Chapter 17

Next Generation Sequencing in Alzheimer's Disease

Lars Bertram

Abstract

For the first time in the history of human genetics research, it is now both technically feasible and economically affordable to screen individual genomes for novel disease-causing mutations at base-pair resolution using "next-generation sequencing" (NGS). One popular aim in many of today's NGS studies is genome resequencing (in part or whole) to identify DNA variants potentially accounting for the "missing heritability" problem observed in many genetically complex traits. Thus far, only relatively few projects have applied these powerful new technologies to search for novel Alzheimer's disease (AD) related sequence variants. In this review, I summarize the findings from the first NGS-based resequencing studies in AD and discuss their potential implications and limitations. Notable recent discoveries using NGS include the identification of rare susceptibility modifying alleles in *APP*, *TREM2*, and *PLD3*. Several other large-scale NGS projects are currently underway so that additional discoveries can be expected over the coming years.

Key words Alzheimer's disease, Next-generation sequencing, Rare variant association, Genome-wide association study, GWAS

1 Introduction

Similar to many other adult-onset human disorders, Alzheimer's disease (AD)—a slowly progressive neurodegenerative disease of the brain eventually resulting in cognitive impairment and dementia—represents a "genetically complex trait". This term alludes to the fact that a person's liability for AD is the result of a combination of heritable (e.g. genetic) and non-heritable (e.g. environmental) factors. Twin studies suggest that the contribution of the former probably outweighs the latter for the vast majority of AD cases [1]. In some rare familial forms of AD, genetics plays the predominant role through the effect of extremely infrequent amino acid substituting mutations in genes such as *APP* (β-amyloid precursor protein [APP]), *PSEN1* (presenilin 1 [PS1]), and *PSEN2* (presenilin 2 [PS2]). Despite their rarity, mutations in these genes have been instrumental in clarifying the molecular mechanisms underlying AD pathogenesis where the aberrant production of the β-amyloid (Aβ)

peptide represents a crucial event [2]. Intracellularly, Aβ is cleaved from APP by the sequential action of two enzymatic activities, i.e. β-secretase (encoded by the *BACE1* [β-amyloid cleavage enzyme 1] gene) and γ-secretase (whose catalytic site is made up of PS1 and PS2). By identifying AD-causing mutations in both the precursor of Aβ (i.e. *APP*) and the enzymes involved in its production (i.e. *PSEN1* and *PSEN2*), genetics has supported the "amyloid hypothesis" of AD which posits that dysregulated Aβ triggers the development and perhaps progression of the disease [2]. For recent reviews on AD genetics *see* refs. 3–5.

Mutations in *APP*, *PSEN1*, and *PSEN2* only account for a small fraction (≪5 %) of all AD cases, which I will refer to as "Mendelian AD" due to the almost complete penetrance and mostly autosomal-dominant mode of transmission of implicated DNA sequence changes. The vast majority of AD cases, however, is actually of a "non-Mendelian" nature. The predisposition for this type of AD is the result of a combined action of dozens, if not hundreds or thousands, of common DNA sequence variants (i.e. polymorphisms) of small effect (i.e. odds ratios [ORs] typically ≪2) and, hence, incomplete penetrance. Over 30 years of research have investigated thousands of DNA polymorphisms in hundreds of putative AD candidate genes to find genetic risk factors underlying non-Mendelian AD [6]. With one notable exception, i.e. the apolipoprotein E gene (*APOE*) on chromosome 19 [7], these studies have not resulted in establishing firm disease associations until the advent of high-throughput microarray genotyping technology allowed the completion of genome-wide association studies (GWAS) in ~2008. These GWAS have finally resulted in a number of highly convincing AD association findings (*see* the 'AlzGene' database for a summary of these studies: http://www.alzgene.org [8]). Collectively, however, GWAS associations currently explain no more than half of the disease heritability, i.e. the proportion of phenotypic variance that can be explained by genetic or epigenetic factors. Interestingly, this situation—i.e. that the combined contribution to liability for disease is only partially explained by findings to emerge from even the most powerful GWAS—is observed for many more genetically complex diseases (and non-disease traits). Several potential hideouts for this "missing heritability" have been proposed [9], including the possibility that it may altogether represent a "phantom" phenomenon [10].

