Molecular and Cellular Probes xxx (2016) 1–7

Contents lists available at ScienceDirect

Molecular and Cellular Probes

journal homepage: www.elsevier.com/locate/ymcpr

Genomics of Alzheimer's disease: Value of high-throughput genomic technologies to dissect its etiology

Giuseppe Tosto [a, b, c], Christiane Reitz [a, b, c, d, *]

[a] The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY, USA
[b] The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
[c] The Department of Neurology, Columbia University, New York, NY, USA
[d] The Dept. of Epidemiology, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form
8 September 2016
Accepted 8 September 2016
Available online xxx

Keywords:
Genetics
Genome-wide association study
Alzheimer's disease
High-throughput technologies
Rare variants
Epidemiology of neurodegenerative disorders

ABSTRACT

Late-onset Alzheimer's disease (AD), the most common neurodegenerative disorder in western countries, is clinically defined by progressive worsening in cognitive functions along with function and behavioral impairment. This ultimately results in complete incapacity and death. AD is a clinically and pathologically heterogeneous disease, and this is reflected by the numerous genetic findings that point to several diverse molecular mechanisms and pathways. Linkage, genome-wide association and next-generation sequencing studies have led to the identification of more than 20 novel susceptibility loci for AD. While these observations have significantly increased the knowledge of pathogenic mechanisms and potential therapeutic targets, a large part of the genetic component underlying AD is still unexplained. This review will summarize and discuss the major genetic findings and their potential impact on AD diagnosis and prediction of prognosis.

1. Introduction

In the US alone, about 5.5 million individuals are affected with late-onset Alzheimer's disease (AD) [1] and this number is destined to increase dramatically in the next few decades. The predominant clinical features of AD are a progressive, and ultimately fatal, worsening in cognition, function and behavior, but a broad phenotypic profile is commonly observed both in terms of timing of onset and rapidity of progression [2,3]. Pathological findings at brain autopsy are intracellular deposits of hyperphosphorylated tau protein in the form of neurofibrillary tangles, and extracellular β-amyloid (Aβ) protein in diffusible oligomers and insoluble plaques. Neuropathological findings are heterogeneous as well, with common observation of Lewy bodies and/or vascular changes [2]. Available drugs mostly belong to two classes of compounds, namely acetylcholinesterase inhibitors and NMDA receptor antagonists. Both are widely used and are relatively safe to use, but systematic reviews (i.e. Cochrane reviews) failed to demonstrate any significant benefit in terms of disease prevention or in slowing of disease progression. Mapping the genes and molecular pathways linked to the disease will be critical for identifying more effective preventive and therapeutic targets.

2. Identification of *APP*, *PSEN1* and *PSEN2* as AD risk genes by early linkage studies

Autosomal dominant mutations in the *APP*, *PSEN1* and *PSEN2* genes [4–6], were discovered by linkage studies conducted in large families multiply affected by early-onset AD (onset age: <50 years old) in the 1990's. A genetic linkage study is a family-based method used to map a disease or trait to a genomic location by demonstrating its co-segregation with genetic markers throughout a pedigree. This approach leverages the fact that co-segregating genomic regions are more likely to contain a causal genetic variant compared to regions without co-segregation.

* Corresponding author. The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, 630 W 168th Street, New York, NY 10032, USA.
E-mail address: cr2101@cumc.columbia.edu (C. Reitz).

http://dx.doi.org/10.1016/j.mcp.2016.09.001
0890-8508/

The successful mapping of *APP*, *PSEN1* and *PSEN2* genes as disease loci helped to formulate the "amyloid cascade hypothesis". The latter indicated as a key pathogenic step the enhanced generation of Amyloid beta protein (Aβ) fragments generated through alternative pathogenic cleavage of the amyloid precursor protein (APP) by β- and ϒ-secretases, as opposed to the non-amylogenic cleavage by α-secretase. Aβ peptides generated through the amyloidogenic pathway are prone to aggregation, leading to toxic Aβ oligomers and amyloid plaque formation. As of to date (March 2016), 49 pathogenic mutations in *APP*, 216 mutations in *PSEN1* and 16 mutations in *PSEN2* have been mapped (http://www.molgen.ua.ac.be/ADMutations/).

Linkage studies in multiplex AD families also led to identification of the first susceptibility gene identified in the common, late-onset form of AD (LOAD), i.e. Apolipoprotein E (*APOE)* in 1993 [7,8]. In humans, the APOE gene has 3 exons and is located on chromosome 19; it has three isoforms exhibiting major differences in effects on structure and function of APOE. These isoforms are encoded by the APOE-ε2, ε3, and ε4 alleles.

A single *APOEe4* allele increases LOAD risk by ~3 fold [9,10] and lowers memory function as well as the age at onset of AD [7,11–14]. The mechanisms through which APOE increases AD risk is not fully clarified. In peripheral tissues, APOE is primarily produced by the liver and macrophages, and mediates blood lipids and cardiovascular risk profile in an isoform-dependent manner. Bennet et al. showed that *APOE- ε 2* is associated with lower LDL and reduced risk of coronary artery disease, whereas the presence of the ε 4 allele increased both [15]. In the central nervous system, APOE traffics cholesterol resulting from neurodegeneration to neurons requiring them for membrane repair, proliferation, or remyelinization. In addition, APOE affects glutamate receptor function and synaptic plasticity by modulating neuronal APOE receptor recycling [16]. *APOE-ε4* dosage at autopsy correlates strongly with Aβ load and plaque accumulation [17,18]. Correspondingly, in *in vitro* studies the *APOE-ε4* isoform binds Aβ peptides more vigorously than *APOE*-ε3 [19] and leads to increased Aβ aggregation [*20,21*]; knockout *APOE*−/− mice develop less nonfibrillar Aβ deposits [22,23]. APOE may also indirectly affect clearance of Aβ by competing for the same astrocyte receptor, low-density lipoprotein receptor-related protein 1 (LRP1), which helps clear Aβ [24], Finally, *APOE-ε4* may stimulate tau phosphorylation leading to pre-neurofibrillary tangles [25].

