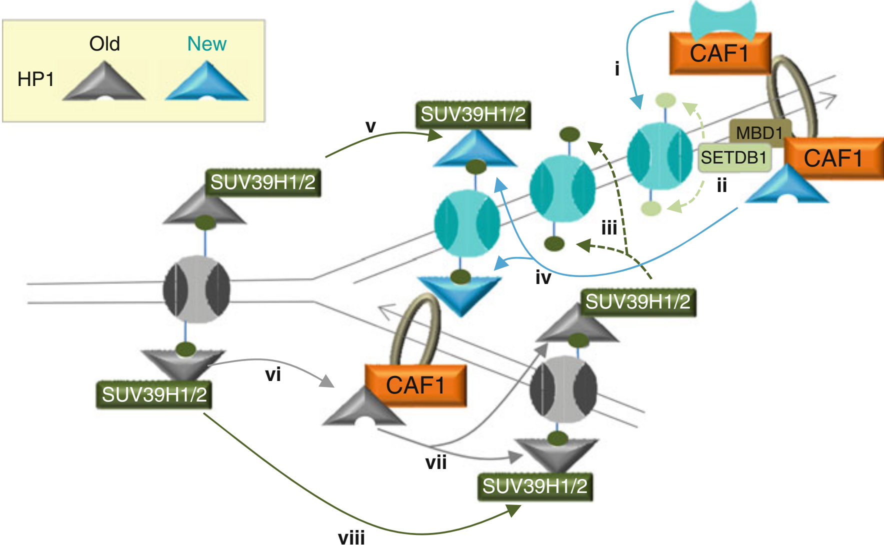



**Fig. 2** Multiple roles for CAF1 at heterochromatin. *Solid arrows* indicate protein movements, *dotted arrows* indicate methylation events. Old and new nucleosomes, PCNA, and DNA are shown as in Fig. 1. (i) CAF1 tethered to PCNA deposits new H3-H4 tetramers onto DNA (see Fig. 1 for details). (ii) A separate CAF1 complex tethers SETDB1 to the fork through an interaction with MBD1, and this enzyme monomethylates H3K9. This CAF1 complex also contains new HP1 to deposit on trimethylated H3K9. (iii) SUV39H1, tethered to HP1, di and trimethylates H3K9. (iv) New HP1, tethered to CAF1, is deposited on H3K9me3. (v) SUV39H1 is transferred to newly deposited HP1. (vi) Old (parental) HP1 is transferred to a third CAF1 complex at the fork. (vii) Old HP1 is transferred (in this example) onto old, already trimethylated H3K9 (see Fig. 1 for details on how old nucleosomes are transferred at the fork). (viii) SUV39H1 is transferred to old HP1. It is likely that old and new HP1 is distributed randomly among old and new nucleosomes, although this has not been tested. The *top right* illustrates two complexes of CAF1 bound to one PCNA trimer, which is theoretically feasible, but whether this occurs, or whether CAF1-$(H3\text{-}H4)_2$ and CAF1-MBD1-SETDB1-HP1 complexes bind sequentially (i.e. separately), is unknown. New and old SUV39H1/2 is not differentiated, but new enzyme is likely drawn from the nucleoplasm

fraction to replicating DNA on the surface of heterochromatin domains in mouse fibroblasts [76]. These data suggest that when DNA emerges at the surface of a heterochromatic domain, HP1 is displaced and tethered at the replication fork by Chaf1a. Indeed, while Chaf1a/p150 or Chaf1b/p60 knockdown reduces nucleosome deposition in S-phase DNA, only the former causes cell cycle arrest in mid-S-phase, which is not associated with a DDR, but rather an inability to replicate pericentric heterochromatin [76]. This RNAi-induced defect is complemented by wild-type Chaf1a, but not by mutants that do not bind HP1. Moreover, Chaf1a knockdown does not arrest *Suv39h1/h2* double null fibroblasts where HP1 is absent from DAPI-rich heterochromatin [77]. Altogether, these data suggest that CAF1 displaces HP1 during heterochromatin replication and holds it at the replication fork ready for redeposition (Fig. 2).

CAF1-HP1 [76] exists separately from CAF1-H3.1-H4, thus a single CAF1 complex deposit newly synthesized histones or handles old/new HP1 separately. CAF1-HP1 binds other proteins critical for heterochromatin maintenance, including Methyl Binding Domain 1 (MBD1) [78], which recruits the H3K9 methyltransferase SETDB1 [79]. While bound to CAF1, SETDB1 stimulates mono-, but not di- or tri-, methylation of H3K9 [80]. Monomethylated K9 is an excellent substrate for di- and tri-methylation by Suv39h1 (Fig. 2).

In summary, CAF1 performs many functions relevant to the epigenetic inheritance of heterochromatin (Fig. 2): i. As at other loci it loads new histone H3.1-H4 tetramers onto replicated DNA; ii. It displaces and redeposits old parental HP1 at the surface of heterochromatin domains and brings in new HP1 to maintain heterochromatin on replicated DNA; iii. CAF1-HP1-MDB1-SETDB1 mono-methylates H3K9 on new nucleosomes, allowing Suv39h1/h2—tethered to old nucleosomes—to convert H3K9me to H3K9me3, and the latter can now receive old or new HP1 from CAF1.

As well as maintenance, CAF1 is also important for the *de novo* formation of murine heterochromatin as *Chaf1a* null mouse embryos arrest at the 16 cell stage and fail to form DAPI-rich spots [81]. In ES cells, where chromatin is more plastic than in differentiated cell types, Chaf1a knockdown does not cause arrest, perhaps because HP1 is easier to displace in these cells than fibroblasts. However, these depleted ES cells die after 4 days of knockdown [81], perhaps because the failure to form heterochromatin disrupts chromosome segregation, although the latter was not tested. Whether this survival function for CAF1 depends on interaction with HP1 has not been tested.

The above model of Chaf1-HP1 mediated regulation of heterochromatin was established primarily in mouse cells. The picture is less clear in other animal species/cell types. In human, mouse, and chick cells removing Chaf1a/p150 or Chaf1b/p60 impairs nucleosome deposition in S-phase [49, 77, 82–84]. In chick DT40 cells, like mouse fibroblasts, there is no DDR when CAF1 is disrupted (indeed the response to UV or HU is dampened) [84]. However, whereas only Chaf1a knockdown perturbs S-phase progression in mouse cells, knockout of Chaf1a or Chaf1b delays S-phase in chick cells, and also causes extensive cell death by 48 h [77, 84]. Binding of chick Chaf1a to PCNA and Chaf1b is required for survival, but the interaction with HP1 is dispensable [84]. Whether Chaf1a HP1-binding mutants affect heterochromatin replication in chicks is unknown. Unlike chick/mouse cells, CHAF1A small interfering RNA (siCHAF1A) causes a DDR in the human cancer cell lines RTK, HeLa, and U2OS cells [49, 82], as does siCHAF1B in HeLa cells [83]. Viability was compromised in both the U2OS/siCHAF1a and HeLa/siCHAF1B assays. Intact heterochromatin is important for proper nucleation of spindles at mitosis; thus cell death in some of these scenarios might follow disruption of this key process, but may also be the result of an S-phase DDR in some cases. However, neither scenario applies in chicks as there is no DDR, and HP1 binding is dispensable for survival. The extent to which the above variable responses to CAF subunit disruption reflect differing chaperone redundancy, species/tissue specificity, and/or degree of neoplastic transformation is unclear. Notably, an RNAI screen in HeLa

cells identified CHAF1A as critical to maintain silencing of an integrated GFP reporter gene [85]. Other heterochromatin regulators were also identified, such as HP1 and SETDB1, consistent with the model discussed above.

Beyond its well-established role in regulating HP1 bound heterochromatin, CAF1 may regulate other repressive mechanisms. The Polycomb group (PcG) of proteins represses transcription at multiple loci, including homeobox transcription factors required for specific developmental fates [86]. Heterozygous mutations in the large Drosophila CAF1 subunit (p180, equivalent to p150/CHAF1A in humans) enhance the effect of heterozygous mutations affecting the PcG protein, Pc [70]. This genetic evidence suggests that CAF1 may facilitate the epigenetic effects of PcG complexes. But how this operates and the specific PcG complexes affected are unknown. The CAF-1 subunit, RBBP4 (Rbap48) or p55 (Nurf55) in *Drosophila*, is also a subunit of Polycomb Repressive Complex 2 (PRC2) *, but these functions are separate because structural studies show that Nurf55 uses the same region to interact with histones H3 and H4 (as part of its role in CAF1) and with SUZ12 (a component of PRC2) [33]. There is no biochemical evidence that CAF-1 and PcG interact directly, but conceivably CAF-1 might recruit other proteins to the fork that influence PcG activity. Apart from well-known interactions with PCNA, ASF1, histones, and HP1-MBD-SETDB1 (see above), CAF1 binds several other proteins [87], but their role in epigenetic inheritance is unclear.

## CAF1 and Cancer

In view of their central role in regulating nucleosome density, histone modifications, chromatin structure, genome stability, and epigenetics, it seems logical that cancer cells might manipulate histone chaperones to promote mutagenesis and/or alter gene expression. The field is young, and while most links are indirect, there are some striking examples of how chaperones influence cancer progression. We review some key examples below.

Disrupting the CAF1 specific subunits p150 (CHAF1A) or p60 (CHAF1B), or their equivalents in lower organisms, disrupts nucleosome frequency, and, in many cases, causes spontaneous DNA damage and/or increased susceptibility to DNA damage inducing agents [28, 49, 82]. Moreover, because CAF1 regulates heterochromatin, disrupting this function could lead to large-scale defects in chromosome alignment and segregation at mitosis. However, inactivating mutations in CAF1 subunits have not been reported in cancer, likely because their depletion is lethal in many circumstances [81, 83, 84]. Nevertheless, subtle sequence variants, changes in posttranslational modifications and/or altered levels of chaperones could affect genome stability or gene expression. Indeed, excess CHAF1B/p60 correlates with poor outcome in some cancers [88]. The genetic connection between CAF1 and Polycomb phenotypes in *Drosophila* is also intriguing since excess Polycomb activity is common in human tumors [86].

## ASF1, HIRA, and Cancer

As discussed earlier, ASF1A and B are critical to buffer excess histones during replication stress [89], and disrupting this process promotes a delayed DNA damage response [62]. In addition, ASF1B has unique roles in long-term growth, gene regulation, and chromosomal stability, and elevated ASF1B expression is associated with high mitotic index, poor tumor grade, and worse outcomes in breast cancer [62]. Also, ASF1A is part of a network of factors critical to repress the pro-apoptotic gene Fas in K-ras transformed cells [90]. ASF1 stimulates H3K56 acetylation to promote transfer to the CAF1 complex and subsequent deposition on DNA (see above). This histone modification promotes longevity in yeast [42], and is elevated together with ASF1 in several cancer cell types [40]. H3K56ac is also important at sites of DNA repair [91]. Normally, this mark is rapidly removed upon H3 deposition, but is maintained at repair sites to increase nucleosome “breathing” and facilitate repair complex access to DNA [92, 93]. The connection between ASF1, CAF1, and H3K56ac provides a logical explanation for increased DNA damage in the absence of one or more of these components [91].