The recent development of extremely powerful, massively parallel DNA sequencing technologies now allows to systematically screen individual genomes for DNA sequence variation at base-pair resolution, enabling researchers to address the "missing heritability" question (and many other questions) empirically (Fig. 1; Table 1). Owing to their potential to revolutionize human genetics research, the term "next-generation sequencing" (NGS) has been coined for these methods. In this review, I summarize the findings of the first studies specifically applying NGS to the field of AD
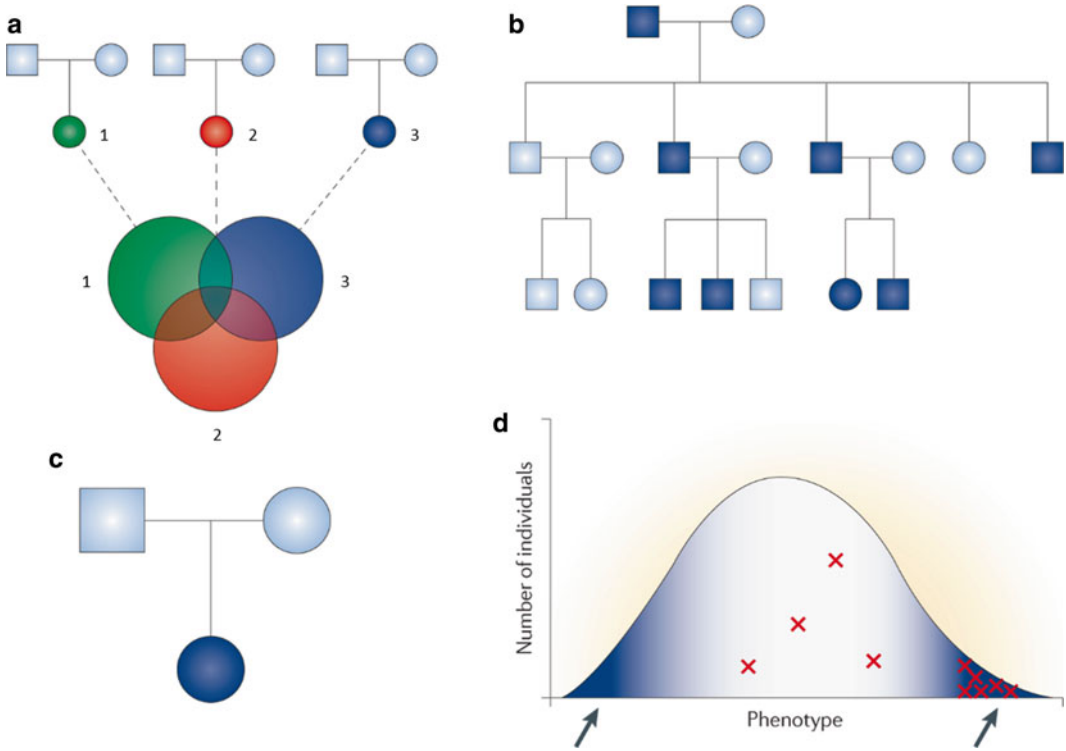
Fig. 1 Strategies for finding disease-causing rare variants using exome sequencing. Four main strategies are illustrated. (**a**) Sequencing and filtering across multiple unrelated, affected individuals (indicated by the *three colored and numbered circles*). This approach is used to identify novel variants in the same gene (or genes), as indicated by the *shaded region* that is shared by the three individuals in this example. (**b**) Sequencing and filtering among multiple affected individuals from within a pedigree (*shaded circles and squares*) to identify a gene (or genes) with a novel variant in a shared region of the genome. (**c**) Sequencing parent–child trios for identifying de novo mutations. (**d**) Sampling and comparing the extremes of the distribution (*arrows*) for a quantitative phenotype. As shown in panel **d**, individuals with rare variants in the same gene (*red crosses*) are concentrated in one extreme of the distribution. Figure and legend reprinted with permission from Macmillan Publishers Ltd: Nature Reviews Genetics [18], copyright (2011)

Table 1
Variant detection in complex traits classified by allele frequency

Variant class	Minor allele frequency	Implication for analysis	Example in AD
Very common	Between 5 and 50 %	GWAS	*APOE, CLU, CR1, PICALM*
Less common	Between 1 and 5 %	GWAS	n.a. (not available)
Rare (but not private)	Less than 1 % but still polymorphic in one or more major human population	GWAS; co-segregation in families	*APP* (Ala673Thr), *PLD3, TREM2*
Private	Restricted to probands and immediate relatives	Co-segregation in families; linkage	AD-causing mutations in *APP, PSEN1, PSEN2*

GWAS genome-wide association study. See text for explanation of gene symbols. Adapted for this review with permission from Macmillan Publishers Ltd: Nature Reviews Genetics [19], copyright (2010)

genetics research. The focus lies exclusively on studies using NGS for genome resequencing (whole or in part) and does not cover projects utilizing other NGS applications, such as transcriptome or methylome sequencing.

2   Next-Generation Sequencing to Identify Novel Disease Genes

There already exist a large number of excellent reviews on the technical details of the currently available NGS platforms [11, 12], as well as on relevant theoretical (e.g. [13–16]), practical (e.g. [17–20]), and analytical (e.g. [21–23]) considerations when embarking on an NGS-based resequencing project (*see* also Fig. 1). Accordingly, these topics will be skipped here except for a few general remarks. For instance, one important issue to keep in mind is that the *generation* of NGS-based large-scale sequencing data has become relatively straightforward thanks to the availability of highly optimized operating procedures developed by the manufacturers of today's NGS instruments. On the other hand, efficient and appropriate *management* and *interpretation* of the resulting sequence data is not nearly as straightforward for most laboratories outside highly specialized genome centers. To a large part this is due to the sheer amount of information created. For instance, a single human genome consists of ~3.2 billion base-pairs (Gbp), each of which needs to be covered at least 30- to 35-fold in order to confidently differentiate between wild-type allele and mutation [20], yielding a minimum of ~100 Gbp per DNA sample per experiment.

Another, possibly even more challenging aspect is that potentially "functional" DNA sequence variants occur at much higher frequency in the general population than originally anticipated (even though overall they may still be classified as "rare", i.e. displaying a minor allele frequency [MAF] $\ll 1$ %) [9, 24]. The important conclusion to draw is that not every amino-acid changing nucleotide substitution found in an affected individual or observed to co-segregate with disease status in a given family also automatically represents the underlying disease-causing variant. This situation has been referred to as the "narrative potential" of individual genomes [15], meaning that assigning a disease-related narrative to potential mutations in anyone's genome sequence is relatively easy (simply owing to the high frequency of these sequence changes) but statistically often poorly justified. Thus, in order to avoid publishing "genomic fairytales", researchers need to go to great lengths to ensure that a presumed connection between a pinpointed DNA sequence variant and onset of disease is in fact genuine. For GWAS findings this typically entails to provide consistent association evidence in several independent data sets which, when analyzed alone or in combination, pass a certain threshold of statistical support (typically a P-value below 5×10^{-8}). Neither of these requirements can be applied to NGS-based genetics studies in a straightforward fashion.

