Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Alzheimer's disease: rare variants with large effect sizes

Jorge L Del-Aguila[1,2], Daniel C Koboldt[3], Kathleen Black[1,2], Rachel Chasse[1,2], Joanne Norton[1,2], Richard K Wilson[3] and Carlos Cruchaga[1,2]

Recent advances in sequencing technology and novel genotyping arrays (focused on low-frequency and coding variants) have made it possible to identify novel coding variants with large effect sizes and also novel genes (*TREM2, PLD3, UNC5C*, and *AKAP9)* associated with Alzheimer's disease (AD) risk. The major advantages of these studies over the classic genome-wide association studies (GWAS) include the identification of the functional variant and the gene-driven association. In addition to the large effect size, these studies make it possible to model these variants and genes using cell and animal systems. On the other hand, the underlying population-variability of these very low allele frequency variants poses a great challenge to replicating results. Studies that include very large datasets (>10,000 cases and controls) and combine sequencing and genotyping approaches will lead to the identification of novel genes for Alzheimer's disease.

Addresses
[1] Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave. B8134, St. Louis, MO 63110, USA
[2] Hope Center for Neurological Disorders. Washington University School of Medicine, 660 S. Euclid Ave. B8111, St. Louis, MO 63110, USA
[3] The Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA

Corresponding author: Cruchaga, Carlos (cruchagc@wustl.edu)

Current Opinion in Genetics & Development 2015, **33**:49–55

This review comes from a themed issue on **Molecular and genetic bases of disease**

Edited by **Dan E Arking** and **Johanna M Rommens**

http://dx.doi.org/10.1016/j.gde.2015.07.008

Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disorder affecting more than 5.3 million people in the U.S. [1]. It is a complex disease characterized by gradual onset and progression of memory loss combined with deficits in executive functioning, language, visuospatial abilities, personality, behavior and self-care. AD can be divided into two subtypes depending on the presence or absence of familial aggregation: respectively, familial and sporadic AD. Sporadic AD is normally characterized by later onset (>60 years) and complex patterns of inheritance. While they differ in familial clustering, both are defined by the same pathological features: neuronal loss and the presence of Aβ plaques and neurofibrillary tangles.

Mutations in the *amyloid-beta precursor protein (APP) presenilin (PSEN1)* and *presenilin 2 (PSEN2)* genes cause Mendelian forms of AD [2]. These mutations have been identified in only a small number of people (500 families worldwide; http://www.molgen.ua.ac.be/ADMutations/), but the identification of such mutations and genes has led to a better understanding of the biology of AD (reviewed in [3]). Much of the recent genetic research of AD has used genome-wide association studies (GWAS) to identify common variants associated with disease risk. Lambert *et al.* presented a meta-analysis including GWAS data for more than 74,000 AD cases and controls [4•]. This study reported more than 20 loci that were significantly associated with AD risk. Despite the success of GWAS in identifying new loci for AD, these studies have two major problems: First, each locus only accounts for a small proportion of the variance in AD susceptibility [5] and second, these studies identify genetic regions, not genes. In some cases, it is not possible to determine which gene or variant is driving the association, especially when the locus is located on a gene-rich region with a high number of genetic variants in high linkage disequilibrium. In any case, these results further establish that AD is a complex and multifactorial disease [6–11].

Because of the intrinsic problems with GWAS, there is great interest in identifying low-frequency and rare variants with a large effect size for AD risk. The identification of such variants unequivocally points to a single gene. For this reason and also because of the large effect size, it is possible to functionally characterize and develop cell and animal models to study the role of the identified variants and genes in AD. Until recently, identification and association of those alleles remained a complex task; however, recent advances in sequencing technology have made it possible to study those alleles on a genome-wide scale and in large populations. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) have proven to be very powerful techniques to identify novel genes associated not only with Mendelian disorders [12–14], but also with complex traits [15]. Additionally, several companies in the last few years have developed genotyping arrays focused on low-frequency and rare coding variants. These new genotyping arrays provide a powerful and

affordable approach to identify novel variants and genes associated with complex traits. However, these studies also have some challenges. Most coding variants are very rare (minor allele frequency [MAF] < 0.01) so extremely large sample sizes are needed to identify significant genome-wide associations. Alternatively, several new statistical methods have been developed [16–19] that analyze entire genes rather than single variants. These tests provide more power than the single-variant analyses and are based on the hypothesis that there will be a significant difference in the frequency of coding variants in cases compared to controls (Table 1).

