EDITORIAL

Probing the Exome in Alzheimer Disease and Other Neurodegenerative Disorders

Lars Bertram, MD; Christine Klein, MD

Alzheimer disease (AD) is a heterogeneous disorder with a substantial genetic component. A small number of cases (ie, early-onset familial AD) are caused by exceedingly rare but pathogenic and highly penetrant mutations, while most cases (ie, late-onset AD) are caused by an intricate—and still only partially understood—interplay of genetic and nongenetic risk factors.[1] The past decade has seen unprecedented progress in deciphering the genetic underpinnings of late-onset AD. This advancement was achieved mostly by the application of high-throughput microarray genotyping in the context of genome-wide association studies (GWASs) comparing the allele status at millions of different base pairs on increasingly large samples of affected and unaffected individuals.[2] Most AD GWAS findings to date were made with common (ie, frequency of the minor allele typically >5%) single-nucleotide polymorphisms (SNPs) typically exerting small genetic effect sizes (ie, odds ratios <1.3). In most cases, the pathogenic mechanisms underlying these associations have been difficult to discern owing to the fact that most common SNPs are located in noncoding regions of the genome.

Related article

The study by Chen et al[3] in this issue is different from most other GWASs in the field because it specifically focused on SNPs located in the coding—and thus potentially functional—portions of the genome (ie, the exome). They achieved this by using a microarray (the HumanExome Array, Illumina, Inc) specifically designed to capture the mostly rare (ie, frequency <5%, applicable to approximately 90% of all markers on the array) exome SNPs in a sample of 224 patients with AD and 224 healthy control subjects. Additional genotyping was completed on 168 patients with frontotemporal dementia and 48 patients with progressive supranuclear palsy. While the analyses of frontotemporal dementia and progressive supranuclear palsy did not yield any noteworthy new results, at least 2 novel putative loci, ie, *DYSF* and *PAXIP1*, were highlighted by the authors for AD, in addition to very strong signals near the well-established *APOE* locus on chromosome 19. Overall, Chen et al[3] estimated that about 44% of the variance in case-control status in their sample could be explained by the SNPs typed on the exome array, a number substantially higher than the approximately 25% reported from an earlier and much larger GWAS meta-analysis mostly studying common variants.[2]

The main strength of the study by Chen et al[3] is that it represents one of the first published applications of the exome array technology to AD (and frontotemporal dementia and progressive supranuclear palsy). This technology is meant to serve as a cost-effective interim solution to studying the genetics of complex traits in times when more comprehensive approaches, such as whole-exome sequencing or whole-genome sequencing, are still relatively expensive. The design of the exome array used by Chen et al[3] resulted from a comprehensive analysis of whole-exome sequencing and whole-genome sequencing data of approximately 12 000 individuals with the aim to specifically capture all nonprivate and putatively functional (ie, nonsynonymous, nonsense, splice) base-pair changes (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). It is estimated that 94% to 98% of these types of variants identified in an average genome by whole-exome sequencing are also captured by the exome array. While this coverage should suffice for most exome variant association studies, this type of microarray technology—by design and unlike whole-exome sequencing or whole-genome sequencing—is not able to discover any novel sequence variants. Instead, the exome array is limited to genotyping polymorphisms previously shown to exist at sufficient frequency in the general population and, conceptually, is therefore no different from a conventional (ie, focusing on common variants not selected based on functionality) GWAS or smaller-scale association study. Despite this limitation, almost 90% of the variants genotyped on the exome array used by Chen et al[3] were not captured or analyzed in the most comprehensive AD GWAS published to date[2] (variant overlap determined for the purpose of this editorial is based on data releases from Lambert et al[2] and http://www.illumina.com). Differently put, approximately 213 000 SNPs, most of which were predicted to be functional, genotyped in the study by Chen et al[3] were not previously tested in patients with AD and controls of European descent, although similar studies are currently under way (see below). The only other exome genotyping study for AD published to date was performed on 400 patients with late-onset AD and 605 controls from Korea using a similar exome microarray (manufactured by Affymetrix, Inc).[4]

While the study by Chen et al[3] thus uses state-of-the-art technology, its scientific implications currently remain limited. This is mostly owing to the small sample size investigated. In their discovery phase, Chen et al[3] analyzed only 416 patients with AD and controls, followed by replication analyses in an independent set of 480 individuals. Thus, the overall sample size analyzed in the AD portion of this study was well below 1000 individuals. It was even smaller for the frontotemporal dementia (n = 168) and progressive supranuclear palsy (n = 48) analyses, for which no replication analyses were

