Biochemical and Biophysical Research Communications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Review

Genomic aspects of sporadic neurodegenerative diseases

Jun Mitsui, Shoji Tsuji *

Department of Neurology, The University of Tokyo, Graduate School of Medicine, Japan

ARTICLE INFO

Article history:
Received 8 April 2014
Available online xxxx

Keywords:
Sporadic neurodegenerative disease
Genome-wide association studies
Next-generation sequencing technologies
Common disease-multiple rare variants hypothesis

ABSTRACT

Sporadic neurodegenerative diseases are complex in nature, that is, they involve multiple genetic and environmental factors that may play roles at the molecular level. In contrast to diseases with Mendelian inheritance, the genomic signatures of common sporadic forms of neurodegenerative diseases largely remain unknown. Over the past decade, genome-wide association studies employing common single-nucleotide polymorphisms have been intensively conducted, in which the theoretical framework is based on the "common disease–common variants" hypothesis. Another paradigm is a sequence-based association study under the "common disease–multiple rare variants" hypothesis. Because current next-generation sequencing technologies enable us to obtain virtually all the variants in human genome irrespective of allele frequencies, it is anticipated that sequence-based association studies will become the mainstream approach. In this review, we present brief overviews of molecular genetic approaches to elucidate the molecular bases of sporadic forms of neurodegenerative diseases, including Alzheimer disease, Parkinson disease, and multiple system atrophy as examples.

Contents

1. Introduction

Neurodegenerative disease is an umbrella term for various types of disease characterized clinically by ages at onset usually in the late adulthood and by specific neurological symptoms such as cognitive decline, ataxia, parkinsonism, or motor weakness, and neuropathologically by the progressive neuronal dysfunctions usually accompanied by cell death of specific groups of neurons. Histopathologically, many neurodegenerative diseases have distinctive inclusion bodies in neurons as well as glial cells that are associated with the accumulation of misfolded and aggregated proteins, such as Lewy bodies in Parkinson disease (PD), senile plaques in Alzheimer disease (AD), TDP-43-positive inclusions in amyotrophic lateral sclerosis (ALS), and glial cytoplasmic inclusions (GCIs) in multiple system atrophy (MSA) [1].

Although palliative therapies that temporarily and partially relieve symptoms are currently available for several neurodegenerative diseases, such as choline esterase inhibitors or *N*-methyl-D-aspartate (NMDA) antagonists for AD, there remains no disease-modifying therapy that directly targets the underlying neurodegenerative processes or halts the progression of these diseases. Elucidation of the molecular bases that underlie the

Abbreviations: AD, Alzheimer disease; PD, Parkinson disease; MSA, multiple system atrophy; ALS, amyotrophic lateral sclerosis; NMDA, *N*-methyl-D-aspartate; GWASs, genome-wide association studies; SNPs, single-nucleotide polymorphisms; GCIs, glial cytoplasmic inclusions.

* Corresponding author. Address: Department of Neurology, University of Tokyo, Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Fax: +81 3 5800 6844.
E-mail address: tsuji@m.u-tokyo.ac.jp (S. Tsuji).

http://dx.doi.org/10.1016/j.bbrc.2014.07.098
0006-291X/

pathogenesis of neurodegenerative diseases is the most fundamental basis for the development of disease-modifying therapies [2].

The majority of patients with neurodegenerative diseases have sporadic diseases without any familial occurrence, however, we infrequently encounter patients with neurodegenerative diseases presenting with rare familial forms that are consistent with Mendelian inheritance. The term "sporadic" may be oversimplistic, because many neurodegenerative diseases, such as AD, PD, or ALS, have shown a familial aggregation to a certain extent, which suggests a contribution of genetic factors to the pathogenesis of sporadic forms [3–5]. Therefore, they are complex in nature, that is, they involve multiple genetic and environmental factors that may play roles at the molecular level [6].

Regarding the familial cases with Mendelian inheritance, positional cloning strategies have been quite successful in identifying the causative genes [7]. As of April 2014, a total of 3195 genes with phenotype-causing mutations have been described in Online Mendelian Inheritance in Man (http://omim.org/statistics/geneMap) [8]. More recently, next-generation sequencing (NGS) technologies have been developed, which has further accelerated the process of finding causative genes for diseases with Mendelian inheritance over the past few years [9–17].

In contrast to diseases with the Mendelian inheritance, the genomic signatures of common sporadic forms of neurodegenerative diseases largely remain unknown. Over the past decade, genome-wide association studies (GWASs) employing common single-nucleotide polymorphisms (SNPs) have been intensively conducted to identify genomic variants associated with complex diseases [18]. The theoretical framework of GWASs is based on the "common disease-common variants" hypothesis [19–21], in which common diseases are attributable in part to relatively common variants present in more than 5% of the population. Although GWASs have successfully revealed numerous SNPs that are associated with neurodegenerative diseases, the odds ratios associated with these risk alleles are generally low and account for only a small proportion of estimated heritability. Given these results, it is expected that low-frequency variants, in particular functional variants in coding sequences, with relatively large effect sizes still remain to be identified [22]. For instance, GWASs have identified risk variants at over two dozen loci affecting the development of PD; however, only 3–5% of phenotypic variance associated with PD can be explained using all SNPs within a region identified by replicated GWASs [23]. These estimates are substantially smaller than those obtained in epidemiological studies [24,25], pointing to the compelling evidence of yet-to-be-discovered additional genetic factors. In this review, we present brief overviews of genetic research on sporadic forms of neurodegenerative diseases, including AD, PD, and MSA as examples.