In this review, we will focus on the rare and low-frequency coding variants with large effect sizes in addition to novel genes that have been identified in the last few years using novel sequencing technologies. We will also discuss the major challenges concerning the validation of those studies.

TREM2: triggering receptor expressed on myeloid cells 2

In 2013, low-frequency coding variants in *TREM2* were identified as being high-risk for AD [20••,21••]. This could be considered the first instance of a novel gene being identified for AD risk using next-generation sequencing technologies.

TREM2 is a type 1 transmembrane receptor protein expressed on myeloid cells including microglia, monocyte-derived dendritic cells, osteoclasts and bone-marrow derived macrophages [22,23]. *TREM2* transduces its intracellular signaling through DAP12 (TYROBP) [22,23]. Although the natural ligands of *TREM2* remain unknown, upon ligand binding *TREM2* associates with DAP12 to mediate downstream signaling. In the brain, *TREM2* is primarily expressed on microglia and has been shown to control two signaling pathways: regulation of phagocytosis and suppression of inflammatory reactivity [24,25].

Recessive *TREM2* loss-of-function mutations cause Nasu-Hakola disease (NHD) or polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (PLOSL; MIM 221770) which is characterized by early onset frontotemporal-like (FTD-like) dementia and bone involvement [27,28]. In addition, some families with inherited recessive FTD-like dementia with leukodystrophy but without bone involvement have *TREM2* mutations [28,29]. These findings prompted an effort to identify allelic variants in the *TREM2* coding region in AD patients [30]. This endeavor continued without marked success until recent years when two independent groups reported that a non-synonymous variant in *TREM2*, rs75932628 (encoding R47H), is strongly associated with AD [20••,21••]. In the first study, Jonsson *et al.* [20••] identified the *TREM2* R47H variant [$p = 3.42 \times 10^{-10}$, odds ratio (OR) = 2.92 (95% confidence interval (CI) 2.09-4.09)] by performing WGS sequencing in 2261 Icelandic individuals, and using GWAS data and pedigree information to perform imputation on more than 12,000 individuals. Guerreiro's concurrent study also reported the association of the R47H variant with risk for AD using a candidate gene approach [21••].

Since these two initial reports, the association of the R47H variant has been replicated in multiple populations of European-descent [31,32•,33–35]. However, other studies have failed to replicate this association in Asian populations [36]. These findings raise the important question of which strategy could best replicate rare or low-frequency variants. These particular variants present high variability of MAF among different populations and even within populations of same origin. For example, the MAF of the R47H variant varies from 0.63% in Icelandic population to 0.12–0.26% [20••] in the USA, 0.9% in France, 1% in North Europe and England, and 2% [21••] in some British populations. If cases and controls are not matched carefully, association and replication studies could lead to both false positive and false negative findings because of these differences in MAF, even if all populations are strictly of European origin.

Additionally, diverse populations with different genetic backgrounds can harbor several variants within the same gene that contribute to AD risk (as in the case of *APP*, *PSEN1* and *PSEN2*). As a result, it has been proposed that gene-based and resequencing studies rather than single-variant analyses should be performed to replicate these

Table 1

Summary of the major findings in new variants in genes related to AD.

Gene	SNP	Codon Variation	Chr.	OR (95% CI)	*P* value	References
TREM2	rs75932628	R47H	6	2.92 (2.09–4.09)	3.42×10^{-10}	[20••]
	rs143332484	R62H		2.36 (1.47–3.08)	2.36×10^{-4}	[34]
PLD3	rs145999145	V232M	19	2.1 (1.47–2.99)	2.93×10^{-5}	[47••]
UNC5C	rs137875858	T835M	4	2.15 (1.21–3.84)	9.5×10^{-3}	[56••]
AKAP9	rs144662445	—	7	2.75	2.2×10^{-3}	[58••]
	rs149979685	—		3.61	2.2×10^{-3}	

Odd ratio and *p* values for meta-analysis.