First, the variants identified are often exceedingly rare (if not altogether "private", i.e. restricted to one founder and its family members), so that a sufficient number of carriers—affected by disease or not—may be difficult to come by for any individual laboratory (*see* also Table 1). Second, there currently exist no firm guidelines on the statistical interpretation of "rare variant" association findings. As a matter of fact, there currently exist no firm rules on how to establish evidence in favor of a genetic association between disease status and specific rare variants. Possible analysis strategies include variant-specific tests (similar to those used for common variants in GWAS), gene-specific tests (e.g. by pooling discovered variants within loci in affected vs. unaffected individuals) with and without pre-defined allele frequency thresholds, and network/gene-set analyses (e.g. by pooling association evidence across genes by their presumed or proven functional connection).

As will become clear in the following paragraphs, many of the issues briefly touched upon above have already been encountered in the few available NGS-studies conducted in the field of AD genetics. Thus, some of the reported gene findings outlined below can be assigned greater credibility than others. In a sense, this situation is not unlike that encountered during the pre-GWAS candidate gene era of AD genetics research. The field will likely remain in this state until firm criteria on the analysis and interpretation of NGS resequencing data have been established. I close this introduction by quoting from a recently published, highly interesting essay on the theoretical framework of rare-variant associations in human diseases [13], which concluded that "for very late-onset diseases like Alzheimer's […] common variant association studies would likely be the better strategy" to identify the most important disease genes. This is due to the fact that rare variant associations will be a comparatively infrequent occurrence in these diseases for a number of reasons discussed in [13]. Time will tell, whether or not these considerations and conclusions will prove to be correct.

3   Next-Generation Sequencing in Alzheimer's Disease Research to Date

Owing to the fact that NGS technologies have only become available (and affordable) outside highly specialized genome centers from ~2010, the literature reporting first results of their application to AD is still rather limited. For the purpose of this review, NCBI's "PubMed" database (http://www.ncbi.nlm.nih.gov/pubmed) was queried using the keywords "[alzheimer* AND ((next generation sequencing OR NGS OR deep sequencing) OR (exome sequencing OR WES) OR (whole genome sequencing OR WGS))]" which yielded a total 103 publications on April 15, 2014. Of these, only 15 reported data relevant to this review, i.e. bona fide NGS-based resequencing in at least one cohort of AD patients. Together with a few additional relevant publications identified via

other sources, these studies represent the "core" findings discussed in more detail below (Table 2). Notwithstanding the relative paucity of NGS studies in AD, the already available papers have collectively applied the full range of different NGS strategies and thus provide a timely starting point for a first critical assessment. Note, that this field is advancing very rapidly, so that readers are advised to consult the 'AlzGene; database' (http://www.alzgene.org [8]) or 'AD and FTD Mutation Database' (http://www.molgen.ua.ac.be/ADMutations/ [25]) for updated summaries of relevant studies published after the day of writing.

3.1 NGS for Resequencing of Alzheimer's Disease Candidate Genes

In an attempt to resolve the contribution of putative functional DNA sequence variants in genes *known* to be associated with disease risk, most early NGS studies in AD either performed deep resequencing of the established early-onset Mendelian AD (and other forms of dementia, particularly FTLD) genes, i.e. *APP*, *PSEN1*, *PSEN2*, *MAPT* and *GRN*, or of loci recently implicated by GWAS, in particular *CLU*, *CR1*, and *PICALM* (Table 2). In addition, there are also a few publications reporting deep resequencing results of loci that had emerged during the "candidate gene" era of AD genetics, i.e. *ABCA1* (encoding ATP-binding cassette, sub-family A [ABC1], member 1) and *NCSTN* (encoding the γ-secretase component nicastrin; Table 2). As will be discussed below, the knowledge gained from these focused, early-adopter NGS studies remains limited. This is in contrast to more systematic projects performing whole exome (WES; Subheading 3.2) or whole genome sequencing (WGS; Subheading 3.3).

3.1.1 NGS for Resequencing Mendelian AD Genes

Several of the first projects have applied NGS to resequence known Mendelian genes in late-onset, non-Mendelian AD cases [26–28]. All of these studies identified novel "rare variants" across the investigated loci leading some authors even to conclude that "rare variants in these genes could explain an important proportion of genetic heritability of AD" [26]. It remains debatable whether this conclusion is indeed justified. First, many of the variants identified within target genes were deemed as "non-pathogenic", i.e. there is currently no compelling evidence from in silico or in vitro assessments to support an impact on pathogenicity (see also 'AD & FTD Mutation Database' for more details). Second and more importantly, sequencing technologies and bioinformatic variant-calling algorithms (and correlated measures such as base-pair coverage, which determine the false-positive and false-negative rate) in AD cases often differed from those applied to controls, potentially biasing the discovery of "rare variants" towards AD populations. This situation arises when AD cases are resequenced in-house and then compared to control genomes derived from public databases, such as the 1000 Genomes Project website (http://www.1000genomes.org) or the ExomeVariant Server (http://evs.gs.washington.edu/EVS/). Possibly the most interesting result

