



ARTICLE

Polygenic risk score analysis for amyotrophic lateral sclerosis leveraging cognitive performance, educational attainment and schizophrenia

Restuadi Restuadi¹ · Fleur C. Garton¹ · Beben Benyamin^{1,2} · Tian Lin¹ · Kelly L. Williams³ · Anna Vinkhuyzen¹ · Wouter van Rheenen⁴ · Zhihong Zhu¹ · Nigel G. Laing^{5,6} · Karen A. Mather^{7,8} · Perminder S. Sachdev^{7,9} · Shyuan T. Ngo^{10,11,12} · Frederik J. Steyn^{12,13} · Leanne Wallace¹ · Anjali K. Henders¹ · Peter M. Visscher^{1,10} · Merrilee Needham^{14,15,16} · Susan Mathers¹⁷ · Garth Nicholson^{3,18} · Dominic B. Rowe³ · Robert D. Henderson^{10,11,19} · Pamela A. McCombe^{12,19} · Roger Pamphlett²⁰ · Ian P. Blair³ · Naomi R. Wray^{1,10} · Allan F. McRae¹

Received: 17 September 2019 / Revised: 24 March 2021 / Accepted: 1 April 2021 / Published online: 27 April 2021
© The Author(s), under exclusive licence to European Society of Human Genetics 2021

Abstract

Amyotrophic Lateral Sclerosis (ALS) is recognised to be a complex neurodegenerative disease involving both genetic and non-genetic risk factors. The underlying causes and risk factors for the majority of cases remain unknown; however, ever-larger genetic data studies and methodologies promise an enhanced understanding. Recent analyses using published summary statistics from the largest ALS genome-wide association study (GWAS) (20,806 ALS cases and 59,804 healthy controls) identified that schizophrenia (SCZ), cognitive performance (CP) and educational attainment (EA) related traits were genetically correlated with ALS. To provide additional evidence for these correlations, we built single and multi-trait genetic predictors using GWAS summary statistics for ALS and these traits, (SCZ, CP, EA) in an independent Australian cohort (846 ALS cases and 665 healthy controls). We compared methods for generating the risk predictors and found that the combination of traits improved the prediction (Nagelkerke- R^2) of the case–control logistic regression. The combination of ALS, SCZ, CP, and EA, using the SBayesR predictor method gave the highest prediction (Nagelkerke- R^2) of 0.027 (P value = 4.6×10^{-8}), with the odds-ratio for estimated disease risk between the highest and lowest deciles of individuals being 3.15 (95% CI 1.96–5.05). These results support the genetic correlation between ALS, SCZ, CP and EA providing a better understanding of the complexity of ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease, with death typically occurring within 3–5 years of symptom onset [1, 2]. Currently, there is no effective treatment for ALS, likely in part due to a limited understanding of its underlying causes [1–3]. Approximately 10% of ALS cases are considered to have a Mendelian form of the disease [1, 3] carrying a highly penetrant

variant, however the remaining and vast majority of cases, likely have a more complex aetiology reflecting both genetic and environmental susceptibility risk factors [4].

Cognitive impairment affects $\approx 30\%$ of ALS patients, with $\sim 10\%$ diagnosed with frontotemporal dementia (FTD) [5, 6]. However, most studies consider cognitive impairment as a consequence of the ALS-FTD spectrum, rather than discussing attenuated cognitive performance as a risk factor of ALS [7–9]. Recent studies that link ALS and cognitive performance remain inconclusive. For example, lower educational attainment was reported to increase ALS-FTD risk in Italian and the United States (US) ALS cohorts [10, 11]. However, a larger cohort from the Swedish Conscript Register reported the opposite result; higher IQ in young conscripts was associated with higher ALS risk in later life [5]. Traditional prospective studies designed to investigate if cognitive decline is a risk factor that precedes

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-021-00885-y>.

✉ Allan F. McRae
a.mcrae@uq.edu.au

Extended author information available on the last page of the article

the diagnosis of ALS are hard to establish since ALS is a relatively low-frequency disorder with a lifetime risk of ~0.3% (ref. [1, 2]). Demonstrating a genetic relationship between ALS and cognitive traits would provide more conclusive evidence for the direction of association between them. Cognitive related traits have high heritability estimates (~60%) [12, 13] and are frequently measured in large population or community samples, particularly using proxies such as educational attainment. In contrast, while ALS has a substantial genetic component (heritability of 40–45%) [14], it is a late onset disease of relatively low frequency, making the collection of large cohorts difficult.

