



# Polygenic Risk Scores in Neurodegenerative Diseases: a Review

Laura Ibanez<sup>1</sup> · Fabiana H. G. Farias<sup>1</sup> · UMBER DUBE<sup>1</sup> · Kathie A. Mihindukulasuriya<sup>1</sup> · Oscar Harari<sup>1</sup>

Published online: 17 January 2019

© The Author(s) 2019

## Abstract

**Purpose of the Review** This review summarizes the current state of the art of polygenic risk scores (PRSs) in the assessment of risk for neurodegenerative diseases.

**Recent Findings** Polygenic risk scores have been used to identify the shared genetic architecture between comorbid complex traits, disease presentations, and disease endophenotypes.

**Summary** The pathological and symptomatologic overlap between neurodegenerative diseases is strikingly high. In some cases, the diagnostic decision is arbitrary depending on the first appearance of symptomatology. Genetic studies have demonstrated that the genetic architecture of each of these diseases is different, but has a high degree of overlap. The creation of polygenic risk scores has allowed a more accurate calculation of this overlap. However, the power of the PRS is dependent on the power of the genome-wide association studies (GWAS) used to describe the genetic architecture. Even though not all neurodegenerative disease GWAS have the same sample size, and thus the same power, the use of polygenic risk scores has been successful in demonstrating the genetic overlap that has been observed phenotypically.

**Keywords** PRS · Polygenic risk score · Neurodegeneration · Genetic architecture

## Introduction

Neurodegeneration is defined as progressive neuronal vulnerability and loss of function. To date, this progression is unstoppable and the diseases are incurable. Neurodegenerative diseases are proteinopathies; there are the accumulation and aggregation of a pathogenic protein in the brain. The pathogenesis of neurodegenerative diseases (ND) is characterized by the aggregation of specific proteins in intracellular inclusions or extracellular aggregates. The hallmark protein tends to be different for each disease [1] but the overlap is striking (Table 1). Amyloid-beta plaques (A $\beta$ ) and tau tangles characterize Alzheimer's disease (AD) affected brains, and tau, but not A $\beta$ , is also aggregated in frontotemporal dementia (FTD). Alpha-synuclein is aggregated in Parkinson's disease (PD) and dementia with Lewy bodies (DLB). The proteins FUS,

TDP-43, and SOD-1 have been found aggregated in amyotrophic lateral sclerosis (ALS), but TDP-43 has been found also in FTD brains. Amyloid-beta plaques have been found in PD brains, and alpha-synuclein aggregates are also present in several AD brains [2]. In fact, abnormal cortical A $\beta$  deposition is present in PD patients with dementia [3, 4].

A combination of multiple genetic, lifestyle, and environmental factors modulate the risk of ND. A very low percentage of cases show Mendelian inheritance patterns, while the majority of the cases have complex genetic architectures that define the ND genetic predisposition. This genetic burden interacts with lifestyle and environmental factors to predispose patients to ND.

AD is the most common neurodegenerative disease. Aside from the accumulation of amyloid beta plaques and neurofibrillary tangles [5], it is also characterized by the degeneration of the subcortical hippocampal regions and the medial temporal lobe, which are associated with memory impairment [6–8]. Even though AD and FTD are both characterized by tau aggregates, in FTD, the degeneration happens in the frontal and anterior temporal lobes, rather than the hippocampus and medial temporal lobe [9]. In fact, FTD and AD are sometimes difficult to distinguish at onset and even during disease progression [10••]. FTD also overlaps genetically, pathologically, and neuropsychologically with ALS. ALS is characterized by

This article is part of the Topical Collection on *Neurogenetics and Psychiatric Genetics*

✉ Oscar Harari  
harario@wustl.edu

<sup>1</sup> Department of Psychiatry, School of Medicine, Washington University in Saint Louis, 660 South Euclid Avenue, Campus Box 8134, St. Louis, MO 63110, USA

**Table 1** Protein aggregates found in neurodegenerative diseases

Disease	Neuropathological hallmark	Additional aggregates
Alzheimer's disease	Amyloid beta Tau	Alpha-synuclein
Parkinson's disease	Alpha-synuclein	Amyloid beta Tau
Dementia with Lewy body Frontotemporal dementia	Alpha-synuclein Tau TDP-43 FUS	Amyloid beta
Amyotrophic lateral sclerosis	SOD-1 TDP-43 FUS	

the loss of motor neurons in the brain and spinal cord that consequently results in muscle wasting, spasticity, and death, usually within three years [8, 11]. PD is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta and the presence of Lewy bodies in the surviving neurons. The presence of Lewy bodies, or aggregates of alpha-synuclein, is a characteristic shared with DLB, but the diagnosis of either PD or DLB depends on the order of symptom manifestation. If motor symptoms are the first to manifest, followed by dementia symptoms within a year of PD diagnosis, the individual will be diagnosed with PD with Parkinson's disease dementia (PDD). If dementia manifests first, the diagnosis will be DLB, even though the person could develop motor symptoms later on in the disease. This is an arbitrary definition that was reached by consensus due to the overlap of these two diseases [12, 13].