Table 2
Findings from NGS-based resequencing studies in AD

Ref.	Author, year	NGS approach	Study design	Main finding	Overall credibility
[26]	Cruchaga, 2012	Candidate (APP, PSEN1/2, MAPT, GRN)	Family-based (LOAD)	RVs in APP, PSEN1, GRN, MAPT to confer risk	Unclear
[27]	Jin, 2012	Candidate (APP, PSEN1/2, MAPT, GRN)	Family-based (EOAD, LOAD)	RVs in PSEN1, GRN, MAPT to confer risk	Unclear
[28]	Benitez, 2013	Candidate (APP, PSEN1/2, APOE, MAPT, GRN)	Extreme CSF levels	PSEN1-E318G linked to CSF-tau	Low
[29]	Lord, 2012	Candidate (CLU, CR1, PICALM)	AD cases only	Technical issues	–
[36]	Lupton, 2011	Candidate (NCSTN)	Case–control (LOAD)	RVs to confer risk	Low
[37]	Lupton, 2014	Candidate (ABCA1)	Case–control (LOAD)	RVs to confer protection	Low
[39]	Guerreiro, 2012	WES	Family (1 exome)	NOTCH3-R1231C in index patient	Low
[40]	Pottier, 2012	WES	Families (EOAD)	RVs to confer risk	Low/ medium
[43]	Cruchaga, 2014	WES	Families and case–control data sets	PLD3 (several RVs to confer risk)	Medium
[50]	Jonsson, 2012	WGS (for GWAS)	Case–control	APP-Ala673Thr to confer protection	High
[51]	Jonsson, 2013	WGS (for GWAS)	Case–control	TREM2-Arg47His to confer risk	High
[58]	Guerreiro, 2013	WGS, WES	Case–control	TREM2-Arg47His (+ other RVs) to confer risk	High

Simplified summary of lead findings of NGS studies discussed in main text. "Overall credibility/relevance" as (subjectively) judged by the author of this review based on the data presented in original publications. See main text for more details on studies and findings. *RVs* rare variants, *LOAD* late-onset AD [also referred to as "non-Mendelian AD" in main text], *EOAD* early-onset AD [also referred to as "Mendelian AD" in main text], *GWAS* genome-wide association study. See text for explanation of gene symbols

from these early NGS studies is the comparatively high proportion of pathogenic *MAPT* and *GRN* variants in individuals clinically diagnosed as "AD" suggesting a larger than anticipated misdiagnosis rate [26] or—less likely—pleiotropy at these loci.

3.1.2 NGS for Resequencing AD Susceptibility Genes

At the day of writing, there is only one published *bona fide* NGS study investigating the established AD GWAS loci [29]. This project performed pooled NGS of sequence capture products targeting *CLU*, *CR1*, and *PICALM* in 96 AD patients and compared the

frequency of identified variants to those listed in public databases. Owing mostly to technical difficulties encountered during the course of the project, no firm conclusion regarding the prevalence of putative functional variants (common or rare) in these loci could be reached. This is in line with a number of conventional, Sanger-based resequencing projects of either all or a subset of the same genes [30–35]. These studies were also unable to conclusively pinpoint "functional" variants within previously reported GWAS loci. The same is true for NGS-based resequencing studies of non-GWAS AD candidate genes, i.e. *NCSTN* [36] and *ABCA1* [37], which have produced little more than anecdotal results (Table 2).

Collectively, this situation is reminiscent of the pre-GWAS "candidate gene era" of AD genetics which has produced a flurry of proclaimed results essentially none of which survived the test of time [38]. The current lack of convincing NGS results in the established AD loci may be due to the same combination of factors that had already bedeviled genetic association studies back then, including small sample size, smaller-than-anticipated effect sizes, failure to provide replication evidence from independent cohorts, and over-interpretation of borderline statistical evidence.

3.2   Whole-Exome Sequencing Studies in Alzheimer's Disease

With respect to "outcome", NGS studies following a more systematic approach—e.g. those focusing on the exome (WES) or whole genome (WGS)—have thus far produced more convincing results than those investigating only a few loci (*see* Subheading 3.1). While a number of publications report to have performed WES on AD cases and sometimes control populations, only the three studies discussed below applied this technology for *bona fide* exome-wide discovery of novel disease genes.

The first study by Guerreiro et al. [39], used WES in a single clinically diagnosed AD patient from Turkey originating from a consanguineous family with a complex medical history, including the presence of both neurological and immunological disorders. Bioinformatic analysis and filtering of the WES data identified 178 candidate missense variants, one of which (p.R1231C in *NOTCH3*; Table 2) was reported as the molecular culprit for the clinical AD-like picture seen in this patient. This mutation, along with more than 130 other mutations in this gene, has previously been reported to cause another neurological disorder: CADASIL (cerebral arteriopathy autosomal dominant with subcortical infarcts and leukoencephalopathy) [39], an inherited cerebrovascular disease with a number of different neurological symptoms. Typically, patients with CADASIL show pronounced white matter abnormalities resulting in characteristic MRI findings. These MRI findings, however, were absent from the patient subject in this study. Unfortunately, DNA from other affected relatives was not available to assess segregation of the variant with disease status. The only other carrier of the p.R1231C mutation was a cognitively normal

son of the index patient, approximately 20 years younger than the dementia onset age in his father. The putative AD-causing *NOTCH3* mutation was absent in more than 300 AD cases and controls from Turkey and elsewhere. While this study demonstrated the power of WES to efficiently generate a near complete status of the "mutational landscape" in a single patient, it did lack some essential supporting evidence to conclusively imply the p.R1231C *NOTCH3* variant as the cause of the complex clinical phenotype observed in this particular patient: e.g. lack of within family disease segregation, independent replication, exclusion of causality of other missense variants present in WES data. Further, owing to the unclear and complex clinical picture it is too early to add *NOTCH3* to the list of AD genes. An alternative interpretation of the available data could simply be—as the authors themselves concede—that p.R1231C is neither pathogenic for CADASIL nor AD.