2. Alzheimer disease

AD is the most common neurodegenerative disease characterized by progressive cognitive impairment, including gradual loss of memory, judgment, and ability to maintain social or occupational functioning. AD usually develops in people over 65 years of age, but the less-prevalent early-onset AD can occur much earlier in adulthood. From the viewpoint of genetics, AD occurs under two conditions: a familial AD, which is determined by mutations in a single gene, and sporadic AD. Familial AD is rare and the majority of AD cases are sporadic AD. Regarding familial AD, to date, three causative genes have been identified in autosomal dominantly inherited cases of AD, *APP* [26,27], *PSEN1* [28], and *PSEN2* [29]. Among these causative genes, mutations in *PSEN1* are the most common causes of early-onset familial AD, accounting for 18–50% of autosomal dominantly inherited cases of AD [30]. On the other hand, the molecular bases of sporadic AD are not fully elucidated, which could be caused by genetic susceptibilities, environmental exposures, and gene–environment interactions. Familial aggregation in late-onset AD has been well recognized. For example, the lifetime risk estimates for the ages of 88–90 years were 23.4–25.9% in relatives of AD patients and 19.1% in relatives of healthy controls [31,32]. A study of monozygotic and dizygotic twins without dementia, who were followed up for an average of 5 years, showed that 5.8% of subjects were newly diagnosed as having AD during the follow-up period, and of the pairs in which at least one twin developed AD, the concordance rate was 32.2% for monozygotic pairs and 8.7% for dizygotic pairs [33]. Until recently, however, the only proven genetic risk factor for late-onset AD in various ethnic groups has been the E4 allele of *APOE*, which is mapped to chromosome 19q. A meta-analysis showed that the odds ratio for E4 allele frequency in the AD and healthy control groups was 3.98 (95% confidence interval: 3.44–4.61) [34]. Originally, *APOE* conferring strong susceptibility to late-onset AD was identified by the linkage study of several late-onset AD families showing familial aggregation [35]. Because the mode of inheritance could not be determined with certainty, the affected-pedigree-member method of linkage analysis was selected [35]. Following the nonparametric linkage studies, association analysis first confirmed the role of the *APOE* E4 allele as a strong genetic risk factor for late-onset AD [36,37], as well as early-onset AD [38].

After the discovery of *APOE*, many studies highlighted the potential role of other genes in late-onset AD, and eventually a recent meta-analysis of 74,046 individuals by the International Genomics of Alzheimer's Project (IGAP) confirmed 20 AD susceptibility loci, including *CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/HLA-DRB1*, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4/RIN3*, *SORL1*, *ZCWPW1*, *CR1*, *BIN1*, *CD2AP*, *EPHA1*, *CLU*, *MS4A6A*, *PICALM*, *ABCA7*, and *APOE* loci [39]. Despite the identification of an increasing number of AD susceptibility variants that are common in the general population, these variants confer a mere 0.10-fold to 0.15-fold increase or decrease in AD risk in carriers versus noncarriers of the risk alleles except for *APOE* [40]. In contrast to these finding, there could be a substantial number of rare variants with relatively large effect sizes because they were hardly detected by GWASs. For example, whole-exome sequencing analysis of multiplex families with AD and subsequent analyses of the candidate variants in large case-control data sets have recently revealed that a rare variant (an allele frequency of less than 1%), V232M in *PLD3*, segregated with AD in two unrelated families and that V232M carriers have a four-fold increased risk of developing AD compared with individuals aged over 70 years without dementia [41].

3. Parkinson disease

PD, characterized by tremor, rigidity, bradykinesia, and postural instability, is the second most common neurodegenerative disease after AD, with a median age at onset of 60 and a risk that increases with age. The prevalence of PD is estimated to be 1% in individuals older than 60 years [42]. Mendelian forms of PD show both the autosomal dominant and recessive modes of inheritance. The autosomal dominant causative genes are *SNCA* [43,44], *LRRK2* [45,46], *EIF4G1* [47], and *VPS35* [48]. The first identified causative gene *SNCA*, which encodes α-synuclein, is the principal structural component of Lewy bodies. Notably, an increased gene dosage of *SNCA* causes an autosomal dominant form of PD [44]. These findings suggest that aberrant deposition of α-synuclein in dopaminergic neurons predominantly in the substantia nigra of midbrain is the primary driving force in PD pathogenesis. The autosomal recessive PD genes are *PARK2* [49], *PINK1* [50], *DJ1* [51], *PLA2G6* [52], and *FBXO7* [53]. Recent evidence suggests that the autosomal recessive PD genes *PARK2* and *PINK1* play roles in the clearance of damaged mitochondria by autophagy [54]. Recessively transmitted

Please cite this article in press as: J. Mitsui, S. Tsuji, Genomic aspects of sporadic neurodegenerative diseases, Biochem. Biophys. Res. Commun. (2014), http://dx.doi.org/10.1016/j.bbrc.2014.07.098