Advances in genotyping technology allow estimation of the genetic contribution to traits associated with SNPs measured genome-wide. These estimates of so-called SNP-based heritability use genome-wide association study (GWAS) data to capture the contribution from common genetic variants and so are smaller than heritability estimates from family studies. The meta-analysis of the ALS published GWAS study reported a SNP heritability estimate of ~8% [15]. Moreover, we can use GWAS summary statistics to estimate the genetic correlation (r_g) between traits using independently collected samples; for other diseases and traits correlation estimates made from genome-wide SNPs have been found to be similar to estimates from traditional epidemiology [16]. Using these approaches ALS has been reported to be negatively correlated with fluid intelligence ($r_g = -0.34$) and academic or professional qualifications ($r_g = -0.25$) [17], and positively correlated with schizophrenia ($r_g = 0.14$) [18]. Combining information from genetically correlated traits can improve genetic predictors of disease risk [19, 20], particularly for diseases such as ALS where cohort sizes are relatively small [21].

Here, we investigate the genetic relationship between ALS and over 700 traits using linkage disequilibrium score regression (LDSR) as implemented in LDhub ver1.90 platform [22], confirming previously identified correlations with ALS. We provide independent evidence for a genetic relationship between ALS and these traits using out-of-sample polygenic risk prediction into an independent data set of 846 ALS cases and 665 controls to demonstrate improvements in genetic prediction when combining multiple traits.

Materials and methods

Australian ALS GWAS data

We present new data from an Australian ALS GWAS cohort comprising 836 cases and 665 controls, and independent of all published ALS GWAS data. The sample includes the University of Sydney's Australian Motor Neuron Disease DNA Bank (MND Bank) cohort recruited

between April 2000 and June 2011 (462 cases, 449 controls), with study protocol approved by the Sydney South West Area Health Service Human Research Ethics Committee (HREC). The remainder of the cases ($N = 374$) comprised ALS patients recruited from clinics across Australia between 2015 and 2017 under HREC approvals from University of Sydney, Western Sydney Local Health District, Royal Brisbane and Women's Hospital and Macquarie University. The ALS cases were diagnosed with definite or probable ALS according to the revised El Escorial criteria [23]. Those with a recorded family history of ALS or had tested positive for genetic variants with strong support for ALS causality were excluded. Some controls ($N = 127$) were recruited as either partners or friends of patients, healthy individuals free of neuromuscular diseases. Additional controls were included from the Older Australian Twin Study (OATS) [24] comprising 89 monozygotic (MZ) twin pairs from QIMR Berghofer Medical Research Institute, University of New South Wales and the University of Melbourne, and was approved by their respective HRECs. Twin pair data helped in quality control checks but only one twin from each pair was used in analyses.

DNA was extracted using standard protocols and was genotyped using Infinium CoreExome-24 version 1.1 producing ~300,000 informative whole genome SNP markers. Standard GWAS quality control (QC) steps were performed, including sex-checks (incompatible sex between genotyping result on the X-chromosome and the individual's clinical record) and the removal of SNPs that were genotyped <95% of individuals, had a low minor allele frequency (MAF < 0.01) or deviated from Hardy–Weinberg Equilibrium (HWE) $p < 1 \times 10^{-6}$, using PLINK version 1.9 [25]. A total of ~250,000 SNPs passed quality control and were imputed to the Haplotype Reference Consortium reference panel (Version r1.1 2016) [26] implemented in Sanger Institute Imputation Server. SNPs with poor imputation accuracy (info score < 0.8) and low frequency SNPs (MAF < 0.01) were removed, leaving 6,681,912 SNPs for later analysis. The QC on individuals included filtering related individuals (Identity By Decent, IBD > 0.05, PLINK 1.9 "--genome" command) and individuals known to harbour Mendelian-like variants associated with ALS. To remove ancestry outliers, we projected our case–control cohort onto the first two principal components (PCs) of the 1000 Genomes cohort [27] using GCTA's PC loading method [28]. We removed the ancestry outliers that deviated more than four standard deviations from the European population mean (calculated using 1000 Genomes Northern European (CEU), British (GBR), Finnish (FIN), Iberian Spanish (IBS), and Toscani Italian (TSI) samples), leaving 1501 individuals (836 cases and 665 controls) to be used for further analyses. SNP data is available from dbGaP reference phs002068.v1.p1.