The second publication [40] highlighted in this section used WES in 14 index cases of autosomal dominant early-onset AD families without mutations in any of the established AD genes *APP*, *PSEN1*, and *PSEN2*. Bioinformatic filtering of the variant calls was based on putative functional impact (i.e. only non-synonymous, splice site, and frameshift indels were retained), "novelty" (i.e. only variants not listed in public databases were retained), and recurrence (i.e. only genes harboring variants resulting from the previous filtering steps in >1 family were retained). This strategy led to a number of potential candidate genes, the most compelling of which was sortilin-related receptor LR11/SorLA (*SORL1*; Table 2) [40]. In its function as a sorting receptor and central regulator of the trafficking and processing of APP, SorLA has represented an AD candidate gene for nearly a decade [41]. Early genetic association analyses suggested *SORL1* to be a late onset Alzheimer's disease (LOAD) risk gene, but were met with a flurry of conflicting data [8]. A recent GWAS meta-analysis revived the topic by reporting common variants in *SORL1* to show genome-wide significant association with AD risk [42], although previous analyses of largely overlapping datasets did not yield such a finding. In the WES study discussed here, *SORL1* was found to harbor previously unknown and putatively functional mutations in 5 of the 14 index patients. Resequencing of *SORL1* in a separate collection of 15 index patients from independent EOFAD families identified two additional novel variants also presumed to be disease-causing [40]. Owing to the lack of available biospecimen, co-segregation with disease could only be demonstrated for one of the seven *SORL1* variants. If indeed genuine, these findings would imply that *SORL1* would be on a par with *PSEN2* as the third most frequently mutated gene responsible for autosomal dominant forms of AD. Furthermore, it would represent the first gene in AD genetics to harbor both rare and disease-causing as well as common susceptibility variants. The lack of functional data clearly and directly supporting an impact of

the identified mutation on protein expression or function, the absence of conclusive segregation evidence in all but one family, and the current nonexistence of independent replication results suggests that more time and scientific evidence is needed before *SORL1* can be counted as an established causal AD gene.

The third and most recent paper to employ WES in AD [43] sequenced a total of 29 individuals from 14 AD families with four or more affected individuals. Filtering based on minor allele frequency (<0.5 %), segregation of candidate variants with disease status within families, and occurrence of the same variant in >1 family, revealed a missense change in the gene encoding phospholipase D3 (*PLD3*; Table 2) on chromosome 19q13.2 [43]. The protein represents a hitherto poorly characterized member of the PLD superfamily of phospholipases. Other members of this superfamily, i.e. PLD1 and PLD2, have been reported to be involved in APP trafficking and synaptic dysfunction [48, 49], making PLD3 a reasonable candidate as well. The PLD3 variant identified by Cruchaga and colleagues, rs145999145, elicits a valine to methionine change at residue 232 (Val232Met) and is present in up to 0.5 % of non-AD individuals of European descent. In AD cases, the frequency of the Met-allele is between 0.6 and 1.3 %, thus approximately doubling the risk for AD in carriers vs. non-carriers [43]. The original family-based finding was subsequently extended to an independent series of more than 11,000 AD cases and controls where it was also found to be associated with risk for AD (OR ~2) and a significant reduction in onset age (between 3 and 8 years). Resequencing of the *PLD3* coding region in ~4,300 AD cases and controls of European descent revealed potential additional rare PLD3 variants which, by means of aggregate analysis ("burden test"), occurred significantly more often in cases than healthy controls. This finding may indicate the presence of other disease associated functional variants at this locus beyond Val232Met. Finally, additional analyses revealed an excess of rare *PLD3* coding variants in AD cases vs. controls in a small collection of individuals of African descent, further supporting the notion that *PLD3* represents a genuine AD locus. The genetic results were accompanied by a range of supporting functional data from human brain samples and transgenic mouse neuroblastoma cell lines suggesting that *PLD3* may exert its pathogenic effects by directly affecting APP processing and Aβ42 and Aβ40 production. Based on these data, the most likely effect of Val232Met (and possibly other variants in *PLD3*) is a loss of function, e.g. by reduced gene expression, and as a result an increase in Aβ42 and Aβ40 production. Since the original publication [43], a number of studies [44–47] were published in an attempt to independently validate the reported findings. However, despite having excellent power to detect the previously reported effect size, none of these papers was able to replicate the

association between AD risk and Val232Ala (or other polymorphisms in PLD3) shedding serious doubt on the notion that PLD3 is in fact a genuine AD gene [44–47].

3.3 Whole-Genome Sequencing Studies in Alzheimer's Disease

At the time of writing, there are no published studies directly applying WGS for the discovery of novel AD genes, but several such studies are currently underway. For instance, WGS data generated as part of the "AD sequencing project" (ADSP) on 584 individuals from 111 multiplex AD families were recently released to the community and are in the process of being analyzed for the presence of AD-associated rare variants in these families. First results from these efforts are expected to be reported early in 2015 (*see* ADSP website for more details: https://www.niagads.org/adsp/). A more extensive WGS study is being conducted as part of the "Alzheimer's Genome Project" (AGP, P.I. Rudolph E. Tanzi at Harvard Medical School), which has recently completed the generation of sequencing data in over 1,500 individuals from 437 multiplex AD families. Bioinformatic workup of the sequence reads is currently underway, and first results are expected to be released in 2015 (for more details *see*: http://www.curealz.org/projects/whole-genome-sequencing).