Notably, there are important differences among ethnic groups: *APOE-ε4* association with AD is in fact weaker among African Americans and Hispanics [26]. Our group reported that Caribbean Hispanics have an increased prevalence of LOAD regardless of their *APOE* genotype; in the absence of an ε4 allele, African American and Hispanics show 2 to 4 times fold higher incidence of LOAD as compared to the Whites [26]. Moreover, we also showed that, among Caribbean Hispanics, familial LOAD is strongly associated with APOE-Ɛ4 whereas sporadic LOAD has only weak association with the risk allele [27].

Two recent linkage studies in Caribbean Hispanics and non-Hispanic Whites conducted as part of the Alzheimer's Disease Sequencing Project (ADSP), identified several additional genomic regions with strong evidence for linkage to LOAD [28,29], with the strongest signals at 11q12.3 (a locus located ~2 Mb upstream of the membrane-spanning 4A gene cluster) in both ethnic groups, 7p14.3 (harboring several genes associated with the nervous system (GARS, GHRHR, and NEUROD6)) in Caribbean Hispanics, and 3q25.31, 4q34.1, 8q22.3, 11q12.2–14.1, and 19q13.41 in non-Hispanic Whites. These findings are currently followed up by whole genome sequencing analyses in order to identify potential rare disease-associated variants.

3. Susceptibility loci associated with late-onset AD identified by high-throughput technologies

Common disorders such as the late-onset form of AD are caused by a combination of multiple environmental risk factors and common and rare genetic variants in genes belonging to various pathogenic pathways ("polygenic disease", "complex disorder"). Under this scenario, a genetic association study can demonstrate higher statistical power than the linkage approach if the sample size is sufficiently large.

Over the past decade, the development of high-throughput massively parallel technologies for disentangling disease at the molecular level has revolutionized medical science and has significantly advanced the understanding of the molecular underpinnings of AD. These "next generation" genomic technologies include genome-wide genotyping arrays for genome-wide association studies (GWAS), whole-exome (WES) or whole-genome (WGS) sequencing for detection of disease-associated DNA sequence variants ("genomics"), and RNA sequencing (RNA-seq) for transcriptome and non-coding RNA analysis ("transcriptomics").

3.1. AD susceptibility loci identified by GWAS studies

By definition, genotyping chips for GWAS are designed to examine common variation with a minor allele frequency (MAF) > ~1–5% across the whole genome. As we recently reviewed [30], the large AD GWAS performed to date focused mostly on individuals of non-Hispanics Whites of European ancestry [31–35]. In addition to *APOE*, over 20 novel loci have been identified by these studies. Common to all these loci are the associated small effect sizes with Odd Ratios between 1.1 and 1.3, suggesting that a large part of the genetic component of LOAD is still unexplained. The largest GWAS to date, conducted in 2013 by the International Genomics of Alzheimer's Project (IGAP) aggregated 74,046 non-Hispanic Whites [36] through a multi-center collaboration. The study confirmed at genome-wide significance for *CR1*, *BIN1*, *CD2AP*, *EPHA1*, *CLU*, *MS4A6A*, *PICALM*, *ABCA7*and *CD33*, previously identified by the first set of GWAS, and identified several novel associations: *HLA-DRB5/HLA-DRB1*, *PTK2B*, *SORL1*, *SLC24A4/RIN3*, *DSG2*, *INPP5D*,*MEF2C*, *NME8*, *ZCWPW1*, *CELF1*, *FERMT2* and *CASS4* [37]. Most of these genes cluster in a specific set of pathways, namely immune response (*HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*), APP processing (*SORL1*, *CASS4*), Tau pathology (*CASS4*, *FERMT2*), cell migration (*PTK2B*) and lipid transport and endocytosis (*SORL1*), strongly reinforcing the importance of these molecular mechanisms in AD etiology. Of note, *SORL1* (sortilin-related receptor, L (DLR class) 1) had previously been demonstrated to modulate trafficking and processing of APP in a candidate gene approach [37]. In addition, consistent with the notion of a complex disease, the findings of this multicenter project further strengthen the evidence for additional molecular mechanisms including hippocampal synaptic function (MEF2C, *PTK2B)*; cytoskeletal function and axonal transport (*CELF1*, *NME8*, CASS4); regulation of gene expression and post-translational modification of proteins, microglial and myeloid cell function (*INPPD5*).

Due to differences in allele frequencies and LD patterns, loci detected by genomic studies can differ between ethnic groups. In the largest GWAS performed to date in African Americans, *ABCA7* was confirmed as an AD susceptibility locus, notably with an effect size similar to that of *APOE* [38]. In addition, *CR1*, *BIN1*, *EPHA1* and *CD33* were confirmed in gene-based analyses. Our group recently published the largest GWAS in Caribbean Hispanics to date: a novel locus in the *FBXL7* gene was found associated with AD, along with other known-loci previously identified in GWAS of European

Please cite this article in press as: G. Tosto, C. Reitz, Genomics of Alzheimer's disease: Value of high-throughput genomic technologies to dissect its etiology, Molecular and Cellular Probes (2016), http://dx.doi.org/10.1016/j.mcp.2016.09.001