Selection of correlated traits

The genetic correlation between a range of traits with ALS was estimated using Linkage Disequilibrium Score Regression (LDSR) [29] between the European ALS GWAS summary statistics (20,806 ALS cases and 59,804 healthy controls of European ancestry) [30] to over 700 traits as implemented in the LDHub platform (v1.9.0) [22]. Despite the previous report of LDHub results for ALS [18], the GWAS data held within LDHub are regularly updated. Here, we report all genetic correlations significant at $p < 0.05$ for the test of null hypothesis $r_g = 0$, with traits that have SNP-based heritability estimate $>10\%$. We chose the minimum SNP-based heritability of 10% because our interest was to improve out-of-sample prediction and hence correlated traits need to have sufficient genetic contribution, noting that the SNP-based heritability of ALS estimated from the results of the latest GWAS [30] is only 1.76% (SE = 0.38%), calculated using a lifetime risk of 0.003 (ref. [26]).

Polygenic risk scores

We calculated a polygenic risk score (PRS) for all individuals in our Australian ALS case–control sample. The SNPs taken into the PRS calculations were limited to those SNPs found in HapMap 3 (HM3) [31] as these were common across the summary statistics of all traits analysed. PRSs were calculated using different methods to decide SNPs included and their effect sizes, but in each case, the PRS is the sum of risk alleles weighted by SNP effect sizes calculated using the PLINK 1.9 “--score”. The efficacy of the predictor was measured by the Nagelkerke- R^2 of the logistic regression of PRS on case–control status (R glm package [32] for logistic regression and fmsb function [33] for Nagelkerke- R^2 calculation) and by comparison of the odds of being a case in the 10th decile vs 1st decile or ordered PRS.

In the basic PRS approach, the SNPs were clumped (PLINK --clump), which selects a quasi-independent SNP set by taking the most associated SNP in a genomic region and excluding any SNP with $r^2 > 0.01$ with already selected SNPs. We considered a range of p -value thresholds for selection of SNPs into the PRS (see Supplementary Figure 1), but report in results from PRS using all HapMap3 SNPs, we call this standard PRS (STD_PRS). Including all SNPs in our prediction model rather than selecting the p -value threshold based on results from the data prevents the variance captured from PRS being biased due to winner’s curse [34], allowing fairer comparison across the methods.

Since the clumping r^2 threshold is arbitrary, we also used BLUP (Best Linear Unbiased Prediction) estimates of all SNPs to calculate a PRS_{BLUP}, an approach that

appropriately accounts for linkage disequilibrium (LD) of the SNPs, but assumes SNP effects are normally distributed (which is a valid assumption for highly polygenic traits). Approximate BLUP estimates were derived from GWAS summary statistics using the SBLUP [35] method implemented in the GCTA software with the Human Retirement Study (HRS) cohort [36] used as the reference sample to calculate the LD structure. We also calculated SNP effects using LDPred-Funct (LDPF) [37] a method that includes functional annotation to weight SNPs effects. We used the Baseline-LD functional annotation provided by Gazal et al. [38]. and the HRS cohort for the LD structure reference to calculate LDPF-inf SNP weightings. Lastly, we calculated SNP effects using Summary-based BayesR (SBayesR) [39], a method that models effect sizes using a mixture of normal distributions with different variances. This allows greater flexibility in the underlying model, potentially providing a better reflection of the underlying genetic architecture of ALS. We used sparse LD-matrix built from 10,000 UK Biobank [40] unrelated individuals for the LD reference. [39].