The above data imply oncogenic roles for ASF1. Intriguingly, however, ASF1A together with the H3.3-H4 chaperone HIRA is also critical for Ras-induced senescence, an anticancer response to DNA damage that forces permanent cell cycle exit [94]. Cancer cells overcome senescence by inactivating the RB and p53 tumor suppressor pathways, and it is presumably beyond this point that the putative oncogenic roles of ASF1A discussed above come into effect.

The ability of ASF1 to promote senescence is intimately linked with the histone chaperone HIRA. Normally, HIRA deposits H3.3 at genic or telomeric regions [13], but in senescing cells it is required for the formation of Senescence Associated Heterochromatic Foci (SAHF) that silence RB-E2F regulated cell cycle genes [95]. Early in senescing cells, PML bodies, also linked to senescence, contain HIRA and HP1, and although ASF1A is not found in these structures, a HIRA-ASF1A interaction is critical for SAHF formation [94]. Late in senescence the repressive H2A variant macroH2A (which also has a separate role in silencing the X chromosome) is recruited to SAHF, which also requires the HIRA-ASF1A interaction. ASF1B cannot bind HIRA and is not involved in senescence. Neither HIRA nor ASF1A bind macroH2A, thus the chaperone for the latter in senescing cells is unclear, although **APLF** (aprataxin and PNKP like factor) chaperones macroH2A after DNA damage [96]. Whether HIRA-ASF1A dependent macroH2A deposition is linked to H3.3 and/or the ability of ASF1 to dislodge preexisting nucleosomes (see above) is unknown. In yeast, where there is no macroH2A, interaction of the HIRA (Hir1 and Hir2) and Asf1 equivalents promotes silencing at telomeres and repression of histone gene expression [97–101]. Consistent with its pro-senescence function, low levels of ASF1A are associated with longevity in humans [102], but as noted earlier, Asf1 is actually required for longevity in yeast [42].

## DEK, DAXX, H3.3, and Cancer

Apart from HIRA, DEK and DAXX also chaperone H3.3, and strikingly both are mutated in a variety of tumor types, providing direct evidence linking chaperones to cancer.

DEK is a highly abundant protein (almost as abundant as nucleosomes) with several roles in regulating chromatin structure. The *Drosophila* ecdysone hormone binds the nuclear receptor EcR and its ability to promote gene activation depends on DEK [103]. In the same study, DEK was shown to promote nucleosome formation by chaperoning H3.3, which was dependent on CKII phosphorylation of DEK. Disrupting the latter blocked H3.3 binding, nucleosome formation in vitro, and ecdysone-mediated gene activation.

DEK also has roles in repressing transcription. Thus, independent of CKII (and thus its H3.3 chaperone function), DEK promotes heterochromatin through an RNA-dependent interaction with HP1. Loss of DEK results in the concomitant depletion of HP1 and H3K9me3 from either constitutive pericentric heterochromatin or promoters silenced by the latter repressive epigenetic mark [10]. Fruit fly DEK is in the same class as the H3K9 methyl transferase Su(var)3-9, because its inactivation relieves position effect variegation (PEV), the repression linked to translocation of a previously active gene to a location near heterochromatin [10]. Thus, both CAF1 and DEK have key roles in binding HP1 and regulating chromatin.

In line with a repressive function for DEK, it restricts access of complexes required for transcription to chromatin templates in vitro, and it is dislodged by SET, a protein required for Pol II-mediated transcription on such templates [104]. Intriguingly, SET, like DEK, is found as a CAN/NUP214 fusion protein in leukemia [105].

DEK is linked to cancer in two ways. First translocations have been found in a subset of AML that fuse its N-terminus to the C-terminus of the nucleoporin CAN (NUP214) [105]. This translocation switches the location of CAN from the nuclear pore to the nucleoplasm [106]. The DEK-CAN leukemic fusion protein cannot bind CKII, abolishing its histone chaperone activity, arguing that defects in H3.3 deposition may be important for transformation [103], although the critical targets are unclear. The effect of the translocation on HP1 or SET activity is also unclear.

Second, DEK over-expression is observed in various cancers, either through gain of its chromosomal location 6p22 [107–110] or through induction by E2F, such as in cells expressing the RB-inactivating human papilloma virus E7 oncoprotein [111, 112]. DEK over-expression elevates H3K9me3 at silenced chromatin, which could be oncogenic in some contexts [10]. However, the effect of DEK over-expression on chaperoning H3.3 or the ability of SET to open chromatin is unknown. DEK influences a variety of cancer hallmarks, such as survival, senescence, DNA repair, and invasiveness [107, 112–116], but exactly which of these biological functions, if any, require its activating and/or repressive molecular chromatin roles remains unclear. Complicating matters further, DEK has also been linked to other activities such as splicing and protein translation [117–119].

DAXX was first identified as a FAS-binding protein and interacts with other factors that regulate cell survival [120]. Three recent studies have shown that it is a

histone H3.3 chaperone [13–15]. Drane et al. report that DAXX is more tightly associated with H3.3 than HIRA, and that it promotes nucleosome deposition on plasmids in vitro, which was also demonstrated by Lewis et al. Consistent with prior work [121, 122], DAXX and the SWI/SNF-like chromatin remodeling enzyme ATRX (α-thalassemia mental retardation X-linked syndrome) were found at pericentric satellite repeats. These repetitive regions are transcribed, which Drane et al. found requires DAXX, H3.3, and ATRX. DAXX-mediated deposition of H3.3 on pericentric DNA is replication independent and H3.3 co-localizes with DAXX and ATRX in PML bodies, which is blocked if DAXX is missing. DAXX/ATRX-mediated recruitment of H3.3 to PML bodies and pericentric heterochromatin is reminiscent of the requirement for HIRA to transport H3.3 to PML bodies and senescence associated heterochromatin (see above).

Goldberg et al. [13] focused on the genome-wide localization of H3.3, demonstrating that it is enriched at known regulatory elements (active or poised, proximal, or distal), across the bodies of active genes, and at telomeres. HIRA was essential for enrichment at promoters and gene bodies, but not at remote regulatory elements or telomeres. Like Drane et al. they found that H3.3 associates with DAXX/ATRX, and while ATRX was dispensable for H3.3 enrichment at genic regions or regulatory elements, it was essential at telomeres. Moreover, Lewis et al. found that DAXX, like ATRX, is essential for H3.3 deposition at telomeres [15]. Finally, ATRX loss elevates transcription of the telomeric repeat-containing transcript TERRA [13]. Thus, paradoxically, DAXX/ATRX/H3.3 recruitment is linked to induction or downregulation of pericentric or telomeric transcription, respectively [13, 14]. The structure of DAXX bound to a H3.3-H4 dimer has been solved and reveals specific interactions that explain why this chaperone preferentially binds H3.3 rather than replicative H3 isoforms [26].

Recent deep sequencing efforts have exposed direct evidence for a link between DAXX/ATRX and cancer. First, mutations in DAXX or ATRX were reported in almost half of pancreatic neuroendocrine tumors [123]. Consistent with the key role of DAX/ATRX in depositing H3.3 at telomeres, these mutations or loss of nuclear DAXX/ATRX are linked to abnormal telomere structures, and analysis of other cancers with aberrant telomeres exposed ATRX, although not DAXX, mutations in some brain tumors [124]. A subsequent study found both ATRX and DAXX mutations in pediatric glioblastoma samples, but also a high frequency of H3.3 mutations that target Lysine 27 or Glycine 34 [125]. These variants were linked to perturbed telomeres and altered transcription profiles, which is consistent with the H3.3 chaperone function discussed above, and with the link between lysine 27 methylation and Polycomb-mediated repression. Finally, DAXX mutations have also been described in a subset of AML patients [126].

## Future Directions

The term chaperone conjures up rather pedestrian images, yet these proteins have diverse functions, and their deregulation has drastic effects on cell homeostasis. The last decade has seen dramatic advances in our understanding of their roles in

replicating nucleosomes, histone exchange, deposition of histone subtypes, epigenetic inheritance, transcription, and DNA repair, yet there remain many unanswered questions. How, for example, are old and new $(H3\text{-}H4)_2$ tetramers kept apart at the replication fork? What precise mechanisms are used to replicate old H2-H2B dimers or deposit new dimers? Why does CAF1 perturbation have such different effects in distinct species/cell types on DNA damage or survival? Which functions of CAF1 (e.g., recruiting H3-H4, HP1, SETDB1 etc.) are critical in these various contexts? What is the molecular link connecting CAF1 to Polycomb function, and how is this function played out in organisms other than *Drosophila*? The ties between histone chaperones, cancer, and longevity are also fascinating, but we only have a superficial understanding. What is the relevant molecular effect in these biological contexts: DNA damage, gene expression, Polycomb function, heterochromatin, and/or another role, and what are the key molecular players that chaperones influence? What are the signals that switch HIRA from regulation of genic or telomeric regions to silencing of cell cycle genes destined to be buried in SAHF or that control the ability of ASF1 to promote senescence in collaboration with HIRA, versus silencing of pro-apoptotic genes? And what are the precise functions of DEK, DEK fusions, and DAXX that are so critical in driving cancer progression? These are only a few of the questions that will keep researchers busy in the chaperone field for the next decade.

## References


1. Laskey RA, Honda BM, Mills AD, Finch JT. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature. 1978;275:416–20.
2. English CM, Adkins MW, Carson JJ, Churchill ME, Tyler JK. Structural basis for the histone chaperone activity of Asf1. Cell. 2006;127:495–508.
3. Akey CW, Luger K. Histone chaperones and nucleosome assembly. Curr Opin Struct Biol. 2003;13:6–14.
4. Loyola A, Almouzni G. Histone chaperones, a supporting role in the limelight. Biochim Biophys Acta. 2004;1677:3–11.
5. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. Science. 2010;330: 612–6.
6. Hake SB, Allis CD. Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis”. Proc Natl Acad Sci USA. 2006;103:6428–35.
7. Xu M et al. Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. Science. 2010;328:94–8.
8. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389:251–60.
9. Wu RS, Tsai S, Bonner WM. Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell. 1982;31:367–74.
10. Kappes F et al. The DEK oncoprotein is a Su(var) that is essential to heterochromatin integrity. Genes Dev. 2011;25:673–8.
11. Ray-Gallet D et al. HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Mol Cell. 2002;9:1091–100.
12. Graham PJ. Developments in occupational health and safety qualifications. Ann Occup Hyg. 1992;36:663–8.
13. Goldberg AD et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell. 2010;140:678–91.