### Shared Genetic Architecture among Neurodegenerative Diseases

Early genetic studies were focused on the identification of variants within the coding regions of proteins associated with each neurodegenerative trait. These studies allowed the identification of variants in genes such as *apolipoprotein E* (*APOE*) for AD and *alpha-synuclein* (*SNCA*) for PD among others. The study of families with extreme phenotypic characteristics, such as early age at onset for AD, allowed the identification of additional variants with Mendelian inheritance patterns. However, these fully penetrant mutations in general are present in a low percentage of ND cases.

Genome-wide association studies (GWAS) enabled the systematic screening of the genome. The analyses and meta-analyses of large cohorts identified additional variants with smaller effects on risk that were more common in the population. Currently, the largest GWAS is the meta-analysis of PD that includes a very large number of participants (Table 2)

**Table 2** Largest GWAS for neurodegenerative diseases

Disease	Sample size	Year	Number of GWAS hits
Alzheimer's disease [14]	74,046	2013	22
Parkinson's disease [15]	416,518	2017	44
Frontotemporal dementia [16]	12,928	2014	10
Dementia with Lewy body [17]	6197	2018	5
Amyotrophic lateral sclerosis [18]	41,398	2016	4

from the 23andMe PD cohort [15]. This study detected 44 loci, of which 17 are new findings missed in previous meta-analysis [19]. The International Genomics of Alzheimer's Project (IGAP) is the latest and largest GWS meta-analysis published for AD (Table 2). This effort analyzed 74,046 individuals and identified 22 genetic loci. The estimated proportion of variation tagged by all SNPs is 0.24, while the genetic heritability of AD is 0.74 [20, 21]. The largest GWAS for FTD is a two-stage meta-analysis with a total number of 12,928 participants (Table 2) [16]. Interestingly, additional FTD subtype stratified analyses identified additional loci. This evidence supports that different FTD subtypes have a distinct genetic architecture. A two-stage meta-analysis for DLB analyzed 6197 individuals [17]. In accordance with the already known overlap with AD, the top GWAS hit was the *APOE* loci. The largest genetic analysis for ALS analyzed 41,398 individuals and reported four loci with genome-wide significant association [18]. GWAS chips captured 8.5% of the genetic heritability, while the total is estimated at 65%. Additional modeling of the data (linear mixed models) identified four additional loci. This may be indicating that the genetic architecture of ALS is extremely heterogeneous, and it might be more informative to subclassify it into subtypes, similar to what was done in the FTD GWAS.

Remarkably, these studies demonstrated what was already observed during pathological examination of ND brains. The extent of overlap among the genetic architecture shared among neurodegenerative diseases is surprisingly high. Genetic studies of late-onset AD identified the  $\epsilon 4$  allele of *APOE*, which increases the risk for AD (OR = 3.1 for heterozygotes, OR = 12 for homozygotes [22]), and it is present in approximately 15% of individuals of European ancestry. Even though it was described for the first time in AD, the  $\epsilon 4$  allele has been associated with DLB severity [23], ALS age at onset [24], and its association with PD has also been reported [25, 26].

Autosomal-dominant AD accounts for between 1 and 5% of total AD cases [27] and presents a dominant inheritance pattern with variants in *amyloid precursor protein* (*APP*), *presenilin 1* (*PSEN1*), and *presenilin 2* (*PSEN2*) [28]. Several rare variants in *PSEN1* have also been reported in PD patients [29, 30]. More recent genetic studies have

identified low frequency coding variants in *TREM2* associated with AD [31, 32]. Afterwards, variants in the same gene were identified in PD [33] and in FTD [34].

Mutations in *LRRK2*, a gene associated with PD [35], have also been found in two families with AD [36]. FTD has been linked to mutations in the *MAPT* and *GRN* genes [37] which have also been found to be involved in AD and PD [38, 39••]. Finally, mutations in *C9ORF72*, *SOD1*, *FUS*, and others are associated with both FTD and ALS [16, 18, 40–43].

## Polygenic Risk Scores

Polygenic risk scores (PRSs) are simple models that have been instrumental to analyze genetic architecture and predict disease risk of complex traits, such as schizophrenia and bipolar disorder [44]. These scores aggregate genome-wide information to account for the phenotypic variation observed in complex traits, by assuming an additive, non-multiplicative, effect of multiple variants with variable effect sizes. This allows a more accurate assessment of an individual's risk for a disease, given their genetic background, than evaluating each genetic variation independently. PRS can highlight at-risk individuals for closer examination and allow for the application of early intervention strategies. In addition, PRS can be applied as inclusion criteria for targeted clinical trials. Furthermore, the genetic overlap between comorbid diseases, previously identified only by epidemiological or clinical studies, can be evaluated by PRS, to determine whether the pleiotropic effects of variants identified in one disease leads to increased risk for another disease. For example, major depressive disorder and current psychological distress positively moderate the effect of polygenic risk for obesity on body mass index [45]. Many examples of the use of PRS can be found in the literature, such as the identification of the genetic overlap between schizophrenia and cognitive ability [46], as well as, major depressive disorder and body mass index [45]. Our group has employed PRS to study the extent of the overlap of the genetic architecture among distinct clinical manifestations of Alzheimer's disease [47••].

## Polygenic Risk Score Calculation

Polygenic risk scores are an estimate of disease risk carried by the individual based on the risk alleles and the corresponding effect sizes obtained from the GWAS summary statistics. The GWAS summary statistics will be referred as the base and the dataset to be evaluated as the target. There are three important factors for PRS construction: the base and target must be from independent datasets; quality control has to be performed on both, the base and target GWAS; and the selection of the significance threshold has to be evaluated to optimize prediction power.

Several quality controls need to be applied to the base and target datasets. First, the genome build and affected alleles must be matched between the reference and target datasets. Ambiguous SNPs (A/T and C/G variants) should be removed from the datasets, since it is not possible to match them with certainty. Because GWAS are typically performed one SNP at a time, the identification of independent genetic signals is challenging. In consequence, it is necessary to control for the genetic architecture of the population defined by the linkage disequilibrium (LD). Clumping is the standard approach to deal with LD, as it selects the SNP with lowest  $p$  value retaining independent associations for an area of LD. This is preferable to pruning, which is a random process and thus results in a representative SNP that may explain less of the total variance of the region of LD than a clumped SNP [48]. It is also possible to calculate PRS using variants with association significance that pass alternative cutoff values. In this way, PRSs allow to evaluate not only the variants with genome-wide stringent  $p$  values, but also the suggestive ones, or even variants with marginal  $p$  values (for example threshold =  $(0.05 \times 10^{-5}, 1 \times 10^{-4}, \dots, 0.05, 0.1, \dots, 0.5)$ ) [49•]. Then, the association with additional traits is usually evaluated using logistic regressions. Additional covariates (for example sex and age) and confounding factors (such as principal components from population stratification) are also modeled. Given the multiple thresholds evaluated, it is usually considered significant statistical association when the  $p$  value  $< 1.00 \times 10^{-3}$  [49•].

The calculation of PRS can be performed using PLINK [50], PRSice [49•], lassosum [51•], LDpred [52•], or Multi-trait analysis of GWAS (MTAG) [53•]. Since the PLINK score function uses a linear scoring system for calculating PRS, all quality control needs to be performed on the input data before running the calculation. Unlike PLINK, PRSice performs clumping and removes ambiguous SNPs as a default. Also, it allows the selection of different  $p$  value thresholds giving the best fit scores for the data. The other three methods introduce additional methodological aspects to calculate the PRS. Lassosum uses penalized regression to correct for LD structure and adjust effect sizes, while LDpred assumes a prior for genetic architecture and LD information from a reference panel. In consequence, the choice of methodology to calculate PRS also has to take into consideration the characteristics of the data available for the base and target GWAS. Several manuscripts provide guidelines for selecting optimal approaches [52•, 54•].

Estimates of the shared genetic architecture among traits can be also calculated using summary statistics and not individual genotype data. PRSice provides a convenient model to test it based on the inverse-variance method that corresponds to the instrumental variable method that uses individual-level data for Mendelian randomization approaches [55]. In addition, there are additional approaches to estimate the extent of

overlap among traits. The method linkage disequilibrium (LD) score [56] regression estimates the genetic correlation of two traits, analyzing all SNPs in LD. It regresses the  $\chi^2$  statistic against the LD scores, which is estimated summarizing the LD  $r^2$  in a predefined region. The method called *coloc* tests [57] whether two association signals share a common causal variant. It employs Bayesian approaches, and it was conceived to integrate gene analysis for one disease with expression quantitative locus, but it can also be employed to analyze two trait-associated analyses. Another alternative method is *GNOVA* (genetic covariance analyzer), which allows the estimation of genetic covariance using the method of moments [58], while allowing to stratify the variants analyzed. Finally, MTAG analyzes summary statistics to estimate genetic correlations among traits using bivariate linkage disequilibrium score regression while correcting for the possibility of overlap between samples. The research questions to be answered, the data available, and the planned analyses should also guide the choice of methodology.

### Polygenic Risk Score in Neurodegenerative Diseases

Polygenic risk scores have been used in neurodegeneration for both testing the genetic overlap between characteristics of the same neurodegenerative disease, such as risk and age at onset, and testing the genetic overlap among neurodegenerative diseases.