Out-of-sample prediction for a trait can be improved by using information from correlated traits [19–21], with multiple-trait prediction implemented in MTAG [19] and SMTpred [20] software. MTAG and SMTpred use similar methodologies to develop a multi-trait predictor. Here, we use MTAG to combine basic SNP effects of ALS and correlated traits generating a single effect size per SNP from which to generate a PRS. We use SMTpred (with --blup option) to combine single trait scores generated by SBLUP (PRS_{BLUP}), LDPred-Funct (PRS_{LDPF}), and SBayesR (PRS_{SBayesR}) for each individual using the estimated genetic correlation and SNP-based heritabilities of each single trait.

Results

Selection of correlated traits

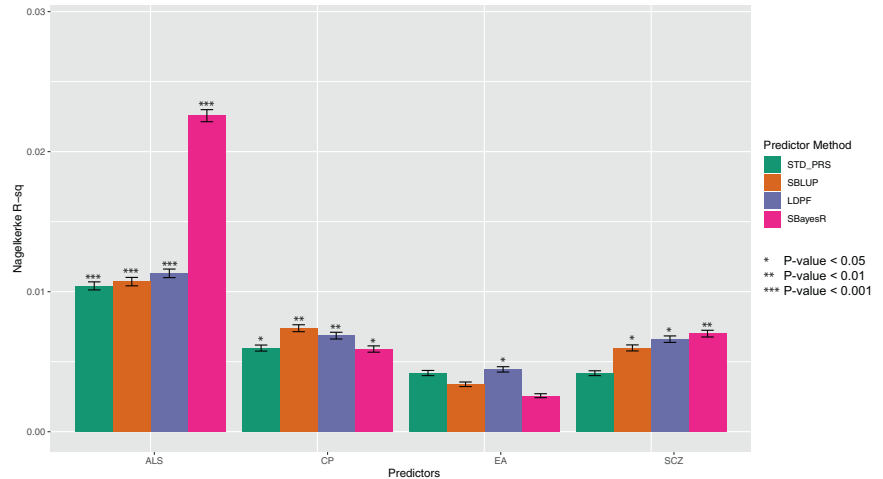
A genetic correlation analysis between ALS and more than 700 traits available in the LD Hub ver1.90 platform identified 85 traits that were significantly correlated with ALS (Supplementary table 1) at P value < 0.05 . After applying additional filters of SNP-based heritability estimate $>10\%$ and a 5% False Discovery Rate (FDR) (nominal P value $< 5.6 \times 10^{-3}$), three traits remained, all related to cognition (Table 1). As larger GWAS sample sizes are available for both education attainment (EA) and cognitive performance (CP) [41], and the results from these studies show a significant r_g with ALS of -0.28 and -0.24 respectively (Table 1), we took these GWAS summary statistics forward to improve statistical power in our prediction analysis. While the genetic correlation between ALS and

Table 1 Genetic correlations between ALS and traits used in prediction analysis.

Traits	rG	SE	SNP-h ²	P value	N (Case + Control)
Cognitive performance (CP) [41]	−0.28	0.06	0.2	1.11E−06	257,828
Educational attainment (EA) [41]	−0.24	0.05	0.11	1.10E−06	766,345
Schizophrenia (SCZ) [43]	0.15	0.05	0.42	2.60E−03	105,318

Fig. 1 Prediction accuracy of single-trait predictors of ALS in the Australian cohort.

Predictors constructed using GWAS summary statistics of CP, EA and SCZ had small but significant predictive ability for ALS case–control status.



schizophrenia [42] was not significant after correction for multiple testing in our LDhub analysis ($r_g = 0.14$, $p = 1.2 \times 10^{-2}$), the genetic correlation estimated with the latest schizophrenia GWAS [43] results was significant (Table 1). This observation combined with a previous report of a genetic correlation between ALS and SCZ [18], led us to take SCZ through to our prediction analysis.