In addition to *directly* sequencing specific subjects and families of interest for disease gene discovery, WGS can also be used *indirectly*. This strategy was followed by researchers from deCODE genetics who utilized WGS data on a collection of ~2,000 Icelandic individuals in two related projects investigating the role of rare functional variants on AD risk [50, 51]. In both studies, WGS data were used to impute allele status at sites harboring putative functional variants (n ~192,000, including nonsynonymous, frameshift, splice site, and stop gain-loss variants) onto microarray based genome-wide genotype data in more than 3,000 AD patients. These were compared to genotype data imputed from the same WGS panels in ~80,000–110,000 Icelandic healthy control individuals. In essence, the analytic strategy applied in these projects comes down to a GWAS on imputed genotype data specifically enriched for putative functional variants originating from WGS data in the population of interest. Apart from *APOE*, the analyses of the resulting data pinpointed two previously known rare missense variants showing evidence for genome-wide significant association with AD risk: (1) the Ala673Thr substitution in *APP* (decreasing the risk for developing AD in carriers by approximately fivefold) [50], and (2) the Arg47His substitution in *TREM2* on chromosome 6p21.1 (increasing the risk for developing AD by approximately threefold) [51]. Unlike the WES-based results reported in the previous section, both findings have been confirmed by independent laboratories following the original publications.

The first AD association to be revealed by deCODE's "WGS-enriched GWAS strategy" highlighted a previously known SNP

(rs63750847) in *APP* eliciting a threonine to alanine substitution at residue 673 (Ala673Thr) [50]. Interestingly, the minor A-allele was found to be *under* represented in the Icelandic AD cases under study, suggesting a protective effect when compared to the reference G-allele (translating into an OR reduction of about fivefold). In addition to these AD specific findings, the authors also investigated the decline in cognitive function over time in non-AD individuals between 80 and 100 years of age [50]. In line with the observed protection from AD, carriers of the A-allele at rs63750847 performed significantly better in cognitive testing than non-carriers. First in vitro functional data reported in the same study revealed that the protective A-allele (coding for amino acid threonine), which is located within APP's β-cleavage site, significantly reduces the production of sAPPβ and $A\beta_{40/42}$ by approximately 50 % relative to the wild-type G-allele (coding for alanine) [50]. Overall, these data suggest that Ala673Thr exerts a direct effect on BACE1 cleavage of APP. Independent follow-up studies in various populations either confirmed the protective effect of Ala673Thr (e.g. in a Finnish population [52]), or were unable to identify any carriers of the minor threonine allele (e.g. in South-East Asian [53, 54] or North American [55] samples). Of note, A673 was also found in one familial AD case in the NGS-based assessment of known dementia genes by Cruchaga et al. (*see* above and [26]), and in a patient suffering from ischemic cerebrovascular disease [56], a finding that is reminiscent of the protective *APOE* ε2-allele which has also been found—albeit at reduced frequency—in patients suffering from AD and other neurodegenerative diseases [57].

The second AD association identified by the deCODE group was with SNP rs75932628 leading to a an arginine to histidine substitution at position 47 (Arg47His, or R47H) in the gene encoding triggering receptor expressed on myeloid cells 2 (*TREM2*) [51]. Combining association results from a number of different datasets, the minor T-allele at this site was found to significantly increase the risk for AD by approximately threefold. Opposite to the effects observed for the protective variant in *APP* (see above), the AD-associated allele was associated with poorer cognitive function in cognitively healthy individuals aged between 80 and 100 years. Independent confirmation of the association between *TREM2* and AD was reported by another group alongside the deCODE findings [58]. This latter study utilized a combination of DNA sequencing approaches (including reanalysis of WES and WGS data) and reported evidence for the presence of other rare *TREM2* variants in addition to Arg47His to show association with AD. Since these original reports, the Arg47His association with AD has been replicated by several other groups [59–61]. More recently, Arg47His has also been—albeit less consistently—associated with other forms of neurodegenerative disorders (such as PD [62, 63], FTLD [63] and ALS [64]), although these latter findings could not be replicated in independent

datasets [65]. Functionally, TREM2 is likely involved in the body's innate immune system response based on evidence suggesting that the encoded receptor protein has been shown to regulate the phagocytic ability and inflammatory response of microglia [66]. The primary form of resident macrophages in the central nervous system (CNS).

4   Conclusions and Outlook

For the first time in the history of human genetics research, it is now both technically feasible and economically affordable to systematically screen individual genomes for novel disease-causing mutations at base-pair resolution using "next-generation sequencing". Thus far, only relatively few studies have applied these powerful new technologies to search for novel AD-related variants. Notable NGS-based discoveries until early 2014 include the identification of rare susceptibility-modifying alleles in *APP*, *TREM2*, and *PLD3*. Of these, the latter two gene findings are "novel" in the sense that these loci had not previously been linked to AD predisposition and pathophysiology. Several additional large-scale NGS projects are currently underway and many more "*TREM2*-like" discoveries can be expected to emerge from these and other studies over the coming years.

Despite this very exciting and highly promising outlook on AD genetics research made possible by the application of NGS, a few cautionary notes appear justified. First, despite their powerful and maximally exhaustive nature, utilizing NGS in gene discovery efforts does not preclude devising a careful and typically multi-pronged study design that includes experiments aimed at establishing a firm link between the identified DNA variants and the disease under study. Examples include proving segregation within affected families, independent replication of suspected findings, and carrying out functional experiments. Otherwise, the community will be left with situations currently encountered for *SORL1* or *NOTCH3*, where some evidence suggests an involvement in AD pathogenesis, while other crucial evidence in support of these hypotheses is still lacking. Actually, it can be argued that precisely *because* of the powerful and exhaustive nature of NGS a careful study design and execution is more direly needed than ever before: every single genome analyzed thus far by WGS has been found to contain dozens to hundreds of rare and apparently functional DNA sequence that proved to be without pathogenic consequences [15]. Second, it should be emphasized that in contrast to highly penetrant and disease-causing mutations, e.g. those encountered in *APP* or *PSEN1*, rare-variant associations of modest effect size, e.g. those reported for *TREM2*, are of little value as predictors or diagnostic tools in a clinical setting [61]. This is due to their incomplete penetrance, meaning that a sizable fraction of the

population carries the disease-associated variants without ever developing AD. This situation is not unlike that observed for common variants, e.g. those identified by GWAS. Third, as outlined in the introduction, this review only covers studies utilizing NGS for genome resequencing which is only one of several possible NGS applications. This, of course, does not mean that all or even most of the "missing heritability" in AD can solely be attributed to alterations in the genomic sequence. As a matter of fact, there is both theoretical and empirical evidence suggesting that much of the underlying heritability in AD may be due to alterations beyond the genome sequence, e.g. in epigenetic DNA profiles. Their identification and characterization, however, is more complex and requires the application of other NGS-based technologies, including bisulfite sequencing (to assess DNA methylation patterns), as well as RNA and chromatin immunoprecipitation (ChIP) sequencing. However, as epigenetic profiles are often specific to tissues or even cell-types the choice of appropriate biomaterial becomes crucial, but is difficult to resolve in a brain disease such as AD.

Notwithstanding these limitations, the increasingly widespread application and further development of NGS over the coming years will undoubtedly lead to a vast extension of our knowledge and understanding of the molecular processes underlying the onset and progression of AD and other neurodegenerative diseases. As such they will hopefully pave the way for developing novel therapeutics and biomarkers allowing to effectively prevent or halt the progression of this devastating disease.

Acknowledgements

References

1. Gatz M, Reynolds CA, Fratiglioni L et al (2006) Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63:168–174

2. Tanzi RE, Bertram L (2005) Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. Cell 120:545–555

3. Bettens K, Sleegers K, Van Broeckhoven C (2013) Genetic insights in Alzheimer's disease. Lancet Neurol 12:92–104

4. Ridge PG, Ebbert MTW, Kauwe JSK (2013) Genetics of Alzheimer's disease. Biomed Res Int 2013:254954

5. Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: back to the future. Neuron 68:270–281

6. Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci 9:768–778

7. Strittmatter WJ, Saunders AM, Schmechel D et al (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90:1977–1981

8. Bertram L, McQueen MB, Mullin K et al (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39:17–23

9. Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

10. Zuk O, Hechter E, Sunyaev SR et al (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 109:1193–1198

11. Mardis ER (2013) Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto, Calif) 6:287–303

12. Eisenstein M (2012) The battle for sequencing supremacy. Nat Biotechnol 30:1023–1026

13. Zuk O, Schaffner SF, Samocha K et al (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111:E455–E464

14. Vinkhuyzen AA, Wray NR, Yang J et al (2013) Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet 47:75–95

15. Goldstein DB, Allen A, Keebler J et al (2013) Sequencing studies in human genetics: design and interpretation. Nat Rev Genet 14:460–470

16. Kiezun A, Garimella K, Do R et al (2012) Exome sequencing and the genetic basis of complex traits. Nat Genet 44:623–630

17. Thomas DC, Yang Z, Yang F (2013) Two-phase and family-based designs for next-generation sequencing studies. Front Genet 4:276

18. Bamshad MJ, Ng SB, Bigham AW et al (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12:745–755

19. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415–425

20. Sims D, Sudbery I, Ilott NE et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132

21. Wang S, Xing J (2013) A primer for disease gene prioritization using next-generation sequencing data. Genomics Inform 11:191–199

22. Normand R, Yanai I (2013) An introduction to high-throughput sequencing experiments: design and bioinformatics analysis. Methods Mol Biol 1038:1–26

23. Ladouceur M, Dastani Z, Aulchenko YS et al (2012) The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. PLoS Genet 8:e1002496

24. 1000 Genomes Project Consortium, Abecasis GR, Auton A et al (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65

25. Cruts M, Theuns J, Van Broeckhoven C (2012) Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat 33:1340–1344

26. Cruchaga C, Haller G, Chakraverty S et al (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset alzheimer's disease families. PLoS One 7:e31039

27. Jin SC, Pastor P, Cooper B et al (2012) Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res Ther 4:34

28. Benitez BA, Karch CM, Cai Y et al (2013) The PSEN1, p.E318G variant increases the risk of Alzheimer's disease in APOE-ε4 carriers. PLoS Genet 9:e1003685

29. Lord J, Turton J, Medway C et al (2012) Next generation sequencing of CLU, PICALM and CR1: pitfalls and potential solutions. Int J Mol Epidemiol Genet 3:262–275

30. Guerreiro RJ, Beck J, Gibbs JR et al (2010) Genetic variability in CLU and its association with Alzheimer's disease. PLoS One 5:e9510

31. Bettens K, Brouwers N, Engelborghs S et al (2012) Both common variations and rare non-synonymous substitutions and small insertion/deletions in CLU are associated with increased Alzheimer risk. Mol Neurodegener 7:3

32. Ferrari R, Moreno JH, Minhajuddin AT et al (2012) Implication of common and disease specific variants in CLU, CR1, and PICALM. Neurobiol Aging 33:1846.e7–e18

33. Yu J-T, Ma X-Y, Wang Y-L et al (2013) Genetic variation in clusterin gene and Alzheimer's disease risk in Han Chinese. Neurobiol Aging 34:1921.e17–e23

34. Jiang T, Yu J-T, Tan M-S et al (2014) Genetic variation in PICALM and Alzheimer's disease risk in Han Chinese. Neurobiol Aging 35:934.e1–e3

35. Ma X-Y, Yu J-T, Tan M-S et al (2014) Missense variants in CR1 are associated with increased risk of Alzheimer' disease in Han Chinese. Neurobiol Aging 35:443.e17–e21

36. Lupton MK, Proitsi P, Danillidou M et al (2011) Deep sequencing of the Nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease. PLoS One 6:e17298

37. Lupton MK, Proitsi P, Lin K et al (2014) The role of ABCA1 gene sequence variants on risk

of Alzheimer's disease. J Alzheimers Dis 38: 897–906

38. Bertram L (2011) Alzheimer's genetics in the GWAS era: a continuing story of "replications and refutations". Curr Neurol Neurosci Rep 11:246–253

39. Guerreiro RJ, Lohmann E, Kinsella E et al (2012) Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease. Neurobiol Aging 33:1008.e17–e23

40. Pottier C, Hannequin D, Coutant S (2012) High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset alzheimer disease. Mol Psychiatry 17:875–879

41. Andersen OM, Reiche J, Schmidt V et al (2005) Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. Proc Natl Acad Sci U S A 102:13461–13466

42. Lambert J-C, Ibrahim-Verbaas CA, Harold D et al (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45:1452–1458

43. Cruchaga C, Karch CM, Jin SC et al (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505:550–554

44. Heilmann, Stefanie, Dmitriy Drichel, Jordi Clarimon, Victoria Fernández, André Lacour, Holger Wagner, Mathias Thelen, et al. 2015. "PLD3 in Non-Familial Alzheimer's Disease." Nature 520 (7545): E3–5. doi:10.1038/nature14039.

45. Hooli, Basavaraj V., Christina M. Lill, Kristina Mullin, Dandi Qiao, Christoph Lange, Lars Bertram, and Rudolph E. Tanzi. 2015. "PLD3 Gene Variants and Alzheimer's Disease." Nature 520 (7545): E7–8. doi:10.1038/nature14040.

46. Lambert, Jean-Charles, Benjamin Grenier-Boley, Céline Bellenguez, Florence Pasquier, Dominique Campion, Jean-Francois Dartigues, Claudine Berr, Christophe Tzourio, and Philippe Amouyel. 2015. "PLD3 and Sporadic Alzheimer's Disease Risk." Nature 520 (7545): E1. doi:10.1038/nature14036.

47. Van der Lee, Sven J., Henne Holstege, Tsz Hang Wong, Johanna Jakobsdottir, Joshua C. Bis, Vincent Chouraki, Jeroen G. J. van Rooij, et al. 2015. "PLD3 Variants in Population Studies." Nature 520 (7545): E2–3. doi:10.1038/nature14038.

48. Cai D, Zhong M, Wang R et al (2006) Phospholipase D1 corrects impaired betaAPP trafficking and neurite outgrowth in familial Alzheimer's disease-linked presenilin-1 mutant neurons. Proc Natl Acad Sci U S A 103: 1936–1940

49. Oliveira TG, Chan RB, Tian H et al (2010) Phospholipase d2 ablation ameliorates Alzheimer's disease-linked synaptic dysfunction and cognitive deficits. J Neurosci 30: 16419–16428

50. Jonsson T, Atwal JK, Steinberg S et al (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96–99

51. Jonsson T, Stefansson H, Steinberg S et al (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116

52. Kero M, Paetau A, Polvikoski T et al (2013) Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. Neurobiol Aging 34:1518.e1–e3

53. Liu Y-W, He Y-H, Zhang Y-X et al (2014) Absence of A673T variant in APP gene indicates an alternative protective mechanism contributing to longevity in Chinese individuals. Neurobiol Aging 35:935.e11–e12

54. Ting SKS, Chong M-S, Kandiah N et al (2013) Absence of A673T amyloid-β precursor protein variant in Alzheimer's disease and other neurological diseases. Neurobiol Aging 34: 2441.e7–e8

55. Bamne MN, Demirci FY, Berman S et al (2014) Investigation of an amyloid precursor protein protective mutation (A673T) in a North American case-control sample of late-onset alzheimer's disease. Neurobiol Aging 35:1779.e15–e16

56. Peacock ML, Warren JT Jr, Roses AD et al (1993) Novel polymorphism in the A4 region of the amyloid precursor protein gene in a patient without Alzheimer's disease. Neurology 43:1254–1256

57. Lill CM, Liu T, Schjeide BMM et al (2012) Closing the case of APOE in multiple sclerosis: no association with disease risk in over 29 000 subjects. J Med Genet 49:558–562

58. Guerreiro R, Wojtas A, Bras J et al (2013) TREM2 variants in Alzheimer's disease. N Engl J Med 368:117–127

59. Pottier C, Wallon D, Rousseau S et al (2013) TREM2 R47H variant as a risk factor for early-onset alzheimer's disease. J Alzheimers Dis 35:45–49

60. Benitez BA, Cooper B, Pastor P et al (2013) TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiol Aging 34:1711.e15–e17

61. Bertram L, Parrado AR, Tanzi RE (2013) TREM2 and neurodegenerative disease. N Engl J Med 369:1565

62. Benitez BA, Cruchaga C, United States–Spain Parkinson's Disease Research Group (2013) TREM2 and neurodegenerative disease. N Engl J Med 369:1567–1568

63. Rayaprolu S, Mullen B, Baker M et al (2013) TREM2 in neurodegeneration: evidence for association of the p.R47H variant with fronto-temporal dementia and Parkinson's disease. Mol Neurodegener 8:19

64. Cady J, Koval ED, Benitez BA et al (2014) TREM2 Variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis. JAMA Neurol 71:449–453

65. Lill CM, Rengmark A, Lasse P, Fogh I, Shatunov S, Sleiman PM, Wang LS et al. (2015) The role of TREM2 R47H as a risk factor for Alzheimer's Disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's sisease." Alzheimer's & Dementia; doi:10.1016/j.jalz.2014.12.009

66. Neumann H, Takahashi K (2007) Essential role of the microglial triggering receptor expressed on myeloid cells-2 (TREM2) for central nervous tissue immune homeostasis. J Neuroimmunol 184:92–99