14. Drane P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev. 2010;24:1253–65.
15. Lewis PW, Elsaesser SJ, Noh KM, Stadler SC, Allis CD. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci USA. 2011;107:14075–80.
16. Campos EI et al. The program for processing newly synthesized histones H3.1 and H4. Nat Struct Mol Biol. 2010;17:1343–51.
17. Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell. 2006;24:309–16.
18. Richardson RT, Alekseev O, Alekseev OM, O'Rand MG. Characterization of the NASP promoter in 3T3 fibroblasts and mouse spermatogenic cells. Gene. 2006;371:52–8.
19. Song JJ, Garlick JD, Kingston RE. Structural basis of histone H4 recognition by p55. Genes Dev. 2008;22:1313–8.
20. Wang H, Walsh ST, Parthun MR. Expanded binding specificity of the human histone chaperone NASP. Nucleic Acids Res. 2008;36:5763–72.
21. Finn RM, Browne K, Hodgson KC, Ausio J. sNASP, a histone H1-specific eukaryotic chaperone dimer that facilitates chromatin assembly. Biophys J. 2008;95:1314–25.
22. Ejlassi-Lassallette A, Mocquard E, Arnaud MC, Thiriet C. H4 replication-dependent diacetylation and Hat1 promote S-phase chromatin assembly in vivo. Mol Biol Cell. 2011;22:245–55.
23. Zhang H, Han J, Kang B, Burgess R, Zhang Z. Human histone acetyltransferase 1 protein preferentially acetylates H4 histone molecules in H3.1-H4 over H3.3-H4. J Biol Chem. 2012;287:6573–81.
24. Cook AJ, Gurard-Levin ZA, Vassias I, Almouzni G. A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain. Mol Cell. 2011;44:918–27.
25. Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell. 2004;116:51–61.
26. Elsasser SJ et al. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. Nature. 2012;491:560–5.
27. Smith S, Stillman B. Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell. 1989;58:15–25.
28. Kaufman PD, Kobayashi R, Stillman B. Ultraviolet radiation sensitivity and reduction of telomeric silencing in Saccharomyces cerevisiae cells lacking chromatin assembly factor-I. Genes Dev. 1997;11:345–57.
29. Bulger M, Ito T, Kamakaka RT, Kadonaga JT. Assembly of regularly spaced nucleosome arrays by Drosophila chromatin assembly factor 1 and a 56-kDa histone-binding protein. Proc Natl Acad Sci USA. 1995;92:11726–30.
30. Kamakaka RT, Bulger M, Kaufman PD, Stillman B, Kadonaga JT. Postreplicative chromatin assembly by Drosophila and human chromatin assembly factor 1. Mol Cell Biol. 1996;16:810–7.
31. Tyler JK et al. Interaction between the Drosophila CAF-1 and ASF1 chromatin assembly factors. Mol Cell Biol. 2001;21:6574–84.
32. Malay AD, Umehara T, Matsubara-Malay K, Padmanabhan B, Yokoyama S. Crystal structures of fission yeast histone chaperone Asf1 complexed with the Hip1 B-domain or the Cac2 C terminus. J Biol Chem. 2008;283:14022–31.
33. Nowak AJ et al. Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. J Biol Chem. 2011;286:23388–96.
34. Nakano S, Stillman B, Horvitz HR. Replication-coupled chromatin assembly generates a neuronal bilateral asymmetry in *C. elegans*. Cell. 2011;147:1525–36.
35. Winkler DD, Zhou H, Dar MA, Zhang Z, Luger K. Yeast CAF-1 assembles histone (H3-H4)2 tetramers prior to DNA deposition. Nucleic Acids Res. 2012;40:10139–49.

36. Liu WH, Roemer SC, Port AM, Churchill ME. CAF-1-induced oligomerization of histones H3/H4 and mutually exclusive interactions with Asf1 guide H3/H4 transitions among histone chaperones and DNA. Nucleic Acids Res. 2011;40:11229–39.
37. Liu WH, Churchill ME. Histone transfer among chaperones. Biochem Soc Trans. 2012;40: 357–63.
38. Tsubota T et al. Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. Mol Cell. 2007;25:703–12.
39. Han J, Zhou H, Li Z, Xu RM, Zhang Z. The Rtt109-Vps75 histone acetyltransferase complex acetylates non-nucleosomal histone H3. J Biol Chem. 2007;282:14158–64.
40. Das C, Lucia MS, Hansen KC, Tyler JK. CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature. 2009;459:113–7.
41. Li Q et al. Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell. 2008;134:244–55.
42. Feser J et al. Elevated histone expression promotes life span extension. Mol Cell. 2010;39:724–35.
43. Kang B et al. Phosphorylation of H4 Ser 47 promotes HIRA-mediated nucleosome assembly. Genes Dev. 2011;25:1359–64.
44. Burgess RJ, Zhang Z. Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol. 2013;20:14–22.
45. Lopez de Saro FJ. Regulation of interactions with sliding clamps during DNA replication and repair. Curr Genomics. 2009;10(206–15).
46. Shibahara K, Stillman B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell. 1999;96:575–85.
47. Zhang Z, Shibahara K, Stillman B. PCNA connects DNA replication to epigenetic inheritance in yeast. Nature. 2000;408:221–5.
48. Moggs JG et al. A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. Mol Cell Biol. 2000;20:1206–18.
49. Ye X et al. Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. Mol Cell. 2003;11:341–51.
50. Ransom M et al. FACT and the proteasome promote promoter chromatin disassembly and transcriptional initiation. J Biol Chem. 2009;284:23461–71.
51. Winkler DD, Luger K. The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization. J Biol Chem. 2011;286:18369–74.
52. Gambus A et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol. 2006;8:358–66.
53. Lorch Y, Maier-Davis B, Kornberg RD. Chromatin remodeling by nucleosome disassembly in vitro. Proc Natl Acad Sci USA. 2006;103:3090–3.
54. Park YJ, Chodaparambil JV, Bao Y, McBryant SJ, Luger K. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. J Biol Chem. 2005;280: 1817–25.
55. Park YJ, Luger K. Structure and function of nucleosome assembly proteins. Biochem Cell Biol. 2006;84:549–58.
56. Worcel A, Han S, Wong ML. Assembly of newly replicated chromatin. Cell. 1978;15: 969–77.
57. Groth A et al. Regulation of replication fork progression through histone supply and demand. Science. 2007;318:1928–31.
58. Tyler JK et al. The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature. 1999;402:555–60.
59. Natsume R et al. Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. Nature. 2007;446:338–41.
60. Smith S, Stillman B. Stepwise assembly of chromatin during DNA replication in vitro. EMBO J. 1991;10:971–80.
61. Bowman A et al. The histone chaperones Nap1 and Vps75 bind histones H3 and H4 in a tetrameric conformation. Mol Cell. 2011;41:398–408.

62. Corpet A et al. Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. EMBO J. 2011;30:480–93.
63. Mousson F, Ochsenbein F, Mann C. The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways. Chromosoma. 2007;116:79–93.
64. Zeng W, Ball Jr AR, Yokomori K. HP1: heterochromatin binding proteins working the genome. Epigenetics. 2010;5:287–92.
65. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher 3rd FJ. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 2002;16:919–32.
66. Schotta G et al. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. 2002;21:1121–31.
67. Quivy JP, Grandi P, Almouzni G. Dimerization of the largest subunit of chromatin assembly factor 1: importance in vitro and during Xenopus early development. EMBO J. 2001;20:2015–27.
68. Ono T et al. Chromatin assembly factor 1 ensures the stable maintenance of silent chromatin states in Arabidopsis. Genes Cells. 2006;11:153–62.
69. Schonrock N, Exner V, Probst A, Gruissem W, Hennig L. Functional genomic analysis of CAF-1 mutants in Arabidopsis thaliana. J Biol Chem. 2006;281:9560–8.
70. Song Y et al. CAF-1 is essential for Drosophila development and involved in the maintenance of epigenetic memory. Dev Biol. 2007;311:213–22.
71. Dohke K et al. Fission yeast chromatin assembly factor 1 assists in the replication-coupled maintenance of heterochromatin. Genes Cells. 2008;13:1027–43.
72. Enomoto S, McCune-Zierath PD, Gerami-Nejad M, Sanders MA, Berman J. RLF2, a subunit of yeast chromatin assembly factor-I, is required for telomeric chromatin function in vivo. Genes Dev. 1997;11:358–70.
73. Monson EK, de Bruin D, Zakian VA. The yeast Cac1 protein is required for the stable inheritance of transcriptionally repressed chromatin at telomeres. Proc Natl Acad Sci USA. 1997;94:13081–6.
74. Enomoto S, Berman J. Chromatin assembly factor I contributes to the maintenance, but not the re-establishment, of silencing at the yeast silent mating loci. Genes Dev. 1998;12:219–32.
75. Taddei A, Roche D, Sibarita JB, Turner BM, Almouzni G. Duplication and maintenance of heterochromatin domains. J Cell Biol. 1999;147:1153–66.
76. Quivy JP et al. A CAF-1 dependent pool of HP1 during heterochromatin duplication. EMBO J. 2004;23:3516–26.
77. Quivy JP, Gerard A, Cook AJ, Roche D, Almouzni G. The HP1-p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. Nat Struct Mol Biol. 2008;15:972–9.
78. Reese BE, Bachman KE, Baylin SB, Rountree MR. The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Mol Cell Biol. 2003;23:3226–36.
79. Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell. 2004;15:595–605.
80. Loyola A et al. The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep. 2009;10:769–75.
81. Houlard M et al. CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. PLoS Genet. 2006;2:e181.
82. Hoek M, Stillman B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. Proc Natl Acad Sci USA. 2003;100:12183–8.
83. Nabatiyan A, Krude T. Silencing of chromatin assembly factor 1 in human cells leads to cell death and loss of chromatin assembly during DNA synthesis. Mol Cell Biol. 2004;24:2853–62.
84. Takami Y, Ono T, Fukagawa T, Shibahara K, Nakayama T. Essential role of chromatin assembly factor-1-mediated rapid nucleosome assembly for DNA replication and cell division in vertebrate cells. Mol Biol Cell. 2007;18:129–41.

85. Poleshko A et al. Identification of a functional network of human epigenetic silencing factors. J Biol Chem. 2010;285:422–33.
86. Lanzuolo C, Orlando V. Memories from the polycomb group proteins. Annu Rev Genet. 2012;46:561–89.
87. Hoek M, Myers MP, Stillman B. An analysis of CAF-1-interacting proteins reveals dynamic and direct interactions with the KU complex and 14-3-3 proteins. J Biol Chem. 2011;286: 10876–87.
88. Polo SE et al. Clinical significance and prognostic value of chromatin assembly factor-1 overexpression in human solid tumours. Histopathology. 2010;57:716–24.
89. Groth A et al. Human Asf1 regulates the flow of S phase histones during replicational stress. Mol Cell. 2005;17:301–11.
90. Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. An elaborate pathway required for Ras-mediated epigenetic silencing. Nature. 2007;449:1073–7.
91. Avvakumov N, Nourani A, Cote J. Histone chaperones: modulators of chromatin marks. Mol Cell. 2011;41:502–14.
92. Masumoto H, Hawke D, Kobayashi R, Verreault A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature. 2005;436:294–8.
93. Neumann H et al. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell. 2009;36:153–63.
94. Zhang R et al. Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. Dev Cell. 2005;8:19–30.
95. Rai TS, Adams PD. Lessons from senescence: chromatin maintenance in non-proliferating cells. Biochim Biophys Acta. 2012;1819:322–31.
96. Mehrotra PV et al. DNA repair factor APLF is a histone chaperone. Mol Cell. 2011; 41:46–55.
97. Spector MS, Raff A, DeSilva H, Lee K, Osley MA. Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the Saccharomyces cerevisiae cell cycle. Mol Cell Biol. 1997;17:545–52.
98. Sharp JA, Fouts ET, Krawitz DC, Kaufman PD. Yeast histone deposition protein Asf1p requires Hir proteins and PCNA for heterochromatic silencing. Curr Biol. 2001;11:463–73.
99. Sutton A, Bucaria J, Osley MA, Sternglanz R. Yeast ASF1 protein is required for cell cycle regulation of histone gene transcription. Genetics. 2001;158:587–96.
100. Daganzo SM et al. Structure and function of the conserved core of histone deposition protein Asf1. Curr Biol. 2003;13:2148–58.
101. Fillingham J et al. Two-color cell array screen reveals interdependent roles for histone chaperones and a chromatin boundary regulator in histone gene repression. Mol Cell. 2009;35:340–51.
102. Passtoors WM et al. Transcriptional profiling of human familial longevity indicates a role for ASF1A and IL7R. PLoS One. 2012;7:e27759.
103. Sawatsubashi S et al. A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes Dev. 2010;24:159–70.
104. Gamble MJ, Fisher RP. SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery. Nat Struct Mol Biol. 2007;14:548–55.
105. von Lindern M, Breems D, van Baal S, Adriaansen H, Grosveld G. Characterization of the translocation breakpoint sequences of two DEK-CAN fusion genes present in t(6;9) acute myeloid leukemia and a SET-CAN fusion gene found in a case of acute undifferentiated leukemia. Genes Chromosomes Cancer. 1992;5:227–34.
106. Fornerod M et al. Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. Oncogene. 1995;10:1739–48.
107. Khodadoust MS et al. Melanoma proliferation and chemoresistance controlled by the DEK oncogene. Cancer Res. 2009;69:6405–13.
108. Han S et al. Clinicopathological significance of DEK overexpression in serous ovarian tumors. Pathol Int. 2009;59:443–7.

109. Orlic M, Spencer CE, Wang L, Gallie BL. Expression analysis of 6p22 genomic gain in retinoblastoma. Genes Chromosomes Cancer. 2006;45:72–82.
110. Shibata T et al. DEK oncoprotein regulates transcriptional modifiers and sustains tumor initiation activity in high-grade neuroendocrine carcinoma of the lung. Oncogene. 2010;29: 4671–81.
111. Carro MS et al. DEK Expression is controlled by E2F and deregulated in diverse tumor types. Cell Cycle. 2006;5:1202–7.
112. Wise-Draper TM et al. The human DEK proto-oncogene is a senescence inhibitor and an upregulated target of high-risk human papillomavirus E7. J Virol. 2005;79:14309–17.
113. Wise-Draper TM et al. Apoptosis inhibition by the human DEK oncoprotein involves interference with p53 functions. Mol Cell Biol. 2006;26:7506–19.
114. Kavanaugh GM et al. The human DEK oncogene regulates DNA damage response signaling and repair. Nucleic Acids Res. 2011;39:7465–76.
115. Privette Vinnedge LM et al. The human DEK oncogene stimulates beta-catenin signaling, invasion and mammosphere formation in breast cancer. Oncogene. 2011;30:2741–52.
116. Wise-Draper TM et al. Overexpression of the cellular DEK protein promotes epithelial transformation in vitro and in vivo. Cancer Res. 2009;69:1792–9.
117. McGarvey T et al. The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. J Cell Biol. 2000;150:309–20.
118. Soares LM, Zanier K, Mackereth C, Sattler M, Valcarcel J. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. Science. 2006;312:1961–5.
119. Ageberg M, Drott K, Olofsson T, Gullberg U, Lindmark A. Identification of a novel and myeloid specific role of the leukemia-associated fusion protein DEK-NUP214 leading to increased protein synthesis. Genes Chromosomes Cancer. 2008;47:276–87.
120. Salomoni P, Khelifi AF. Daxx: death or survival protein? Trends Cell Biol. 2006;16:97–104.
121. McDowell TL et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc Natl Acad Sci USA. 1999;96:13983–8.
122. Ishov AM, Vladimirova OV, Maul GG. Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. J Cell Sci. 2004;117:3807–20.
123. Jiao R et al. Physical and functional interaction between the Bloom's syndrome gene product and the largest subunit of chromatin assembly factor 1. Mol Cell Biol. 2004;24:4710–9.
124. Heaphy CM et al. Altered telomeres in tumors with ATRX and DAXX mutations. Science. 2011;333:425.
125. Schwartzentruber J et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482:226–31.
126. Ding L et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481:506–10.

# Evaluating Chromatin Regulators in Cancer Cell Lines Through RNAi-Glasses

**Troy Ketela, Kevin R. Brown, and Jason Moffat**

**Abstract** Abnormalities in chromatin organization have been linked to numerous human disease states including cancer, and RNA interference (RNAi) screens are a powerful approach for discovering gene function in transformed mammalian cell cultures. Herein, we review the use of RNAi technologies for characterization of the roles played by chromatin-related genes in cancer cell lines. We also highlight observations on the essentiality of components of epigenetic machinery within a large collection of genome-wide shRNA screens performed in breast, ovarian, and pancreatic cancer cell lines.

**Keywords** Chromatin • Cancer • RNAi • siRNA • shRNA • Screen • Acetylation • Methylation • Histone

## Background

### *Introduction to RNAi*

RNA interference (RNAi), which was discovered in *Caenorhabditis elegans* [1], has been harnessed and exploited for the purpose of studying gene function across a range of model systems. RNAi-based approaches have evolved to become one of

T. Ketela, Ph.D. • K.R. Brown, Ph.D.
Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada, M5S3E1
e-mail: troy.ketela@utoronto.ca

J. Moffat, Ph.D. (✉)
Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada, M5S3E1

Banting and Best Department of Medical Research and Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
e-mail: j.moffat@utoronto.ca

A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4_16,
© Springer Science+Business Media New York 2014

the prominent and powerful tools in functional genetic studies in a variety of eukaryotic organisms [2]. The RNAi machinery processes endogenously or exogenously produced double-stranded RNA (dsRNA) into 21–23 nucleotide small interfering RNAs (siRNAs), which are incorporated into the RNA-induced silencing complex (RISC) [3]. Subsequent recognition of target mRNAs by the RISC-siRNA complex results in suppression of expression of the target gene product by either translational inhibition or mRNA destruction [1, 4, 5]. Once the existing protein molecules translated from target gene transcripts have been turned over by the cell, target gene function is effectively ablated. Critically, large-scale genome sequencing efforts unlocked the potential for leveraging RNAi, advances in molecular biology, robotics, and computational biology for genome-scale construction of RNAi libraries targeting mouse and human cells [6–9].

RNAi technologies are most commonly employed in two ways; targeted knockdown of small numbers of genes (often one or two) to facilitate direct investigation of a hypothesis (e.g., confirm the cellular phenotype for loss of a target gene's activity) or screens of a set of RNAi reagents to identify genes that influence a phenotype of interest. The target gene set is predetermined by the researcher via some sort of selection criteria in focused screens, while unbiased RNAi screens may cover the entire genome. The utility of RNAi for performing fast and facile functional genetic screens in mammalian cells has led to numerous discoveries in a variety of disease areas including cancer [10–13]. Although RNAi has opened the gates for systematic loss-of-function studies and screens in human cells, realizing the full potential of siRNA and shRNA technologies requires additional development to address the pitfalls [14–16].

## *Employing RNAi to Study Chromatin Regulation*

Genomic instability is a hallmark of cancer [17], and it is becoming clear that epigenetic regulation is a key feature of malignant cells which supports plasticity and adaptive evolution of tumors (reviewed in [18]). As a consequence, there have been significant efforts to identify and develop inhibitors of certain chromatin regulators such as histone deacetylases (HDACs) and histone methyltransferases as possible treatments for a variety of types of cancer [19]. New understanding of the roles of specific chromatin related-proteins in tumorigenesis and metastasis and the suitability of certain chromatin regulators as potential targets for therapeutic intervention may be gained by employing functional genetic tools such as RNAi in cancer models.

In principle, there are two main approaches to utilizing RNAi reagents in moderate to high throughput applications with mammalian cells: arrayed or indexed screens and pooled screens. The choice between the methods is dictated by several factors including the number of query genes and cell lines one wishes to screen, the cellular phenotype to be quantified, time constraints, available equipment resources, and available budget. In this chapter, we will review some of the innovative RNAi

screens that have been carried out using either arrayed or pooled screens and will explore the potential of unbiased genome-scale pooled shRNA screens for the identification of potential chromatin regulators that are essential for proliferation in cancer cell lines.

## *Arrayed RNAi Screening*

Systematic application of RNAi reagents to cells that are dispensed into physically separated wells is referred to as arrayed RNAi screening. In general, a single gene is targeted per well in arrayed screens and deconvolution of screening data simply requires a well map that indicates the RNAi contents of each well. Both chemically synthesized siRNAs and virally delivered shRNAs are suitable for arrayed screening applications. The choice between reagents is determined by cell amenability to siRNA transfection or shRNA viral transduction, and whether transient or stable knockdown is required. Simple readouts such as quantitative measurement of cell viability by dye staining may be performed in arrayed RNAi screens; however, arrayed screens are also well suited for high-content screening techniques [20]. For example, cells may be enumerated by nuclear DNA staining, proteins may be labeled by antibodies or conjugated fluorescent proteins (e.g., Alexa-fluor labeled anti-phospho histone H3 antibodies), subcellular structures may be visualized by dye staining (e.g., Mitotracker™), and gross cellular morphology imaged via transmitted light or fluorescent imaging. Sophisticated image analysis software is generally employed to evaluate statistical data compiled from analysis of the microscopic images. Drawbacks of arrayed RNAi screens include high reagent cost (microplates, RNAi reagents, antibodies, and stains), requirement for automated liquid handling and microscopy devices, and relatively low throughput.

There is wide commercial availability for antibodies that recognize epigenetic marks such as DNA methylation and the various states of histone modification by ubiquitin ligases, protein kinases, acetyltransferases, and methyltransferases. These types of antibodies may be utilized as readouts in genetic screens that employ RNAi to identify genes that affect chromatin structure and regulation. For example, Bjorkman et al. [21] systematically assessed RNAi-induced phenotypes for 615 human chromatin-related genes in human prostate cancer cells using an arrayed RNAi approach in search of genes that affect proliferation, chromatin marks, and androgen receptor expression. Using a reverse transfection method, the group spotted 1,328 distinct features (2 siRNAs per gene plus controls) each containing a different siRNA/transfection reagent/Matrigel mixture onto plastic cell culture surfaces. Suspended VCaP human prostate cancer cells were applied to the array, and after a short incubation period, non-adhered cells were washed away. Forty-eight hours after seeding, the cell arrays were fixed, permeabilized, and stained for nuclear DNA content, proliferation (Ki67), apoptosis (cPARP), and androgen receptor (AR) expression. Total cellular levels of H3K4me2 and H3K18ac (which typically accompany active promoters and transcriptional regions), H3K9me2 (typically a mark of transcriptional repression),

and H4K16ac (which affects higher order chromatin structure) were also measured by antibody staining. Z-scores were calculated from signal intensity for all measured aspects of each spot, and the resulting hit data were visualized in a heat map generated in the R statistical language by clustering the Z-scores via the partitioning around medoid (PAM) method. Enrichments for specific protein domains within each cluster were then identified. The authors noted that the cluster where acetylation of H3K18 and H4K16 was strongly reduced contained most genes encoding an HDAC domain [21]. Knockdown of proteins with HDAC function might be expected to result in hyperacetylation and not hypoacetylation of histones; however, consistent with these observations, lack of induced histone hyperacetylation in HDAC-knockdown or HDAC inhibitor-treated cells has been reported [22–24]. The concomitant decrease in cell proliferation as measured by Ki67 staining and expression of the AR in the same cluster may indicate a general lethality induced by siRNAs in this cluster. Other observations include enrichments for JmjC domain containing proteins (histone demethylases) in a cluster where diminished cell proliferation coincided with enhanced H3K9 dimethylation, and enrichment for high mobility group (HMG) proteins and PHD-finger proteins in a cluster where siRNAs induced higher cell proliferation and caused increases in H3K4 methylation, H3K18 and H4K16 acetylation, and AR expression. Reasons as to why knockdown of specific HMG and PHD-finger proteins would enhance cell proliferation are not clear. Nevertheless, these observations lead to the proposal that HDACs and JmjC domain containing lysine demethylases may be attractive enzyme classes to pursue for development of therapeutics to treat prostate cancer, since siRNA knockdown of several of these genes reduced AR expression and cell proliferation [21]. Additional analysis of the histone demethylase gene PHF8 revealed that it was frequently over-expressed in prostate cancer samples [21]. Knock down and overexpression of PHF8 in prostate cancer cells suggests that it influences cancer cell migration as well as invasion and apoptosis, possibly through transcriptional changes.

## *Pooled RNAi Screening*

Although at least one group has generated an shRNA library that features a separate molecular barcode sequence for each shRNA [9], it is not necessary to do so. shRNAs encoded in viral vectors become stably integrated into the host genome; therefore, each cell expressing an shRNA carries a heritable, quantifiable marker—the shRNA sequence itself. This property enables experiments where viral shRNA particle mixtures or pools comprised of anywhere from tens to tens of thousands of unique shRNA sequences may be tested. Pooled screens are suitable for in vitro cell culture systems as well as in vivo xenograft models where parallel examination of gene function in the context of the tumor environment is desirable [25, 26]. In addition to simple cell proliferation under standard culture conditions, pooled screens may be performed to identify shRNAs that confer sensitivity or resistance to any molecule or growth condition, or that alter a phenotype that may be selected (e.g., expression of a cell surface marker trackable by FACS). Comparison of

discrete time points, the endpoint, or sorted populations with an experimental reference population identifies changes within the shRNA population that reflect the phenotypes that each shRNA induces; benign shRNAs remain at stable levels within the population; beneficial shRNAs (e.g., promote proliferation) become more abundant, and deleterious shRNAs (e.g., impede proliferation or kill the cell) become depleted from the cell population. Pooled shRNA screens are generally quantified by preparing genomic DNA from cells, amplifying the shRNA sequences (or co-joined marker sequences) by PCR, and subjecting the samples to custom DNA barcode microarrays or next-generation sequencing approaches.

Acute myeloid leukemia (AML) is a hematological malignancy that has often been found to have disruptions in normal chromatin organization [27]. To identify essential components of the epigenetic regulation machinery in AML, Zuber et al. assembled a custom retroviral shRNA library of ~1,100 shRNAs targeting 243 mouse genes known to be involved in chromatin regulation [28]. The pooled library was transduced into an AML mouse model cell line driven by the MLL-AF9 fusion gene and activated $Nras^{G12D}$, and using a deep sequencing approach, antiproliferative shRNAs were identified by their decrease in abundance at the experimental endpoint compared to the reference population [28]. Of particular note was that depletion of the bromodomain containing BET family member Brd4 caused cell cycle arrest and apoptosis in multiple AML cell lines while counter screens in murine embryonic fibroblasts (MEFs) indicated that Brd4 depletion had only modest effects on cell cycle progression, suggesting that AML may have a special dependency on Brd4 activity [28]. Inhibition of Brd4 also altered the morphology of MLL-AF9 $Nras^{G12D}$ cells, causing a shift to a macrophage-like appearance from myelomonocytic blasts accompanied by changes in cell surface markers consistent with myeloid differentiation [28]. Interestingly, the authors observe that maintenance of the AML state by Brd4 is through Brd4-dependent transcriptional activation of Myc, as Brd4 knockdown causes a decrease in Myc transcript and protein levels, and expression of Myc from a heterologous promoter can sustain the AML phenotype in the presence of a Brd4 knockdown [28]. Intriguingly, the observations connecting Brd4 to AML growth were also confirmed using chemical genetics with JQ1 [29], a compound that is a competitive binder to the acetyl-lysine recognition pocket of BET bromodomains, suggesting that targeting of epigenetic machinery may be a viable strategy for treatment of AML [28].

## Systematic Evaluation of Chromatin Regulators in Cancer Cells

### *Functional Dissection of Lysine Deacetylases in a Cancer Cell Line Model*

Phenotypic screening to detect synthetic genetic relationships has been a powerful approach to identifying functional relationships between genes in model organisms [30, 31]. Although a collection of systematic gene deletions similar to that available

for *Saccharomyces cerevisiae* and other model systems is not currently available for mammalian cell lines, genome-wide RNAi libraries provide a similar resource that may be leveraged for synthetic-genetic screens to study families or classes of genes. For example, using an unbiased genome-wide pooled shRNA approach, a group led by Jef Boeke [32] identified synthetic genetic relationships using a set of query cell lines derived from HCT116 colorectal cancer cells where each cell line or query carried a stable knockdown for 1 of 12 human KDACS (HDAC1-4, HDAC6-9, SIRT1-3, SIRT5) [32]. Genes with multiple shRNAs that significantly inhibited or enhanced proliferation specifically in the context of one of the 12 depleted KDACs were identified and confirmed directly in a cell viability assay [32]. Consistent with model organism findings, query KDACs of the same classes were observed to co-cluster when the genetic interaction data from the screens subjected to hierarchical clustering, and deleterious synthetic genetic interactions were observed for pairs of KDACs that were presumed to have overlapping activities [32]. Notably, the authors speculated that genetic interactions may reveal specific enzyme–substrate relationships and hypothesized that the alpha-1 catalytic subunit of AMP-activated protein kinase PRKAA1 is a substrate for HDAC1 based on the observation that PRKAA1 shows a negative genetic interaction with HDAC1 [32]. Specific acetylation/deacetylation of three lysine residues of PRKAA1 was shown to be reciprocally mediated by p300 and HDAC1, and opposing acetylation and phosphorylation states of PRKAA1 were proposed to mediate its activity level in response to the cell's metabolic state; glucose deprivation promotes PRKAA1 phosphorylation and activation, stimulating catabolic pathways via AMPK kinase activity, while high glucose levels promote PRKAA1 acetylation, inhibiting interaction with its activating kinase LKB1, and thereby reducing AMPK activity [32]. This study demonstrated the potential for RNAi-enabled synthetic genetic screens performed in transformed cell lines to uncover new enzyme–substrate relationships amongst the component machinery of the chromatin regulatory network.

### *Mining Genome-Scale RNAi Screen Data to Identify Chromatin Regulators Important for Cancer Cell Proliferation*

Several groups have employed genome-scale pooled shRNA screens in mammalian cell lines to identify gene knockdowns that reduce cell proliferation or viability [6, 10–13, 33, 34]. Such screens may uncover novel findings in tumor cell biology, and may be useful for the identification of potential new drug targets in specific tumor types. Since genome-scale RNAi screens are largely unbiased and cover many or most known genes in the genome, data for subsets of genes of interest (GOI) may be conveniently mined from these resources and studied in specific contexts.

Recently, we published the results of a collaborative effort where 72 epithelial cancer cell lines derived from human breast, ovarian, and pancreatic tissues were screened with an shRNA pool consisting of ~78,000 shRNAs covering ~16,000 genes [11]. Since aberrations in the posttranslational modification of histones have been associated with tumorigenesis and malignant cell growth and certain

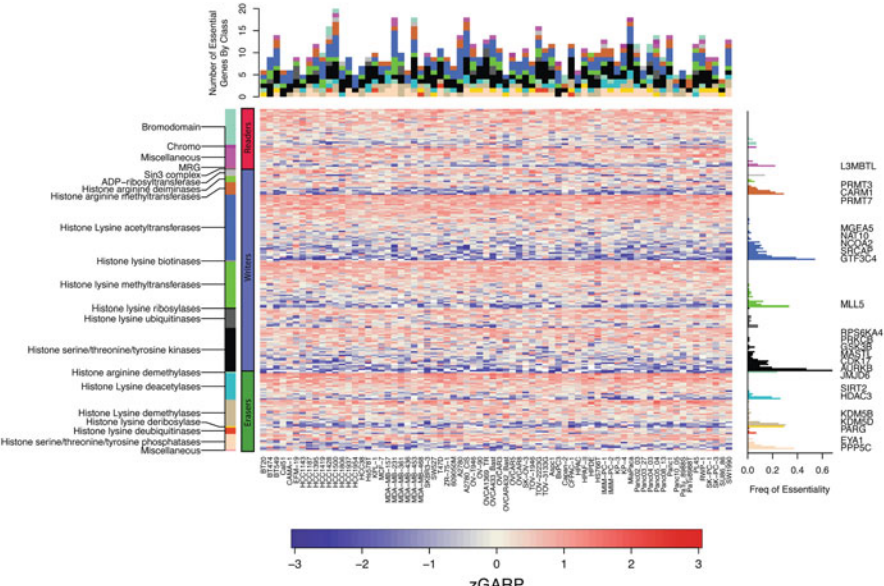



**Fig. 1** A heat map of Z-normalized GARP scores was generated for an annotated set of histone "reader," "writer," and "eraser" genes assembled from various sources [21, 35–37] for a published collection of genome-scale shRNA screens in 72 breast, ovarian, and pancreatic cancer cell lines [11]. Genes with significant zGARP values ($p < 0.05$) in 15 % or more of the cell lines screened are annotated on the right vertical plot axis

components of the chromatin regulatory machinery have been proposed as putative oncology drug targets, we visualized patterns of chromatin-related gene essentiality within the data set described above (Fig. 1 and [11]). All results reported here are derived from GARP scores of primary RNAi screening data [11], where potential hits have not entered systematic validation studies, and false positive and false negative results cannot be ruled out. We examined chromatin-related genes based on their classification as either readers, writers, or erasers and made two general observations. First, the readers consisting mostly of bromo- and chromo-domain containing proteins, were underrepresented for genes that are essential for proliferation in cancer cell lines compared to the writers and erasers (Fig. 1). Second, the writers consisting mostly of histone arginine methyltransferases, histone lysine acetyaltransferases, histone lysine methyltransferases, histone lysine ubiquitinases, and histone serine/threonine/tyrosine kinases, and erasers, consisting mostly of histone lysine deacetylases, histone lysine demethylases, histone lysine deubiquitinases, and histone serine/threonine/tyrosine phosphatases, contained genes that were required for fitness across many cell lines, as well as differentially required for fitness between different cell lines. A major goal moving forward in cancer research is to understand the underlying genomic architectures of cancer cells that drive contextual fitness requirements. This kind of analysis reveals putative fitness genes that can be leveraged for targeted therapeutics.

Amongst genes classified as histone writers, we observe that as a class, lysine methyltransferases appear to be less important for fitness than lysine acetyltransferases and histone serine/threonine kinases across the cell lines screened. The implications of this observation are unclear; however, we speculate that this phenomenon might result from the fact many histone lysine acetyltransferases and kinases have biologically important enzymatic activities towards a plethora of substrates other than histones (reviewed in [38, 39]), while a relatively small number of proteins besides histones have been reported to be modified by lysine methylation (reviewed in [40]). Cells may not be able to compensate for loss of single histone lysine acetyltransferase and serine/threonine kinase gene activities due to multiple important targets requiring the modification. Conversely, except in instances where histone lysine methyltransferases may be acting as oncogenes (e.g., MLL gene rearrangements in leukemias), loss of single gene activities either may not be lethal, or may be compensated for by other family members.

Arginine methylation is mediated by a relatively small group of nine enzymes and occurs on several sites in histones (H2AR3, H3 (R2, R8, R17, R26), and H4R3) and many other nonhistone proteins (reviewed in [41, 42]). Histone arginine methylation is proposed to regulate DNA packaging, transcription regulation, pre-mRNA splicing, mRNA stability, and the propagation of other epigenetic marks [43]. In addition, arginine methylation of nonhistone proteins also influences regulation of transcription [44]. Interestingly, we observe that six out of eight histone arginine methyltransferases targeted in our screens had at least one instance of essentiality, and three genes (PRMT3, PRMT7, and CARM1) were important for fitness in at least 15 % of screens (Fig. 1). Notably, it has been reported that overexpression of CARM1/PRMT4 is required for the development of prostate adenocarcinoma as well as androgen-independent prostate carcinoma [45]. Future studies will determine whether this class is an important driver of cancer cell proliferation across a broad spectrum of tumor types and whether targeting these enzymes with small molecules forms the basis of a potent cancer therapy.

### *Integrating Essential Gene Profiles with Protein Complexes*

Interestingly, many chromatin regulators function as part of large multi-protein complexes, where proteins dynamically interact with each other to regulate changes in chromatin state, DNA replication, and transcription. GOI representing chromatin regulators were compiled from The Histome Infobase [35], published lists of epigenetic regulators [21, 36], and by searching Gene Ontology (GO) categories for known protein complexes to establish protein complexes involved in chromatin regulation. Complexes involving all of the GOI were identified in the CORUM database [46] and GO "cellular component" categories (GO_CC). As the same complex or complex subunit may be named differently in different data sources, the list of protein complexes was manually reduced from 661 annotations to 117 common names. We developed a Java program to convert the list of genes in each complex

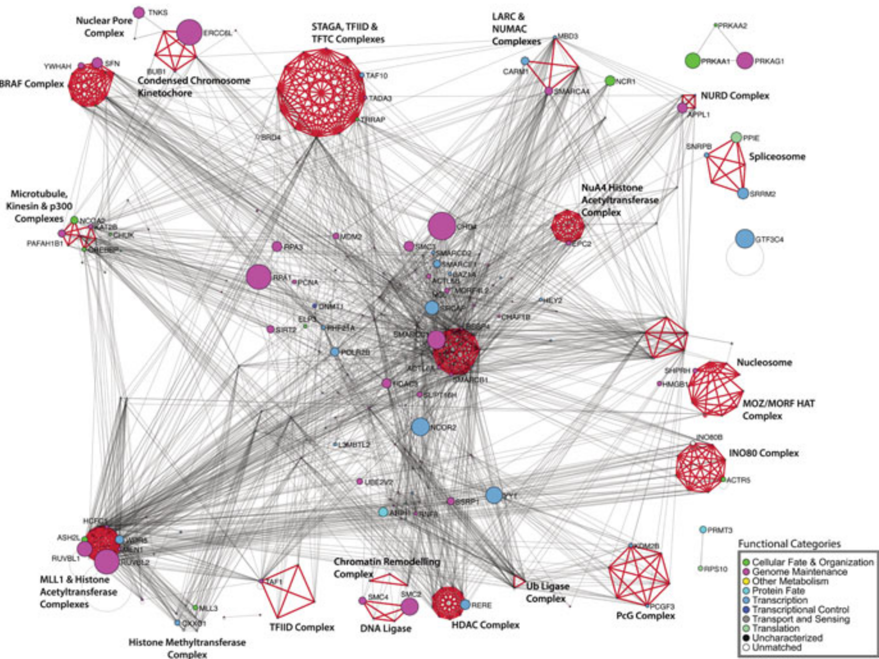



**Fig. 2** Integrated network of chromatin-related protein complexes with essential gene profiles overlaid to highlight chromatin regulators that are highly essential across a compendium of pooled shRNA screens in 72 different breast, pancreas, and ovarian cancer cell lines [11]

into a list of binary protein pairs describing protein co-complex membership. That is, each gene in a protein complex was modeled as being connected to every other gene in the complex. The result is that, when visualized, protein complexes will assemble into tightly interconnected "cliques" using force-directed layout algorithms. The resulting network was visualized using NAViGaTOR v.2.2 (Fig. 2) [47]. NAViGaTOR's clique-finding algorithm identified protein complexes in the graph, which were then manually positioned to improve clarity (Fig. 2). Nodes in the graph, which represent genes/proteins, were automatically colored according to 15 broad GO functional categories, and sized in proportion to the number of shRNA screens where the gene was found to be essential; larger nodes are more often essential than smaller nodes. Nodes essential in at least 5 % of the 72 genome-scale shRNA screens ($p < 0.05$) [11] were labeled with their official gene symbols. A number of complexes have multiple members that are essential for proliferation in cancer cell lines. For example, the INO80 complex contained a number of essential genes for cancer cell proliferation including the transcriptional regulators RUVBL1 and RUVBL2, INO80B, and ACTR5 (Fig. 2). This type of visualization can be informative for selecting targets for secondary assays where one wishes to confirm the potential of a given protein complex for the development of small molecule inhibitors.

# Conclusions

When employed with careful controls and confirmation assays, RNAi technologies are powerful tools for probing gene function in mammalian cancer cells. When combined with well-implemented assays that report on the cellular status of chromatin structure and regulation, RNAi screens have the potential to uncover a wealth of new information about the genes required to maintain and regulate chromatin. Using RNAi, cellular sensitivity to loss of known chromatin-related proteins may be rapidly explored in a variety of cancer models. RNAi-mediated functional genetic studies will be essential in the race to translate small molecule inhibitors of chromatin-related processes into treatments for cancer.

**Acknowledgments** We would like to apologize to authors who have published on this research topic and whose work was not cited due to space constraints. We thank Traver Hart for providing the list of annotated protein complexes displayed in Fig. 2. Research in the Moffat lab is supported by the Canadian Institutes for Health Research, the Ontario Research Fund, Ontario Institute for Cancer Research, the Terry Fox Research Institute, and the Canadian Foundation for Innovation. JM is a Canada Research Chair in the Functional Genomics of Cancer and a Research Scholar at the Canadian Institute for Advanced Research.

# References

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 1998;391(6669): 806–11.
2. Pan Q, van der Laan LJ, Janssen HL, Peppelenbosch MP. A dynamic perspective of RNAi library development. Trends Biotechnol. 2012;30(4):206–15.
3. McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. Nat Rev Genet. 2002;3(10):737–47.
4. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001;411(6836):494–8.
5. Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS. Post-transcriptional gene silencing by siRNAs and miRNAs. Curr Opin Struct Biol. 2005;15(3):331–41.
6. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature. 2004;428(6981):431–7.
7. Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell. 2006;124(6):1283–98.
8. Paddison PJ, Caudy AA, Sachidanandam R, Hannon GJ. Short hairpin activated gene silencing in mammalian cells. Methods Mol Biol. 2004;265:85–100.
9. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, et al. Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet. 2005;37(11):1281–8.
10. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, et al. Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci USA. 2008;105(51): 20380–5.
11. Marcotte R, Brown KR, Suarez F, Sayad A, Karamboulas K, Krzyzanowski PM, et al. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. Cancer Discov. 2012;2(2):172–89.

12. Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, et al. Cancer proliferation gene discovery through functional genomics. Science. 2008;319(5863):620–4.
13. Silva JM, Marran K, Parker JS, Silva J, Golding M, Schlabach MR, et al. Profiling essential genes in human mammary cells by multiplex RNAi screening. Science. 2008; 319(5863):617–20.
14. Echeverri CJ, Beachy PA, Baum B, Boutros M, Buchholz F, Chanda SK, et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Methods. 2006; 3(10):777–9.
15. Kaelin Jr WG. Molecular biology. Use and abuse of RNAi to study mammalian gene function. Science. 2012;337(6093):421–2.
16. Moffat J, Reiling JH, Sabatini DM. Off-target effects associated with long dsRNAs in Drosophila RNAi screens. Trends Pharmacol Sci. 2007;28(4):149–51.
17. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
18. Fullgrabe J, Kavanagh E, Joseph B. Histone onco-modifications. Oncogene. 2011;30(31): 3391–403.
19. Ellis L, Atadja PW, Johnstone RW. Epigenetics in cancer: targeting chromatin modifications. Mol Cancer Ther. 2009;8(6):1409–20.
20. Conrad C, Gerlich DW. Automated microscopy for high-content RNAi screening. J Cell Biol. 2010;188(4):453–61.
21. Bjorkman M, Ostling P, Harma V, Virtanen J, Mpindi JP, Rantala J, et al. Systematic knockdown of epigenetic enzymes identifies a novel histone demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion. Oncogene. 2012;31(29):3444–56.
22. Li M, Jiang L, Kelleher NL. Global histone profiling by LC-FTMS after inhibition and knockdown of deacetylases in human cells. J Chromatogr B Analyt Technol Biomed Life Sci. 2009;877(30):3885–92.
23. Manova V, Singh SK, Iliakis G. Processing of DNA double strand breaks by alternative nonhomologous end-joining in hyperacetylated chromatin. Genome integr. 2012;3(1):4.
24. Patzlaff JS, Terrenoire E, Turner BM, Earnshaw WC, Paulson JR. Acetylation of core histones in response to HDAC inhibitors is diminished in mitotic HeLa cells. Exp Cell Res. 2010;316(13):2123–35.
25. Bric A, Miething C, Bialucha CU, Scuoppo C, Zender L, Krasnitz A, et al. Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. Cancer Cell. 2009;16(4):324–35.
26. Zender L, Xue W, Zuber J, Semighini CP, Krasnitz A, Ma B, et al. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. Cell. 2008;135(5):852–64.
27. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell. 2010;17(1):13–27.
28. Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature. 2011;478(7370):524–8.
29. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, et al. Selective inhibition of BET bromodomains. Nature. 2010;468(7327):1067–73.
30. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327(5964):425–31.
31. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38(8):896–903.
32. Lin YY, Kiihl S, Suhail Y, Liu SY, Chou YH, Kuang Z, et al. Functional dissection of lysine deacetylases reveals that HDAC1 and p300 regulate AMPK. Nature. 2012;482(7384): 251–5.
33. Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA, et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci USA. 2011;108(30):12372–7.

34. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell. 2009;137(5):835–48.
35. Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. HIstome–a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 2012;40(Database issue):D337–42.
36. Kutateladze TG. SnapShot: histone readers. Cell. 2011;146(5):842-e1.
37. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
38. Close P, Creppe C, Gillard M, Ladang A, Chapelle JP, Nguyen L, et al. The emerging role of lysine acetylation of non-nuclear proteins. Cell Mol Life Sci. 2010;67(8):1255–64.
39. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. Gene. 2005;363:15–23.
40. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489(7417):519–25.
41. Di Lorenzo A, Bedford MT. Histone arginine methylation. FEBS Lett. 2011;585(13):2024–31.
42. Lee BB, Lee EJ, Jung EH, Chun HK, Chang DK, Song SY, et al. Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. Clin Cancer Res. 2009;15(19):6185–91.
43. Lakowski TM, Frankel A. Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4. Biochem J. 2009;421(2):253–61.
44. Miao F, Li S, Chavez V, Lanting L, Natarajan R. Coactivator-associated arginine methyltransferase-1 enhances nuclear factor-kappaB-mediated gene transcription through methylation of histone H3 at arginine 17. Mol Endocrinol. 2006;20(7):1562–73.
45. Hong H, Kao C, Jeng MH, Eble JN, Koch MO, Gardner TA, et al. Aberrant expression of CARM1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status. Cancer. 2004;101(1):83–9.
46. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes–2009. Nucleic Acids Res. 2010;38(Database issue):D497–501.
47. Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, et al. NAViGaTOR: network analysis, visualization and graphing Toronto. Bioinformatics. 2009;25(24):3327–9.

# Index


**A**
Acute myeloid leukemia (AML), 245, 289, 290, 301
Akiyoshi, B., 123
Alpha-thalassemia/mental retardation syndrome, 18
Anti silencing function 1 (ASF1), 288
Arginine methylation, 71, 135, 150, 165, 240, 242, 303, 304
Arrowsmith, C.H., 217

**B**
Baker, L.A., 219
Bantscheff, M., 190
Barsyte-Lovejoy, D., 217
Bedford, M.T., 135
Bhanu, N.V., 149
Bjorkman, M., 299
Boeke, J., 302
Boulon, S., 228
Bremner, R., 277
Bromo adjacent homology (BAH), 215
Bromodomain
  chromatin-modifying enzymes interactomes, 55
  histone tail recognition, 212, 214, 216–217, 219–221
  inhibitor, 219–221
  posttranslational modifications (PTMs), 136
  protein−protein interaction (PPI), 55
  reader domain, 212, 214
  target sequences, 216–217
Brown, K., 297
Butala, M., 126
Byrum, S.D., 127

**C**
Cagney, G., 197
Cancer
  anticancer drug discovery (*see* Chromatin modifying enzyme)
  BRCA1/2, 18
  chromatin modifying enzyme (*see* Chromatin modifying enzyme)
  chromosome translocation, 261
  epigenetics, 199–201
  histone chaperone
    ASF1 and HIRA, 288
    CAF1 (*see* Chromatin assembly factor 1)
    DEK, DAXX and H3.3, 289–290
    DNA replication, 277–279
  histone methylation (*see* Histone methylation)
  HMEs (*see* Histone-modifying enzymes)
  KDM4, 12
  KDM3B, colorectal, 11
  mutation identification, 260
  nucleotide change effect, 269–271
  RNAi technology
    chromatin regulator identification, 302–304
    gene profile integration, 304–305
    lysine deacetylase, 301–302
  stem cell property, 198–199
Carroza, M.J., 187
Centromer protein A (CENP-A), 125



A. Emili et al. (eds.), *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, DOI 10.1007/978-1-4614-7931-4,


© Springer Science+Business Media New York 2014

Chaerkady, R., 205
Chaperones. *See* Histone chaperones
Chiarle, R., 267
Chromatin
  histone methylation (*see* Histone methylation)
  in human disease
    alpha-thalassemia/mental retardation syndrome, 18
    BRCA1/2, 18
    ICF syndrome, 18
    Rett syndrome, 18
    Rubinstein–Taybi Syndrome (RSTS), 18
  structure and function
    affinity purifications/mass spectrometry (AP–MS), 3
    DNA methylation, 2–3
    posttranslational histone modifications, 2
Chromatin affinity purification with mass spectrometry (ChAP-MS), 127
Chromatin assembly factor 1 (CAF1)
  cancer, 287
  epigenetic inheritance, 285, 286
  heterochromatin, 284–286
  heterozygous mutation, 287
  RNAi-induced defect, 285
Chromatin-associated proteins
  immunoaffinity purification of
    CENP-A, 125
    disadvantage of, 126
    DNA probes, 124
    mChIP, 125
    micrococcal nuclease digestion, 124
    of protein complex, 126–127
  nucleic acid probes
    DNA-bound complexes in vitro, 121–122
    DNA probes, 119, 120
    DNA–protein complexes generated in vivo, 122–12
  protein–protein interactions, 128
  type of, 118, 119
Chromatin–chromatin interaction, 264–266
Chromatin modifying enzyme (CME)
  arginine methyltransferase, 241, 246–247
  cancer
    BiNGO tool, 56, 57
    Dancer database, 59
    oncogenes enrichment, 58
    "p53 binding" activity, 57
    RNA polymerase II promoter, 57
  histone methylation, 240, 241
  interactomes
    bromodomain, 55
    DNMT1, 54
    epigenetic therapy, 59–60
    H1 and H3.3, 55
    MBD domain, 55
    PHD, 55
    PWWP, 55–56
    SIN3, 55
    SUMO2, 55
    UBC (poly-ubiquitin chain), 55
  nucleosome, 240
  PKMTs, 241
  PMTs (*see* Protein methyltransferases (PMTs))
  PPI networks construction
    CM-PPI network, 53, 54
    CORUM database, 52
    data sources, 51–52
    iRefWeb, 52
  PRMTs, 242
  protein-coding genes, 51
  protein methylation pathway, 242–243
Chromatin readers
  MS-based interactomics technology, 143–144
  posttranslational modifications (PTMs), 134–135
  protein–protein interactions
    complexes assigning, 141–142
    label-free stoichiometry, 142–143
    quantitative immunoprecipitation combined with knockdown (QUICK), 140
    quantitative mass spectrometry (AP-qMS), 140
    tandem affinity purification (TAP), 140
  quantitative mass spectrometry-based proteomics
    DNA and nucleosome readers, 139
    histone PTM readers identification, 136–139
    MS-based interactomics, 135–136
Chromatin regulator
  cancer cell proliferation, 302–304
  gene profile integration, 304–305
  lysine deacetylases, 301–302
Chromatin-remodeling complex
  cancer, 185
  cellular function, 184–185
  chromosomal translocation, 185
  DSB repair mechanism, 186
  histone modification, 178–179
  Hox gene cluster regulation, 95

imprinting and dosage compensation, 94–95
molecular mechanism, 183–184
MudPI (*see* Multidimensional protein identification technology (MudPIT))
mutations, 185
ncRNA expression profiling, 95–96
ncRNA knockdown, 96
quantitative proteomics analysis
multiprotein complexes analysis, 187–189
protein interaction network, 186–187
therapeutic application, 189–190
in stem cells, 96–97
types, 178
Chromosome translocation
affect, genome organization, 269
cancer, 261
genome-wide identification
intrachromosomal rearrangement, 267–268
ionizing radiation, 268
target sequence, 266, 267
translocation frequency, 268–269
Cloutier, P., 227
Collins, S.R., 128
Collision-induced dissociation (CID), 153
Copeland, R.A., 239
Coulombe, B., 227
Cryptic unstable transcripts (CUTS), 89

**D**
Death associated protein six (DAXX), 289–290
Déjardin, J., 122, 162
DEK, 289–290
Dephoure, N., 159
DNA double-strand breaks (DSBs)
chromosomal translocation frequency, 268–269
intrachromosomal rearrangement, 267–268
ionizing radiation, 268
repair mechanism, 186
target sequence, 266, 267
Domecq, C., 227
DOT1L
arginine methyltransferase, 241–242
HOX gene upregulation, 41
PMTs, 244–245
Doucet, A., 117
Drane, P., 290
DSBs. *See* DNA double-strand breaks (DSBs)

**E**
Ekwall, K., 166
Electron capture dissociation (ECD), 153, 156
Electron transfer dissociation (ETD), 153
Emili, A., 33
Epigenetics
cancer, 199–201
chromatin readers, 143–144
histone chaperones, 277–279
histone methylation, 35

**F**
Felletar, I, 217
Figeys, D., 117
Filippakopoulos, P., 211, 217
Fluorescence in situ hybridization (FISH), 105, 262–264
Fodor, B.D., 164
Forget, D., 227
Fujii, H., 127
Fujita, T., 127

**G**
Garcia, B.A., 149
Genome organization, cancer cells
cell nucleus, 262–263
chromatin–chromatin interaction, 264–266
chromosome translocation (*see* Chromosome translocation)
nuclear, 258–259
nucleotide change effect, 269–271
p53 mutation, 258
*TP53* gene, 258
Gingras, A.C., 217
Goldberg, A.D., 290
Greenblatt, J.F., 1, 33, 87
Guruharsha, K.G., 52
Guttman, M., 100
Gygi, S.P., 159

**H**
HDACs. *See* Histone deacetylases (HDACs)
Hi-C technique, 264–266
Higgins, M.E., 57
Histone chaperones
biological regulation, 280–282
cancer
ASF1 and HIRA, 288
CAF1 (*see* Chromatin assembly factor 1)

Histone chaperones (*cont.*)
DEK, DAXX and H3.3, 289–290
DNA replication, 277–279
epigenetics, 277–279
nucleosome replication, 279–280
polycomb function, 291
recycle, 282–284
Histone deacetylases (HDACs), 178, 185, 187–189
Histone demethylases
chromatin structure and function
affinity purifications/mass spectrometry (AP–MS), 3
arginine residues, 3–4
ARID domain, 5
DNA methylation, 2–3
JUMONJI family (*see* Jumonji C-terminal domain)
KDM1 family, 7, 10
lysine 4 (H3K4), 4, 5
nonhistone substrates demethylation, 16–17
posttranslational histone modifications, 2
in human disease, 20–21
Histone methylation
atypical function, 42
differential complex assembly, 42
and differentiation, 42–43
epigenetics, 35
and heterochromatin, 38
lysine residues, 34
MBT domain, 36
noncoding RNA, 41–42
pluripotency, 42–43
polycomb methyltransferase complexes, 38–39
position effect variegation (PEV), 35
recognition sites, 35
SET domain, 36
and transcription, 36–37
trithorax methyltransferase complexes, 39–41
Histone-modifying enzymes (HMEs)
arginine residues, 71
deacetylases, 67, 68
familial mutations of, 80–81
function of
histone acetylation, 71–72
histone modification *vs.* chromatin conformation, 72
malicious brain tumor domain (MBT), 73
plant homeodomain (PHD), 73
TUDOR, 73
histone code
DNA repair, 75
DNA replication, 76
gene expression regulation, 74–75
kinases, 71, 72
lysine acetyltransferases, 67
lysine methyltransferases, 68–70
methylation, 68
phosphatases, 71, 73
phosphorylation, 71
somatic mutations
ANOVA test, 78, 79
COSMIC, 77
CREBBP, 76
EZH2, 77
HDAC, 76
histone acetylation genes, 78
histone methylation genes, 78
histone phosphorylation genes, 78, 79
MOZ-TIF2, 76
PRDM, 76–77
Histone regulator A (HIRA), 288
Histone tail recognition
inhibitor
bromodomain, 219–221
methyl-lysine, 221–223
reader domain
acetyl-lysine, 212–214
BAH, 215
bromodomain, 212, 214
disease, 218–219
methyl-lysine, 214–215
plant homeodomain, 214
PWWP and WD40, 215
tudor and chromodomain, 214
target sequences
bromodomain, 216–217
domain architecture, 216
TRIM33 BRD/PHD, 218, 219
Human chromatin protein states
antibody-based methods
ChIP-Chip, 163
ChIP-qMS, 163
chromatin-centric approach, 161
chromatin immunoprecipitation (ChIP), 161–162
genome-centric approach, 161
for phosphorylated proteins, 162–163
PICh, 162
and cancer, 168

epigenetic factors influencing, 151–152
mass spectrometry
chemical derivatization, 156–157
data analysis, 161
isotope labeling, 159–160
label-free quantification format, 156
metabolic labeling for, 157–159
proteomics challenges, 152–153
sample preparation for, 155
types of, 153–155
workflow instrumentation, 160–161
posttranslational modification
heterochromatin protein 1 (HP1), 164–165
high mobility group A1 (HMGA1) protein, 165
histone PTM-based interactome, 166–168
whole nuclear proteome profiling, 165–166
spatial chromatin
histone code, 151
histone posttranslational modifications, 150–151
histone variants, 150
nucleosome structure, 150
human Embryonic Stem Cell (hESC), 201, 204, 205
Hungerford, D.A., 261

**I**
ICF syndrome, 18

**J**
Jadaliha, M., 205
Jumonji C-terminal domain
ARID domain, 5
FBOX domains, 5
in GIS1, 6
JARID2, 15
JMJC-only demethylases, 15–16
KDM2 group, 10–11
KDM3 group, 11
KDM4 group, 12
KDM5 group, 13–14
KDM6 group, 14
KDM7 group, 14–15
PHD, 5
as recruitment factors, 7
Tudor domains, 5

**K**
Keates, T., 217
Ketela, T., 297
Kingston, R.E., 162
Klein, I., 267
Knapp, S., 211, 217

**L**
Lakshminarasimhan, M., 177
Lambert, J.-P., 126, 217
Lennartsson, A., 166
Lewis, P.W., 290
Liu, C., 65
Li, Y., 87

**M**
Malignant brain tumour (MBT), 222
Malovannaya, A., 141
Mangos, M., 217
Mann, M., 159
Marcon, E., 1
Mass spectrometry
human chromatin protein states
chemical derivatization, 156–157
data analysis, 161
isotope labeling, 159–160
label-free quantification format, 156
metabolic labeling for, 157–159
proteomics challenges, 152–153
sample preparation for, 155
types of, 153–155
workflow instrumentation, 160–161
stem cells differentiation
iTRAQ, 203
label-free MS-based proteomics, 203–204
label-free proteomics, 206–207
labels use, 203
model system, 201
peptides measurements, 202–203
proteomic approaches, 204–206
MBT. *See* Malignant brain tumour (MBT)
Methyl-lysine reader, 214–215, 221–223
MiDAC, 190
Migliori, V., 159
Mitchell, J.A., 257
Modified chromatin immunoprecipitation (mChIP), 125
Moffat, J., 297
Montes de Oca, R., 159
Moyer, M.P., 239

Müller, S, 217
Multidimensional protein identification technology (MudPIT)
  advantages of, 179
  multiprotein complex analysis, 187, 189
  protein analysis, 179–180
  quantitative proteomics
    label-free, 182–183
    MS methodology, 180, 181
    stable isotope labeling-based, 180–182

**N**

ncRNA-protein interactions
  biological roles of
    chromatin-remodeling complex (*see* Chromatin-remodeling complex)
    CUTS, 89
    in epigenetic gene regulation, 89, 90
    PASRs, 89
    posttranslational modifications, 97
    PROMTS, 89
    RNAPII, 91
    in transcriptional regulations (*see* Transcriptional regulations)
    in tumorigenesis, 97–98
    UNTS, 89
  ribonucleoprotein IP coupled with sequencing (RIP-seq)
    ChIRP and CHART, 106
    CLIP-seq and PAR-CLIP, 102
    de novo transcript-assembly strategy, 104
    large-scale cell-based RNAi screen, 105
    large-scale protein-tag or antibody-based systems, 99–100
    MS2-tagged RNA, 106
    ncRNA-protein interaction motif predictions, 107–108
    ncRNA stability, 107
    ncRNA structure predictions, 107
    peak calling strategy, 104–105
    RNA deep sequencing, 100–101
    strand-specific, 101
Nowell, P.C., 261
NTERA-2, 201

**O**

Olsen, J.B., 33

**P**

Pardo, M., 205
Pawson, T, 217
Pewsey, E., 206
PFL. *See* Protein frequency library (PFL)
Picaud, S., 217
PKMTs. *See* Protein lysine methyltransferases (PKMTs)
Plant homeodomain (PHD), 3, 5, 10–12, 14, 213, 214
Pluripotency, 198
PMTs. *See* Protein methyltransferases (PMTs)
Polycomb (PcG), 200, 201
Polycomb methyltransferase complexes, 38–39
Position effect variegation (PEV), 35
Posttranslational modifications (PTMs), 178, 179, 181
  bromodomains, 136
  H/L ratio, 136
  multiple qMS-based interaction screens, 136, 137
  quantitative filtering principle, 136
  TFIID binds, 136
  triple pull-down approach, 138–139
PRMTs. *See* Protein arginine methyltransferases (PRMTs)
Promoter-associated short RNAs (PASRs), 89
Promoter upstream transcripts (PROMTS), 89
Protein arginine methyltransferases (PRMTs), 242, 249
Protein frequency library (PFL), 187
Protein lysine methyltransferases (PKMTs), 241, 249
Protein methyltransferases (PMTs)
  DOT1L and MLL, 243–244
  EZH2 and H3K27 methylation, 244–245
  inhibitors of
    aminonucleoside analogs, 248–249
    selective and drug-like, 246–247
  NSD, 245–246
  structure and function of, 249–251
Protein–protein interaction (PPI). *See also* Chromatin modifying enzyme (CME)
  cancer
    BiNGO tool, 56, 57
    Dancer database, 59
    oncogenes enrichment, 58
    “p53 binding” activity, 57
    RNA polymerase II promoter, 57
  chromatin-associated proteins, 128

chromatin-modifying enzymes interactomes
bromodomain, 55
DNMT1, 54
epigenetic therapy, 59–60
H1 and H3.3, 55
MBD domain, 55
PHD, 55
PWWP, 55–56
SIN3, 55
SUMO2, 55
UBC (poly-ubiquitin chain), 55
chromatin readers
complexes assigning, 141–142
label-free stoichiometry, 142–143
quantitative immunoprecipitation combined with knockdown (QUICK), 140
quantitative mass spectrometry (AP-qMS), 140
tandem affinity purification (TAP), 140
networks construction
CM-PPI network, 53, 54
CORUM database, 52
data sources, 51–52
iRefWeb, 52
protein-coding genes, 51
Proteome analysis. *See* Stem cells differentiation
Proteomics of isolated chromatin (PICh), 162
PTM. *See* Posttranslational modifications (PTMs)
Pu, S., 49

**R**
Rafiei, M., 277
Rett syndrome, 18
Richon, V.M., 239
RNA interference (RNAi) technology
array screening, 299–300
cancer
chromatin regulator identification, 302–304
gene profile integration, 304–305
lysine deacetylases, 301–302
chromatin regulation, 298–299
discovery, 297–298
machinery processes, 298
pooled screening, 300–301
siRNA, 298
RNA polymerase II-associated proteins (RPAP)
interaction, 231–232
interaction network, 235–237
POLR1A, 233–234
regulatory factor, 235
silencing, 232–234
RNA polymerases (RNAP)
antibodies, 229, 231
biogenesis of, 229
immunofluorescence and imaging, 231
nuclear import, 228
protein affinity purification, 229, 230
regulation of, 228
RPAP4/GPN1 and RPAP3
interaction, 231–232
interaction network, 235–237
POLR1A, 233–234
regulatory factor, 235
silencing, 232–234
transcriptional functions, 228
transfection and siRNA silencing, 231
Ross, P.L., 203
Royer, L., 3
RPAP. *See* RNA polymerase II-associated proteins (RPAP)
Rubinstein–Taybi Syndrome (RSTS), 18
Rubio, E.D., 121

**S**
S-adenosyl-L-methionine (SAM), histone methylation, 36
Sarkar, P., 205
Sgarra, R., 165
Shiio, Y., 180
Single nucleotide polymorphisms (SNPs), 269–270
siRNA silencing, 231
Smiley, S., 1
Smith, E., 189
Smits, A.H., 133
SNPs. *See* Single nucleotide polymorphisms (SNPs)
Somatic mutations
ANOVA test, 78, 79
COSMIC, 77
CREBBP, 76
EZH2, 77
HDAC, 76
histone acetylation genes, 78
histone methylation genes, 78
histone phosphorylation genes, 78, 79
MOZ-TIF2, 76
PRDM, 76–77

Spectral counting, 182
Stein, L., 65
Stem cells differentiation
  epigenetic regulation
    cancer, 199–201
    properties of, 198–199
  mass spectrometry-based approaches
    2-D separation methods, 202
    iTRAQ, 203
    label-free MS-based
      proteomics, 203–204
    label-free proteomics, 206–207
    labels use, 203
    model system, 201
    peptides measurements, 202–203
    proteomic approaches, 204–206
Sullivan, B.A., 124
Sullivan, K.F., 124

**T**
*TP53* gene, 258
Transcriptional regulations
  ceRNA/miRNA sponge, 94
  cis gene repression, 92
  cis gene upregulation, 92
  mRNA decay, 94
  trafficking, 93–94
  *trans* gene regulation, 93, 94
  translation regulation, 94
Transfection, 231
Trithorax methyltransferase complexes, 39–41
Turinsky, A.L., 1, 49

**U**
Upstream noncoding transcripts (UNTS), 89

**V**
Van den Berg, D.L.C., 205
Vermeulen, M., 133, 182
Volkmer, R., 217

**W**
Washburn, M.P., 177
Watson, A., 197
Wodak, S.J., 49

**X**
Xu R.M., 280

**Z**
Zhang, Z., 87
Zhao, D.Y., 87
Zhou, H.Y., 257
Zuber, J., 301