Polygenic risk scores

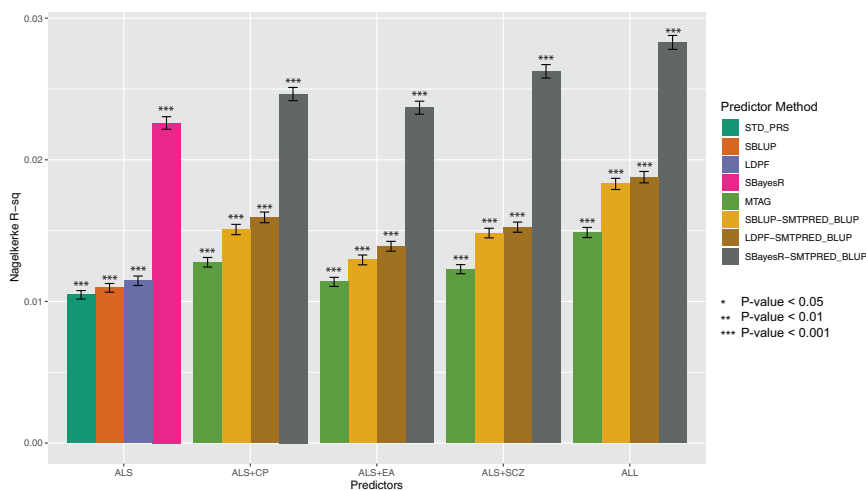
Results from the single-trait PRS prediction into the Australian sample of 836 cases and 665 controls for each of the four selected traits (ALS, CP, EA and SCZ) are summarised in Fig. 1 and Supplementary Table 2. As expected, the ALS discovery sample (HM3 SNP set) gives the best single trait prediction performance with a Nagelkerke- R^2 (NKR2) of 0.010 for standard PRS, 0.010 for SBLUP, 0.011 for LDPF and 0.022 for SBayesR. The CP PRS explained a significant (P value < 0.05) proportion of variance for all four methods, while the association between the EA predictor and ALS case–control status was nominally significant only for the LDPF method. Their regression coefficient had the expected negative sign, providing independent confirmation that ALS is genetically negative correlated with these traits. The predictors calculated from SCZ GWAS statistics are also significantly associated with ALS case–control, and the sign of the regression coefficient was positive as expected. In each single-trait prediction, the highest variance explained was always from either SBLUP, LDPF or SBayesR methods.

Combining these traits into multiple trait predictors of ALS generated higher NKR2 than the single trait predictors (Fig. 2 and Supplementary Table 3). Direct comparison on single trait vs multi-trait predictors is given by the PRS vs MTAG results, both of which do not consider the LD structure between SNPs, and the SBLUP vs SBLUP-SMTPRED, which do account for LD structure. The LDPF and LDPF-SMTPRED results include the functional annotations into SNP weights. The SBayesR and SBAYESR-SMTPRED results demonstrate the utility of a flexible distribution of effect sizes (modelled as a mixture of normal distributions) rather than a single underlying normal distribution as use by SBLUP. In all cases, the NKR2 of multi-traits predictors was higher than the predictors using ALS alone. Combining all correlated traits (ALS, CP, EA and SCZ) gave the best predictor (NKR2 of 0.027) with the SBayesR method. For the best predictor, the calculated risk odds ratio for those in the top 10% of estimated risk when compared to those in the bottom 10% was 3.15 (95% CI 1.96–5.05).

Discussion

Given that ALS is a complex disease, understanding its genetic relationship with other traits provides some insight into this complexity. Analyses using summary statistics from GWAS allow the study of the genetic relationship between traits using independently collected samples. We found that ALS had significant negative genetic correlation

Fig. 2 Prediction accuracy of multi-trait predictors compared to the ALS only predictor. Predictors constructed using combined SNP effects (multi-trait) of ALS and correlated traits, improved predictive ability for ALS case–control status.



(after multiple FDR correction) with cognitive related traits like fluid intelligence, years of schooling, and university/college qualification measured in large cohorts sampled from the general population, for example, $r_g = -0.24$ with educational attainment (Table 1). This observation supports the earlier US [10] and Italian [11] cohort studies. It is notable that the r_g with schizophrenia is positive (0.15) (Table 1), and that a negative r_g is also found between schizophrenia and educational attainment (-0.17) [44].