#### Genetic Overlap within Characteristics of the Same Disease

**Alzheimer's Disease** Even though many methods have been used to calculate PRS for AD, the overall results are coincident among all of the studies. In general, the results from the GWAS in the International Genomics of Alzheimer's Project (IGAP) study have been used to model the PRS [14]. The first PRS was published in 2015. By adding the polygenic burden, they were able to predict disease development with an area under the ROC curve of 78% when age sex and *APOE* genotype were included [59]. In this study, a subset of the IGAP samples was used to investigate the prediction accuracy of models trained with the weights learned from the analysis of the entire IGAP cohort [59]. In another study, the analysis of a subset of the IGAP cohort with neuropathological data produced an increased area under the ROC curve (AUC = 84%) [60]. The authors conclude that most of the missing heritability and the moderate values under the ROC curve may be due to the diagnostic accuracy, and thus, for non-pathologically confirmed AD, there is room for improvement [60, 61••].

The PRS has also been used to detect individuals at greater risk for developing AD, and proved to be successful, even in those individuals that were noncarriers of the *APOE*  $\epsilon 4$  allele. The PRS predicted longitudinal clinical decline in older individuals that showed moderate to high depositions of amyloid

beta and or tau [62] and in clinically diagnosed AD individuals [63]. Moreover, the AD PRS was found to predict the level and rate of memory loss in a sample of non-Hispanic whites from the Health and Retirement Study [64] and in the ADNI (Alzheimer's Disease Neuroimaging Initiative) cohort [65]. In the same work, the authors evaluated the AD PRS in non-Hispanic black participants and found similar results but with a weaker association [64]. The authors argue that distinct factors can limit the power of the PRS in populations with different genetic backgrounds. The genotyping platforms employed for the discovery analyses were designed to capture variation among populations with European ancestry, and the SNPs included might not be as effective in tagging the significant loci in other populations. In addition, most of the GWAS for AD predominantly analyzed participants with European ancestry, and allele frequencies may vary between ethnicities and could alter the detectable effect sizes [64].

PRS has not only been utilized to improve AD diagnosis. It has also been used to test for genetic commonalities between characteristics of AD, like AD risk [47••, 66••, 67], age of disease onset [47••], and AD biomarkers [47••, 68]. PRS has been used to demonstrate that the genetic architecture of sporadic AD is shared with familial AD without Mendelian mutations [66••, 67]. Similarly, the extent of overlap of the known genetic architecture was compared between early (< 65 years at clinical manifestation of symptom) and late onset (> 65 years). The odd ratios between these strata are different (1.40 for sporadic late onset and 1.75 for familial late onset versus 2.27 for sporadic early onset) [47••]. In fact, the genetic factors included in the PRS seem to have additive effects on age at onset [47••]. Finally, PRS has been associated with CSF ptau181-A $\beta$ 42 ratio and CSF tau in autosomal dominant AD [47••]. In patients without clinical dementia, the predictive value of amyloidopathy and tauopathy seems to increase as a function of the PRS [68].

**Parkinson's Disease** The largest meta-analysis performed by Nalls et al. [19], prior the inclusion of the 23andMe data [15], has usually been employed as the reference to build PD-related PRS. These PRSs have been associated with the age at onset of PD [69••], faster motor and cognitive decline [70], and PD status [69••]. An initial study reported that only PRS that included the effect of SNPs with *p values* below nominal significance thresholds were significantly associated with PD in an additional independent PD dataset [71••], implying that the genetic architecture of PD includes many common variants with small effects. Further studies showed that PRS based on more significant (sentinel) SNPs are also associated with PD risk [69••]. In addition, PRSs were employed to show a higher genetic burden in early-onset PD compared to late onset (maximum OR of 4.8 and  $p < 0.001$  [71••]). Thus far, the PD PRS risk has not been successfully associated with CSF alpha-synuclein levels [69••].

**Dementia with Lewy Bodies, Frontotemporal Dementia and Amyotrophic Lateral Sclerosis** At the time of this review, no PRS had been attempted for these diseases. Several facts can explain the lack of a PRS in these ND. First of all, the sample sizes for the GWAS of DLB, FTD, and ALS are not as large as AD or PD (Table 2), not is the amount of explained heritability. In addition, both FTD and ALS are very heterogeneous and can be stratified in different subtypes, reducing the sample size for each group and thus the power of the GWAS and the PRS.

### Genetic Overlap among Neurodegenerative Diseases

PRS provides the mean to compare the genetic architecture of diseases that are suspected to have some genetic overlap. PRS is used to test if the diseases share genetic architecture. The PRS for AD has been found to be associated with amnesic and nonamnesic mild cognitive impairment, whereas the PRS for PD and FTD were only associated with nonamnesic, mild cognitive impairment. However, using these to predict future dementia was unsuccessful, probably due to the heterogeneity of the population with mild cognitive impairment [72]. A recent report [73] shows that even though the association of AD and PD PRSs with case-control status of DLB is highly significant, the amount of variance explained by both PRS is relatively small (for AD is = 1.33% and 0.14% considering or not the *APOE* locus, and for PD is = 0.37%). This adds evidence to the fact the DLB is an entity unto itself, with unique genetic risk factors, and not a mixture of AD and PD [73]. Similarly, PRSs have been used to test whether AD, ALS, and FTD are associated with cognitive function and physical health in healthy individuals [8] and to show that, while the three diseases showed an association with cognitive function, the risk for ALS was not associated with physical function.

PRSs have been used to investigate the extent of overlap of the genetic architecture between ALS and schizophrenia [44] and widely employed to determine the shared genetic risk between AD and PD risk and additional characteristics of the diseases, including age at onset, biomarker levels, and disease progression [8, 10•]. In addition, other methods have been employed to evaluate the genetic overlap among neurodegenerative diseases [10•] including fold-enrichment plots of the nominal  $-\log_{10}$  ( $p$  values) for all FTD-SNPs and a subset of SNPs determined by the significance of their association with PD and AD. The study concluded that there is a genetic pleiotropy between AD, PD and FTD.

### Pathway Specific PRSs

Additional research has been performed to interrogate the extent of overlap of pathway specific genetic architecture among traits, by aggregating biological knowledge into the

calculation of the polygenic risk to derive pathway specific risk. In these studies, the PRS is calculated for variants located in genes that are part of specific pathways, instead of considering the entire genome. For AD, PRSs summarizing the immune, amyloid beta clearance, and cholesterol pathways were created to predict AD-related biomarkers. In this case, the PRSs were poor predictors of cognitive function [74•]. This may be due in part to the incompleteness of pathways, such that variants in genes not known to be part of a pathway are excluded. Two studies have demonstrated the relationship between AD and inflammatory diseases using PRS [75, 76]. AD risk PRS was associated with increased levels of plasma inflammatory biomarkers, adding additional evidence to the involvement of inflammatory processes in AD [75]. While the global increase in inflammation seen in diseases such as multiple sclerosis does not influence age-related cognitive decline, variants that alter peripheral immunity influence microglial density and expression of immune genes in the aging brain. Thus, the influence of peripheral immune function on glial cell activation warrants further study [76].

### Future Directions

The research community has invested a considerable amount of effort generating GWAS data for ND, which has proven instrumental to the discovery of novel variants and genes associated with several disease traits. An ongoing challenge in the field of genetics is determining the relative roles of common variants with small effects and rare variants with large effects on disease phenotypes. Omnigenic model is a promising technique that integrates the effect of both common and rare variants along additional gene regulatory networks to analyze the genetic architecture of complex traits [77]. The power of this approach is constrained by the number of subjects analyzed in whole genome sequencing projects, which thus far is lower than the number of subjects included in GWAS. In addition, it was proposed that participant stratification is a promising strategy for genetic studies [78]. Current analyses of FTD and ALS data support this approach.

PRS combined with pathway analysis are enabling researchers to determine which model better fits complex diseases such as neurodegeneration and can help to determine whether therapeutic and preventative measures will be best targeted to specific genes or more broadly to pathways. Importantly, genetic studies have demonstrated the high complexity of neurodegenerative traits, whose risk is modulated by a large number of variants, with either small effect or very low frequency in the population. Thus, genetic studies need to include a very large number of subjects to pass stringent genome-wide thresholds. However, the sample size collected and meta-analysis varies considerably among diseases (Table 2). The same concept can be applied to the efforts

invested in constructing PRS in the different diseases. Querying PubMed for publications that employed PRS for ND provides a snapshot of the quantitative effort thus far invested for the different traits. A total of 43 manuscripts were retrieved while querying the manuscripts for AD “(PRS or Polygenic Risk Score) AND (Alzheimer),” but for FTD, the search produced only two manuscripts. Furthermore, a search of very active and highly invested research areas, other than ND, shows that PRSs are widely used, as PubMed included 218 publications for heart disease, 141 for cardiovascular disease, and 1127 for cancer.

## Conclusions

Genetic studies have provided valuable insights and novel understanding of ND. Increasing the sample size of the neurodegenerative disease cohorts and performing the meta-analyses on larger studies for FTD, DLB, and ALS will be critical to decipher the genetic structure of each disease and to investigate the genetic architecture shared by these neurodegenerative diseases with overlapping symptomology. In fact, this may also aid in the creation of a clinically useful PRS for neurodegeneration that allows the detection of individuals at risk, so they can enroll in clinical trials of neurodegenerative therapeutics. It is also plausible that a neurodegeneration PRS can be further optimized or combined with other phenotypic or molecular characteristics to specifically predict AD, FTD, DLB, and ALS.

One limitation of PRS is that they can only provide a maximum accuracy [61••], which is bounded to disease prevalence, which varies with the age and heritability. For example, GWAS chips only capture approximately a third of the estimated genetic heritability of AD (0.24 vs 0.76 [20, 21]), and prevalence for AD varies greatly with age (from 3% in the 65–74 age range to > 30% for those older than 85 years). Thus, the genetic risk captured by PRSs, which is lifelong constant, should be combined with additional biomarkers and clinical and environmental data to select at-risk individuals for therapeutic interventions and to produce better diagnostic tools.

To date, no PRS for ND that combines both common and rare variants has been calculated, which may lead to an increase of their accuracy. Novel machine learning methods are being developed to capture compact representations of GWAS [79] and coupled with powerful classification methods, namely deep neural networks, to produce highly accurate predictions [80–82]. The parallel development of larger cohorts, molecular phenotyping and the advancement of novel and more sophisticated methods to represent and operate multi-variant models, will allow a more precise discernment of the overlap of the genetic architecture among ND and predict individuals at risk with the accuracy required in clinical settings.

**Funding Information** This work was supported by grants from the National Institutes of Health (R01AG044546, P01AG003991, RF1AG053303, R01AG058501, U01AG058922, RF1AG058501 and R01AG057777), the Alzheimer Association (NIRG-11-200110, BAND-14-338165, AARG-16-441560 and BFG-15-362540), and the McDonnell Center for Cellular and Molecular Neurobiology.

## Compliance with Ethical Standards

**Conflict of Interest** Laura Ibanez, Fabiana H.G. Farias, Umer Dube, Kathie A. Mihindukulasuriya, and Oscar Harari declare that they have no conflict of interest.

**Human and Animal Rights and Informed Consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Katsnelson A, De Strooper B, Zoghbi HY. Neurodegeneration: from cellular concepts to clinical applications. *Sci Transl Med*. 2016;8(364):364ps18.
2. Meeus B, Theuns J, Van Broeckhoven C. The genetics of dementia with Lewy bodies: what are we missing? *Arch Neurol*. 2012;69(9):1113–8.
3. Meeus B, et al. DLB and PDD: a role for mutations in dementia and Parkinson disease genes? *Neurobiol Aging*. 2012;33(3):629 e5–629 e18.
4. Kotzbauer PT, Cairns NJ, Campbell MC, Willis AW, Racette BA, Tabbar SD, et al. Pathologic accumulation of alpha-synuclein and Abeta in Parkinson disease patients with dementia. *Arch Neurol*. 2012;69(10):1326–31.
5. Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*. 2002;297(5580):353–6.
6. Jack CR Jr, et al. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology*. 2000;55(4):484–9.
7. Jack CR Jr, Lowe VJ, Weigand SD, Wiste HJ, Senjem ML, Knopman DS, et al. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. *Brain*. 2009;132(Pt 5):1355–65.
8. Hagenaars SP, Radaković R, Crockford C, Fawns-Ritchie C, International FTD-Genomics Consortium (IFGC), Harris SE, et al.

- Genetic risk for neurodegenerative disorders, and its overlap with cognitive ability and physical function. *PLoS One*. 2018;13(6):e0198187.
9. Vieira RT, Caixeta L, Machado S, Silva AC, Nardi AE, Arias-Carrión O, et al. Epidemiology of early-onset dementia: a review of the literature. *Clin Pract Epidemiol Ment Health*. 2013;9:88–95.
  10. Ferrari R, et al. Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *J Neurol Neurosurg Psychiatry*. 2017;88(2):152–64 **Unique comparison between AD and FTD using PRSs to date.**
  11. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat Rev Neurol*. 2013;9(11):617–28.
  12. Gomperts SN. Lewy body dementias: dementia with Lewy bodies and Parkinson disease dementia. *Continuum (Minneapolis Minn)*. 2016;22(2 Dementia):435–63.
  13. Irwin DJ, Lee VM, Trojanowski JQ. Parkinson's disease dementia: convergence of alpha-synuclein, tau and amyloid-beta pathologies. *Nat Rev Neurosci*. 2013;14(9):626–36.
  14. Lambert JC, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452–8.
  15. Chang D, Nalls MA, Hallgrímsson IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet*. 2017;49(10):1511–6.
  16. Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JBJ, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol*. 2014;13(7):686–99.
  17. Guerreiro R, Ross OA, Kun-Rodriguez C, Hernandez DG, Orme T, Eicher JD, et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol*. 2018;17(1):64–74.
  18. van Rheenen W, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet*. 2016;48(9):1043–8.
  19. Nalls MA, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46(9):989–93.
  20. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1088–93.
  21. Lee SH, Harold D, Nyholt DR, ANZGene Consortium, International Endogene Consortium, the Genetic and Environmental Risk for Alzheimer's disease (GERAD1) Consortium, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet*. 2013;22(4):832–41.
  22. Kukull WA, Schellenberg GD, Bowen JD, McCormick WC, Yu CE, Teri L, et al. Apolipoprotein E in Alzheimer's disease risk and case detection: a case-control study. *J Clin Epidemiol*. 1996;49(10):1143–8.
  23. Dickson DW, Heckman MG, Murray ME, Soto AI, Walton RL, Diehl NN, et al. APOE epsilon4 is associated with severity of Lewy body pathology independent of Alzheimer pathology. *Neurology*. 2018;91(12):e1182–95.
  24. Li YJ, Pericak-Vance MA, Haines JL, Siddique N, McKenna-Yasek D, Hung WY, et al. Apolipoprotein E is associated with age at onset of amyotrophic lateral sclerosis. *Neurogenetics*. 2004;5(4):209–13.
  25. Federoff M, Jimenez-Rolando B, Nalls MA, Singleton AB. A large study reveals no association between APOE and Parkinson's disease. *Neurobiol Dis*. 2012;46(2):389–92.
  26. Gao J, et al. Apolipoprotein E genotypes and the risk of Parkinson disease. *Neurobiol Aging*. 2011;32(11):2106 e1–6.
  27. Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum Genet*. 1999;65(3):664–70.
  28. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat*. 2012;33(9):1340–4.
  29. Jimenez-Escrig A, Rabano A, Guerrero C, Simon J, Barquero MS, Guell I, et al. New V272A presenilin 1 mutation with very early onset subcortical dementia and parkinsonism. *Eur J Neurol*. 2004;11(10):663–9.
  30. Takao M, Ghetti B, Hayakawa I, Ikeda E, Fukuuchi Y, Miravalle L, et al. A novel mutation (G217D) in the presenilin 1 gene (*PSEN1*) in a Japanese family: presenile dementia and parkinsonism are associated with cotton wool plaques in the cortex and striatum. *Acta Neuropathol*. 2002;104(2):155–70.
  31. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogava E, Majounie E, et al. *TREM2* variants in Alzheimer's disease. *N Engl J Med*. 2013;368(2):117–27.
  32. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of *TREM2* associated with the risk of Alzheimer's disease. *N Engl J Med*. 2013;368(2):107–16.
  33. Benitez BA, Cruchaga C, G. United States-Spain. Parkinson's disease research, *TREM2 and neurodegenerative disease*. *N Engl J Med*. 2013;369(16):1567–8.
  34. Jay TR, von Saucken VE, Landreth GE. *TREM2* in neurodegenerative diseases. *Mol Neurodegener*. 2017;12(1):56.
  35. Paisan-Ruiz C, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron*. 2004;44(4):595–600.
  36. Fernandez MV, et al. Analysis of neurodegenerative Mendelian genes in clinically diagnosed Alzheimer disease. *PLoS Genet*. 2017;13(11):e1007045.
  37. Tang SS, Li J, Tan L, Yu JT. Genetics of frontotemporal lobar degeneration: from the bench to the clinic. *J Alzheimers Dis*. 2016;52(4):1157–76.
  38. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet*. 2009;41(12):1303–7.
  39. Desikan RS, et al. Genetic overlap between Alzheimer's disease and Parkinson's disease at the *MAPT* locus. *Mol Psychiatry*. 2015;20(12):1588–95 **FTD main gene(MAPT) locus comparison in AD and PD.**
  40. van Blitterswijk M, DeJesus-Hernandez M, Rademakers R. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? *Curr Opin Neurol*. 2012;25(6):689–700.
  41. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245–56.
  42. Majounie E, Renton AE, Mok K, Dopper EGP, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol*. 2012;11(4):323–30.
  43. Ferrari R, Kapogiannis D, Huey E, Momeni P. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res*. 2011;8(3):273–94.
  44. International Schizophrenia C, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748–52.
  45. Clarke TK, Hall LS, Fernandez-Pujals AM, MacIntyre DJ, Thomson P, Hayward C, et al. Major depressive disorder and current psychological distress moderate the effect of polygenic risk for obesity on body mass index. *Transl Psychiatry*. 2015;5:e592.
  46. Lencz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM, et al. Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the cognitive

- genomics consortium (COGENT). *Mol Psychiatry*. 2014;19(2):168–74.
- 47.●● Cruchaga, C., et al. Polygenic risk score of sporadic late-onset Alzheimer's disease reveals a shared architecture with the familial and early-onset forms. *Alzheimers Dement*, 2018;14(2):205–14. **Screening for genetic overlap of AD risk and AD characteristics.**
  48. Prive F, et al. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*. 2018;34(16):2781–7.
  - 49.● Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics*. 2015;31(9):1466–8 **Main methodology paper for PRS calculation softwares.**
  50. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
  - 51.● Mak TSH, et al. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol*. 2017;41(6):469–80 **PRS methodology: Penalized regression to adjust for LD structure.**
  - 52.● Vilhjalmsson BJ, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97(4):576–92 **PRS methodology: Use reference panel to correct for LD structure.**
  - 53.● Turley P, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50(2):229–37 **PRS methodology: Bivariate linkage disequilibrium score regression.**
  - 54.● Bogdan R, Baranger DAA, Agrawal A. Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences. *Annu Rev Clin Psychol*. 2018;14:119–57 **PRS methodology: Guidelines and technical review.**
  55. Johnson, T. Efficient calculation for multi-SNP genetic risk scores. American Society of Human Genetics Annual Meeting, San Francisco, November 6–10, 2012, <https://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf>.
  56. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat Genet*. 2018;50(6):857–64.
  57. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
  58. Lu Q, Li B, Ou D, Erlendsdottir M, Powles RL, Jiang T, et al. A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. *Am J Hum Genet*. 2017;101(6):939–64.
  59. Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain*. 2015;138(Pt 12):3673–84.
  60. Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Ann Neurol*. 2017;82(2):311–4.
  - 61.●● Escott-Price V, et al. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging*. 2017;49:214 e7–214 e11 **Analysis of maximum theoretical accuracy for AD-PRS.**
  62. Tan CH, Hyman BT, Tan JJX, Hess CP, Dillon WP, Schellenberg GD, et al. Polygenic hazard scores in preclinical Alzheimer disease. *Ann Neurol*. 2017;82(3):484–8.
  63. Del-Aguila JL, et al. Assessment of the genetic architecture of Alzheimer's disease risk in rate of memory decline. *J Alzheimers Dis*. 2018;62(2):745–56.
  64. Marden JR, Mayeda ER, Walter S, Vivot A, Tchetsgen EJ, Kawachi I, et al. Using an Alzheimer disease polygenic risk score to predict memory decline in black and white Americans over 14 years of follow-up. *Alzheimer Dis Assoc Disord*. 2016;30(3):195–202.
  65. Kauppi K, Fan CC, McEvoy LK, Holland D, Tan CH, Chen CH, et al. Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer's disease. *Front Neurosci*. 2018;12:260.
  - 66.●● Tosto G, et al. Polygenic risk scores in familial Alzheimer disease. *Neurology*. 2017;88(12):1180–6 **Overlap between sporadic and familial AD genetic architecture.**
  67. Chaudhury S, et al. Polygenic risk score in postmortem diagnosed sporadic early-onset Alzheimer's disease. *Neurobiol Aging*. 2018;62:244 e1–8.
  68. Tan CH, et al. Polygenic hazard score: an enrichment marker for Alzheimer's associated amyloid and tau deposition. *Acta Neuropathol*. 2018;135(1):85–93.
  - 69.●● Ibanez L, et al. Parkinson disease polygenic risk score is associated with Parkinson disease status and age at onset but not with alpha-synuclein cerebrospinal fluid levels. *BMC Neurol*. 2017;17(1):198 **Screening for genetic overlap of PD risk and PD characteristics.**
  70. Paul KC, et al. Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol*. 2018;75(3):360–66.
  - 71.●● Escott-Price V, et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol*. 2015;77(4):582–91 **PRS was used to show a larger genetic burden in early onset PD cases compared to late onset ones.**
  72. Adams HH, et al. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement*. 2015;11(11):1277–85.
  73. Guerreiro R, et al. Heritability and genetic variance of dementia with Lewy bodies. 2018 <https://www.biorxiv.org/content/early/2018/10/26/454249>. Accessed 11 Nov 2018.
  - 74.●● Darst BF, et al. Pathway-specific polygenic risk scores as predictors of amyloid-beta deposition and cognitive function in a sample at increased risk for Alzheimer's disease. *J Alzheimers Dis*. 2017;55(2):473–84 **Using a combined PRS designed using biological hypothesis to predict AD status.**
  75. Morgan AR, Touchard S, O'Hagan C, Sims R, Majounie E, Escott-Price V, et al. The correlation between inflammatory biomarkers and polygenic risk score in Alzheimer's disease. *J Alzheimers Dis*. 2017;56(1):25–36.
  76. Felsky D, Patrick E, Schneider JA, Mostafavi S, Gaiteri C, Patsopoulos N, et al. Polygenic analysis of inflammatory disease variants and effects on microglia in the aging brain. *Mol Neurodegener*. 2018;13(1):38.
  77. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177–86.
  78. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. *Cell*. 2018;173(7):1573–80.
  79. Fergus P, Montanez A, Abdulaimma B, Lisboa P, Chalmers C, Pineles B, Utilising deep learning and genome wide association studies for epistatic-driven preterm birth classification in African-American women. *IEEE/ACM Trans Comput Biol Bioinform*, 2018. <https://doi.org/10.1109/TCBB.2018.2868667>.
  80. Mieth B, Kloft M, Rodríguez JA, Sonnenburg S, Vobruba R, Morcillo-Suárez C, et al. Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies. *Sci Rep*. 2016;6:36671.
  81. Lello L, Avery SG, Tellier L, Vazquez AI, de los Campos G, Hsu SDH. Accurate genomic prediction of human height. *Genetics*. 2018;210(2):477–97.
  82. Curbelo CAFP, Curbelo A; Chalmers C. Deep learning classification of polygenic obesity using genome wide association study SNPs. 2018. <https://arxiv.org/ftp/arxiv/papers/1804/1804.03198.pdf>. Accessed 18 Dec 2018.