performed and, accordingly, the results of which are not discussed here. Even for common SNPs of large effect, such as variants in or near *APOE*, the study by Chen et al[3] was not sufficiently sized to detect variant-level evidence for association at genome-wide significance, defined by most authors as $P < 5 \times 10^{-8}$ owing to the large number of common variants in the human genome and the resulting multiple-testing burden.[5] Power is further reduced for alleles of lower frequency and/or lower effect size, ie, precisely the type of variant(s) targeted for analysis in the study by Chen et al.[3] This limitation was acknowledged by the authors and represents a problem that AD genetic association studies have faced for decades.[6] The multiple-testing problem is alleviated somewhat by performing gene-level analyses, a strategy that allows researchers to jointly analyze variants located in the same genes, thereby reducing the number of comparisons. This approach was also used in the study by Chen et al[3] and eventually led to the identification of putative association signals with *DYSF* and *PAXIP1*. However, these observations were made only when applying a relatively liberal false discovery rate threshold of 50% but were no longer apparent when using a more stringent (15%) false discovery rate cutoff. As a result, it will be difficult to deduce any firm conclusions on the role of *DYSF* and *PAXIP1* in AD pathogenesis.

A number of lines of evidence, albeit indirect, suggest that a cautious interpretation of the results by Chen et al[3] may be warranted at this time. First, neither *DYSF* nor *PAXIP1* were reported to show any noteworthy evidence of association in the largest (N = 74 046) AD GWAS published to date (from investigators of the International Genomics of Alzheimer's Project).[2] However, this latter study was targeted mostly toward common SNPs and included only very few rare and/or potentially functional variants in the *DYSF* and *PAXIP1* regions. Second, and probably more important for the purpose of this editorial, investigators of the International Genomics of Alzheimer's Project recently reported the results of a GWAS on a subsample (>30 000 individuals)[7] of the Lambert et al[2] data set using the same exome array also analyzed by Chen et al.[3] In that study, neither *DYSF* nor *PAXIP1* were highlighted as lead findings.[7] However, this much larger exome array GWAS is still in progress and evidence for these 2 loci needs to be reassessed once the final data are published. Finally, following a reverse line of reasoning, the only rare variant known to date to show genome-wide significant association with increased AD risk (ie, p.R47H [rs75932628] located in *TREM2*)[8] was not specifically highlighted in the study by Chen et al.[3] Since *TREM2* must be considered as the only currently available positive control in the context of performing a rare-variant GWAS in AD, failure to detect this signal may also have consequences for the proposed associations with *DYSF* and *PAXIP1*. Regardless, the availability of well-designed and powerful new genotyping and high-throughput sequencing technologies has already led to a diverse range of exciting discoveries in genetics research of AD and other complex traits. All of today's established genes had to undergo several, often painstaking, rounds of independent assessments before admission to the pantheon of AD genetics findings.

The data now published by Chen et al[3] represent a first important step in assessing the role that rare variants play in the genetic architecture underlying AD susceptibility. Once released into the public domain, these data will undoubtedly serve as a valuable resource toward achieving this goal.

ARTICLE INFORMATION

Author Affiliations: Platform for Genome Analytics, Institute of Neurogenetics, University of Lübeck, Lübeck, Germany (Bertram, Klein); Institute of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany (Bertram); School of Public Health, Faculty of Medicine, Imperial College, London, United Kingdom (Bertram).

Corresponding Author: Lars Bertram, MD, Platform for Genome Analytics, Institute of Neurogenetics, University of Lübeck, Maria-Goeppert-Strasse 1 (MFC1), 23562 Lübeck, Germany (lars.bertram@uni-luebeck.de).

Published Online: February 23, 2015. doi:10.1001/jamaneurol.2014.4495.

Conflict of Interest Disclosures: None reported.

Funding/Support: Dr Klein has received a Career Development Award from the Hermann and Lilly Schilling Foundation.

Role of the Funder/Sponsor: The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Bertram L, Tanzi RE. The genetics of Alzheimer's disease. *Prog Mol Biol Transl Sci*. 2012;107:79-100.

2. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452-1458.

3. Chen JA, Wang Q, Davis-Turak J, et al. A multiancestral genome-wide exome array study of Alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy [published online February 23, 2015]. *JAMA Neurol*. doi:10.1001/jamaneurol.2014.4040.

4. Chung SJ, Kim MJ, Kim J, et al. Exome array study did not identify novel variants in Alzheimer's disease. *Neurobiol Aging*. 2014;35(8):1958.e13-1958.e14.

5. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356-369.

6. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39(1):17-23.

7. Naj AC, van der Lee SJ, Vronskaya M, et al. Exome array analysis of 30,582 individuals confirms late-onset Alzheimer's disease (LOAD) risk from common variants and identifies novel rare LOAD susceptibility variants: the International Genomics of Alzheimer's Project (IGAP). Presented at: American Society of Human Genetics 2014 Meeting; October 18-22, 2014; San Diego, CA. Abstract 196.

8. Lill C, Rengmark A, Philstrom L, et al. The role of *TREM2* R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. *Alzheimers Dement*. In press.