We also found that some physical activity traits, such as walking to work (SNP-based $h^2 = 2.2\%$), measured in the UK-Biobank has a significant negative genetic correlation with ALS (Supplementary Table 1), but other measures of exercise (including the duration and frequency of walking and vigorous or moderate exercise) did not show a significant genetic correlation to ALS ($p > 0.05$). Many studies provide support for an association between high levels of physical activity with increased risk of ALS. For example, an increased risk of ALS is reported for professional soccer and football players, and Gulf-war veterans [45–47]. Comprehensive epidemiological studies involving five different populations in the Netherlands, Ireland and Italy found that vigorous and moderate activity are linearly correlated with the risk of ALS [48]. Previous epidemiological studies from Japan and UK also found a similar correlation using a smaller sample size [49, 50]. These traits were not taken forward into our prediction analysis because the physical activity traits show low SNP-based heritability ($<10\%$), and so would not be expected to improve out of sample prediction. In addition, the estimates for genetic correlation between these traits and ALS should be treated with caution, given that ALS also has a small SNP-based heritability estimate [51].

Our goal was to provide independent evidence of the genetic correlation of ALS with schizophrenia and cognitive related traits, through out-of-sample prediction into an independent ALS Australian cohort. We show that there is a

significant out-of-sample prediction for ALS when using PRS built from EA, SCZ and especially CP SNP effect estimates, with the sign of regression coefficients matching the sign of the r_g estimates. These results provide independent validation of the genetic relationship between ALS, CP, EA and SCZ. As expected, out-of-sample prediction is maximised by combining all the traits to make a multi-trait predictor. We compared methods for generating PRS, and found in this context SBayesR gave highest out-of-sample prediction accuracy. Since the best methodology for PRS depends on the genetic architecture of the trait [52], this conclusion may not be true in other disease applications. While the out-of-sample prediction was found to be highly significant (smallest p value = 4.8×10^{-8}), the variance explained by the predictor was still small (maximum NKR2 = 0.027, maximum AUC = 0.580).

This study had several limitations. We used a single cohort to test the out-of-sample prediction. Application of PRS prediction in other disorders, such as schizophrenia [42] and major depression disorder [53], has found variability in results between cohorts. Hence, other, European ancestry ALS cohort/s would be useful to confirm this observation. In addition, the estimated SNP-based heritability of ALS using common SNPs (HM3) was very low (1.76%, SE = 0.38%, assuming lifetime risk of 0.3%) from the latest published GWAS [30], smaller than from the previously published GWAS ($h^2_{\text{SNP}} = 8\%$, SE = 0.52%) [15]. This low SNP-based heritability may reflect the genetic architecture of ALS, and previous analyses have suggested that low-frequency variants may be relatively more important in ALS than other common diseases [15].

In conclusion, we found that ALS had a significant negative genetic correlation with cognitive performance and educational attainment. These correlations were supported by the significant prediction of ALS when using the GWAS summary statistics for both traits, and improvements were made in prediction accuracy for ALS when included in a

multi-trait predictor. However, there is still limited clinical utility in these ALS predictors due to the relatively small proportion of risk they capture. Larger GWAS for ALS is needed in order to provide a stronger baseline from which multi-trait predictors can be built.

Acknowledgements We kindly thank all those who contributed to this research including the participants for providing a blood sample and clinical data, the research nurses and support staff for participant recruitment across clinic sites and the laboratory researchers for their care in generating the DNA data. We acknowledge funding from the National Health and Medical Research Council (NHMRC) (1078901, 1083187, 1113400, 1121962, 1405325, 1084417, and 1079583), a NHMRC/Australian Research Council Strategic Award (401162), the Motor Neurone Disease Research Institute Australia (MNDRIA) Ice Bucket Challenge Grant and the MNDRIA Bill Gole Postdoctoral Fellowship (FCG). The twin study (OATS) was facilitated through Twins Research Australia, a national resource in part supported by a NHMRC Centre for Research Excellence.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

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

References

- Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. *N Engl J Med*. 2017;377:162–72.
- van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH et al. Amyotrophic lateral sclerosis. *Lancet*. 2017. [https://doi.org/10.1016/S0140-6736\(17\)31287-4](https://doi.org/10.1016/S0140-6736(17)31287-4).
- Taylor JP, Brown RH Jr, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature*. 2016;539:197–206.
- Byrne S, Walsh C, Lynch C, Bede P, Elamin M, Kenna K, et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatr*. 2011;82:623–7.
- Longinetti E, Mariosa D, Larsson H, Almqvist C, Lichtenstein P, Ye W, et al. Physical and cognitive fitness in young adulthood and risk of amyotrophic lateral sclerosis at an early age. *Eur J Neurol*. 2017;24:137–42.
- Rippon GA, Scarneas N, Gordon PH, Murphy PL, Albert SM, Mitumoto H, et al. An observational study of cognitive impairment in amyotrophic lateral sclerosis. *Arch Neurol*. 2006;63:345–52.
- Irwin D, Lippa CF, Swearer JM. Cognition and amyotrophic lateral sclerosis (ALS). *Am J Alzheimers Dis Other Dement*. 2007;22:300–12.
- Raaphorst J, de Visser M, Linssen WHJP, de Haan RJ, Schmand B. The cognitive profile of amyotrophic lateral sclerosis: a meta-analysis. *Amyotroph Lateral Scler*. 2010;11:27–37.
- Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology*. 2005;65:586–90.
- Yu Y, Su F-C, Callaghan BC, Goutman SA, Batterman SA, Feldman EL. Environmental risk factors and amyotrophic lateral sclerosis (ALS): a case-control study of ALS in Michigan. *PLOS ONE*. 2014;9:e101186.
- Montuschi A, Iazzolino B, Calvo A, Moglia C, Lopiano L, Restagno G, et al. Cognitive correlates in amyotrophic lateral sclerosis: a population-based study in Italy. *J Neurol Neurosurg Psychiatry*. 2015;86:168–73.
- Deary IJ, Spinath FM, Bates TC. Genetics of intelligence. *Eur J Hum Genet*. 2006;14:690–700.
- Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *PNAS*. 2014;111:15273–8.
- Wingo TS, Cutler DJ, Yarab N, Kelly CM, Glass JD. The heritability of amyotrophic lateral sclerosis in a clinically ascertained United States research registry. *PLoS One* 2011;6. <https://doi.org/10.1371/journal.pone.0027985>.
- van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet*. 2016;48:1043–8.
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47:1236–41.
- Bandres-Ciga S, Noyce AJ, Hemani G, Nicolas A, Calvo A, Mora G, et al. Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. *Ann Neurol*. 2019;85:470–81.
- McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun*. 2017;8. <https://doi.org/10.1038/ncomms14774>.
- Turley P, Walters RK, Maghazian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50:229–37.
- Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA, et al. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun*. 2018;9:989.
- Maier R, Moser G, Chen G-B, Ripke S, Cross-Disorder Working Group of the Psychiatric Genomics Consortium, Coryell W, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet*. 2015;96:283–94.
- Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33:272–9.
- Brooks BR, Miller RG, Swash M, Munsat TL, World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Mot Neuron Disord*. 2000;1:293–9.
- Sachdev PS, Lammel A, Trollor JN, Lee T, Wright MJ, Ames D, et al. A comprehensive neuropsychiatric study of elderly twins: the older Australian twins study. *Twin Res Hum Genet*. 2009;12:573–82.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:1–16.
- Consortium the HR. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–83.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding

- from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291–5.
30. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron.* 2018;97:1268. e6
 31. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467:52–58.
 32. R: the R project for statistical computing. <https://www.r-project.org/> (accessed 29 Jan 2020).
 33. Nakazawa M *fmsb*: functions for Medical Statistics Book with some Demographic Data. 2019 <https://CRAN.R-project.org/package=fmsb> (accessed 29 Jan 2020).
 34. Balding DJ, Moltke I, Marioni J. *Handbook of Statistical Genomics* Fourth Edition. John Wiley & Sons, New Jersey, 2019.
 35. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al. Genetic evidence of assortative mating in humans. *Nat Hum Behav.* 2017;1:0016.
 36. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the health and retirement study (HRS). *Int J Epidemiol.* 2014;43:576–85.
 37. Marquez-Luna C, Gazal S, Loh P-R, Furlotte N, Auton A, Team 23andMe Research et al. Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. e-pub preprint bioRxiv 2018: 375337. <https://doi.org/10.1101/375337>.
 38. Gazal S, Finucane HK, Furlotte NA, Loh P-R, Palamara PF, Liu X, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet.* 2017;49:1421–7.
 39. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun.* 2019;10:1–11.
 40. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–9.
 41. Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* 2018;50:1112.
 42. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511:421–7.
 43. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* 2018;50:381–9.
 44. Trampush JW, Yang MLZ, Yu J, Knowles E, Davies G, Liewald DC, et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry.* 2017;22:336–45.
 45. Chio A, Calvo A, Dossena M, Ghiglione P, Mutani R, Mora G. ALS in Italian professional soccer players: the risk is still present and could be soccer-specific. *Amyotroph Lateral Scler.* 2009;10: 205–9.
 46. Lehman EJ, Hein MJ, Baron SL, Gersic CM. Neurodegenerative causes of death among retired National Football League players. *Neurology.* 2012;79:1970–4.
 47. Weisskopf MG, O'Reilly EJ, McCullough ML, Calle EE, Thun MJ, Cudkovic M, et al. Prospective study of military service and mortality from ALS. *Neurology.* 2005;64:32–37.
 48. Visser AE, Rooney JPK, D'Ovidio F, Westeneng H-J, Vermeulen RCH, Beghi E et al. Multicentre, cross-cultural, population-based, case-control study of physical activity as risk factor for amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry.* 2018: jnnp-2017-317724.
 49. Okamoto K, Kihira T, Kondo T, Kobashi G, Washio M, Sasaki S, et al. Lifestyle factors and risk of amyotrophic lateral sclerosis: a case-control study in Japan. *Ann Epidemiol.* 2009;19:359–64.
 50. Harwood CA, Westgate K, Gunstone S, Brage S, Wareham NJ, McDermott CJ, et al. Long-term physical activity: an exogenous risk factor for sporadic amyotrophic lateral sclerosis? *Amyotroph Lateral Scler Frontotemporal Degener.* 2016;17:377–84.
 51. Rheenen W, van, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet.* 2019;20:567–81.
 52. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park J-H. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet.* 2013;45:400–5.
 53. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50:668–81.

Affiliations

Restuadi Restuadi¹ · Fleur C. Garton¹ · Beben Benyamin^{1,2} · Tian Lin¹ · Kelly L. Williams³ · Anna Vinkhuyzen¹ · Wouter van Rheenen⁴ · Zhihong Zhu¹ · Nigel G. Laing^{5,6} · Karen A. Mather^{7,8} · Perminder S. Sachdev^{7,9} · Shyuan T. Ngo^{10,11,12} · Frederik J. Steyn^{12,13} · Leanne Wallace¹ · Anjali K. Henders¹ · Peter M. Visscher^{1,10} · Merrilee Needham^{14,15,16} · Susan Mathers¹⁷ · Garth Nicholson^{3,18} · Dominic B. Rowe³ · Robert D. Henderson^{10,11,19} · Pamela A. McCombe^{12,19} · Roger Pamphlett²⁰ · Ian P. Blair³ · Naomi R. Wray^{1,10} · Allan F. McRae¹

¹ Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

² Australian Centre for Precision Health, University of South Australia Cancer Research Institute, School of Health Sciences, University of South Australia, Adelaide, SA, Australia

³ Centre for MND Research, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia

⁴ UMC Utrecht Brain Center Rudolf Magnus, Utrecht, Netherlands

⁵ Centre for Medical Research, University of Western Australia, Nedlands, WA, Australia

⁶ Harry Perkins Institute of Medical Research, Nedlands, WA, Australia

⁷ Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW, Australia

- 8 Neuroscience Research Australia Institute, Randwick, NSW, Australia
- 9 Neuropsychiatric Institute, The Prince of Wales Hospital, UNSW, Randwick, NSW, Australia
- 10 Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
- 11 The Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, QLD, Australia
- 12 Centre for Clinical Research, The University of Queensland, Brisbane, QLD, Australia
- 13 School of Biomedical Sciences, The University of Queensland, Brisbane, QLD, Australia
- 14 Fiona Stanley Hospital, Perth, WA, Australia
- 15 Notre Dame University, Fremantle, WA, Australia
- 16 Institute for Immunology and Infectious Diseases, Murdoch University, Perth, WA, Australia
- 17 Calvary Health Care Bethlehem, Parkdale, VIC, Australia
- 18 ANZAC Research Institute, Concord Repatriation General Hospital, Sydney, NSW, Australia
- 19 Department of Neurology, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia
- 20 Discipline of Pathology and Department of Neuropathology, Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia