



OPEN

Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease with a lifetime risk of one in 350 people and an unmet need for disease-modifying therapies. We conducted a cross-ancestry genome-wide association study (GWAS) including 29,612 patients with ALS and 122,656 controls, which identified 15 risk loci. When combined with 8,953 individuals with whole-genome sequencing (6,538 patients, 2,415 controls) and a large cortex-derived expression quantitative trait locus (eQTL) dataset (MetaBrain), analyses revealed locus-specific genetic architectures in which we prioritized genes either through rare variants, short tandem repeats or regulatory effects. ALS-associated risk loci were shared with multiple traits within the neurodegenerative spectrum but with distinct enrichment patterns across brain regions and cell types. Of the environmental and lifestyle risk factors obtained from the literature, Mendelian randomization analyses indicated a causal role for high cholesterol levels. The combination of all ALS-associated signals reveals a role for perturbations in vesicle-mediated transport and autophagy and provides evidence for cell-autonomous disease initiation in glutamatergic neurons.

ALS is a fatal neurodegenerative disease affecting one in 350 individuals. Due to degeneration of both upper and lower motor neurons, patients suffer from progressive paralysis, ultimately leading to respiratory failure within 3–5 years after disease onset¹. In ~10% of patients with ALS, there is a clear family history for ALS, suggesting a strong genetic predisposition, and currently a pathogenic mutation can be found in more than half of these cases². On the other hand, apparently sporadic ALS is considered a complex trait for which heritability is estimated at 40–50% (refs.^{3,4}). There is no widely accepted definition of familial or sporadic ALS⁵, and they are likely to represent the ends of a spectrum with overlapping genetic architectures for which the same genes have been implicated in both familial and sporadic disease^{6–11}. To date, partially overlapping GWASs have identified up to six genome-wide significant loci, explaining a small proportion of the genetic susceptibility to ALS^{11–16}. Indeed, some of these loci found in GWASs harbor rare variants with large effects also present in familial cases (for example, *C9orf72* and *TBK1*)^{6,17,18}. For other loci, the role of rare variants remains unknown.

While ALS is referred to as a motor neuron disease, cognitive and behavioral changes are observed in up to 50% of patients, sometimes leading to frontotemporal dementia (FTD). The overlap with FTD is clearly illustrated by the pathogenic hexanucleotide repeat expansion in *C9orf72*, which causes familial ALS and/or FTD^{17,18} and the genome-wide genetic correlation between ALS and FTD¹⁹. Further expanding the ALS–FTD spectrum, a genetic correlation with progressive supranuclear palsy (PSP) has been described²⁰. Shared pathogenic mechanisms between ALS and other neurodegenerative diseases, including common diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD), can further reveal ALS pathophysiology and inform new therapeutic strategies.

Here, we combine new and existing individual-level genotype data in the largest GWAS of ALS to date. We present a

comprehensive screen for pathogenic rare variants and short tandem repeat (STR) expansions as well as regulatory effects observed in brain cortex-derived RNA sequencing (RNA-seq) and methylation datasets to prioritize causal genes within ALS-risk loci. Furthermore, we reveal similarities and differences between ALS and other neurodegenerative diseases as well as the biological processes in disease-relevant tissues and cell types that affect ALS risk.

Results

Cross-ancestry meta-analysis reveals 15 risk loci for ALS. To generate the largest GWAS of ALS to date, we merged individual-level genotype data from 117 cohorts into six strata matched by genotyping platform. A total of 27,205 patients with ALS and 110,881 control participants of European ancestries passed quality control (including 6,374 newly genotyped cases and 22,526 control participants; Methods and Supplementary Tables 1 and 2). Patients were not selected for a family history of ALS. Through meta-analysis of these six strata, we obtained association statistics for 10,461,755 variants down to a minor allele frequency (MAF) of 0.1% in the Haplotype Reference Consortium resource²¹. We observed moderate inflation of the test statistics ($\lambda_{GC} = 1.12$, $\lambda_{1000} = 1.003$), and linkage disequilibrium (LD) score regression yielded an intercept of 1.029 (s.e.m. = 0.0073), indicating that the majority of inflation was due to the polygenic signal in ALS (LD score regression (LDSC): $h^2 = 0.028$, s.e.m. = 0.003, $K = 350^{-1}$, $P = 5.5 \times 10^{-21}$). The European ancestry analysis identified 12 loci reaching genome-wide significance ($P < 5.0 \times 10^{-8}$; Extended Data Fig. 1). For nine loci, the top SNP or a strong LD proxy ($r^2 = 0.996$) was present in GWAS of ALS in Asian ancestries (2,407 patients with ALS and 11,775 control participants)^{15,16}, and all showed a consistent direction of effects ($P_{\text{binom}} = 2.0 \times 10^{-3}$). The three SNPs that were not present in the Asian ancestry GWAS were low-frequency variants (MAF of 0.6–1.6% in European ancestries, Table 1). The genetic overlap between ALS

A full list of affiliations appears at the end of the paper.

64 risk in European and Asian ancestries resulted in a trans-ancestry
65 genetic correlation of 0.57 (s.e.m. = 0.28) for genetic effect and 0.58
66 (s.e.m. = 0.30) for genetic impact, which were not statistically signif-
67 icantly different from unity ($P = 0.13$ and $P = 0.16$, respectively). We
68 therefore performed a cross-ancestry meta-analysis totaling 29,612
69 cases and 122,656 controls, which revealed three additional loci,
70 totaling 15 genome-wide significant risk loci for ALS risk (Fig. 1,
71 Table 1 and Supplementary Tables 4–18). Conditional and joint
72 analysis did not identify secondary signals within these loci.

73 Of these findings, eight loci have been reported in previ-
74 ous GWASs (*C9orf72*, *UNC13A*, *SCFD1*, *MOBP-RPSA*, *KIF5A*,
75 *CFAP410*, *GPX3-TNIP1* and *TBK1*)^{11,14,15}. The *rs80265967* vari-
76 ant corresponds to the p.D90A mutation in *SOD1* previously
77 identified in a Finnish ALS cohort enriched for familial ALS¹³.

78 Interestingly, we observed a genome-wide significant common
79 variant association signal within the *NEK1* locus, which was pre-
80 viously shown to harbor rare variants associated with ALS⁸. The
81 recently reported association at the *ACSL5-ZDHHC6* locus^{16,22} did
82 not reach the threshold for genome-wide significance (*rs58854276*,
83 $P_{\text{EUR}} = 5.4 \times 10^{-5}$, $P_{\text{ASN}} = 4.9 \times 10^{-7}$, $P_{\text{comb}} = 6.5 \times 10^{-8}$; Supplementary
84 Table 19), despite the fact that our analysis includes all data from the
85 original discovery studies.

86
87 **Rare variant gene-based association analyses in ALS.** To assess
88 a general pattern of underlying architectures that link associated
89 SNPs to causal genes, we first tested for annotation-specific enrich-
90 ment using stratified LDSC. This revealed that 5' UTR regions as
91 well as coding regions in the genome and those annotated as con-
92 served were most enriched for ALS-associated SNPs (Extended
93 Data Fig. 2). Subsequently, we investigated how rare, coding variants
94 contributed to ALS risk by generating a whole-genome sequenc-
95 ing (WGS) dataset of patients with ALS ($n = 6,538$) and control
96 participants ($n = 2,415$), which is a subset of the common vari-
97 ant GWAS cohort. The exome-wide association analysis included
98 transcript-level rare variant burden testing for different models of
99 allele-frequency thresholds and variant annotations (Methods).
100 This identified *NEK1* as the strongest associated gene (minimal
101 $P = 4.9 \times 10^{-8}$ for disruptive and damaging variants at $\text{MAF} < 0.005$),
102 which was the only gene to pass the exome-wide significance
103 thresholds ($0.05 \div 17,994 = 2.8 \times 10^{-6}$ and $0.05 \div 58,058 = 8.6 \times 10^{-7}$
104 for number of genes and protein-coding transcripts, respectively;
105 Supplementary Table 20). This association was independent from
106 the previously reported increased rare variant burden in selected
107 patients with 'familial ALS' (ref. ⁸) who were not included in this
108 study. Polygenic risk score (PRS) analyses did not illustrate a dif-
109 ference in PRSs in patients carrying rare variants in ALS-risk genes
110 (*SOD1*, *C9orf72* repeat expansion, *TARDBP*, *FUS*, *NEK1*, *TBK1*
111 and *CFAP410*) compared to all patients with ALS (Extended Data
112 Fig. 3). Although power was limited, this is compatible with a sce-
113 nario in which the genetic risk of ALS in these patients is a sum of
114 rare variants in ALS genes and other (common) genetic variation.

115
116 **Gene prioritization shows locus-specific underlying architec-
117 tures.** To assess whether rare variant associations could drive the
118 common variant signals at the 15 genome-wide significant loci, we
119 combined the common and rare variant analyses to prioritize genes
120 within these loci. The SNP effects on gene expression were assessed
121 by summary-based Mendelian randomization (MR) (SMR) in blood
122 (eQTLGen²³, $n = 31,648$) and a new brain cortex-derived eQTL
123 dataset (MetaBrain²⁴, $n = 2,970$). Finally, we analyzed methylation
124 quantitative trait loci (mQTL) by SMR in blood-derived ($n = 2,082$)
125 and brain-derived ($n = 522$) mQTL datasets^{25–27}. Through these
126 multi-layered gene-prioritization strategies, we classified each locus
127 into one of four classes of most likely underlying genetic architec-
128 ture to prioritize the causal gene (Supplementary Figs. 1–15).

First, in three GWAS loci, the strongest associated SNP was a
low-frequency coding variant that was nominated as the causal vari-
ant. This was the case for *rs80265967* (*SOD1*, p.D90A; Supplementary
Fig. 14) and *rs113247976* (*KIF5A*, p.P986L; Supplementary Fig. 8),
which are coding variants in known ALS-risk genes. This was
also the most likely causal mechanism for *rs75087725* (*CFAP410*,
formerly *C21orf2*, p.V58L; Supplementary Fig. 15), as the GWAS
variant is a missense variant; no evidence for other mechanisms
including repeat expansions or eQTL or mQTL effects was observed
within this locus, and *CFAP410* itself is known to directly interact
with *NEK1*, another ALS gene^{6,28}. These three loci illustrate the
power of large-scale GWASs combined with modern imputation
panels to directly identify low-frequency causal variants that confer
disease risk.

Second, SNPs can tag a highly pathogenic repeat expansion,
as was observed for *rs2453555* (*C9orf72*) and the known
GGGGCC hexanucleotide repeat in this locus (Supplementary
Fig. 7). Conditional analysis revealed no residual signal after
conditioning on the repeat expansion, which was in LD with the
top SNP ($r^2 = 0.14$, $|D'| = 0.99$, $\text{MAF}_{\text{SNP}} = 0.25$, $\text{MAF}_{\text{STR}} = 0.047$).
Besides the repeat expansion, both eQTL and mQTL analyses
point to *C9orf72* (Supplementary Fig. 7). The HEIDI outlier
test, however, rejected the null hypothesis that gene expression
or methylation mediated the causal effect of the associated SNP
($P_{\text{HEIDI,eQTL}} = 3.7 \times 10^{-23}$ and $P_{\text{HEIDI,mQTL}} = 4.1 \times 10^{-7}$). This is in line
with the idea that pathogenic repeat expansion is the causal vari-
ant in this locus and that eQTL and mQTL effects do not mediate
a causal effect. We found no similar pathogenic repeat expan-
sions that fully explained the SNP association signal in the other
genome-wide significant loci.

Third, in two loci (*rs62333164* in *NEK1* and *rs4075094* in
TBK1), common and rare variants converged to the same gene,
which are known ALS-risk genes^{6,8}. For both loci, the rare vari-
ant burden association was conditionally independent from
the top SNP that was included in the GWAS (Supplementary
Figs. 2 and 9). Here, eQTL and mQTL analyses indicated that
the risk-increasing effects of the common variants were medi-
ated through both eQTL and mQTL effects on *NEK1* and *TBK1*.
Furthermore, a polymorphic STR downstream of *NEK1* was
associated with increased ALS risk (motif, TTTA; threshold = 10
repeat units, expanded allele frequency = 0.51, $P = 5.2 \times 10^{-5}$, false
discovery rate (FDR) = 4.7×10^{-4} ; Extended Data Fig. 4). This
polymorphic repeat was in LD with the top associated SNP within
this locus ($r^2 = 0.24$, $|D'| = 0.70$). There was no statistically sig-
nificant association for the top SNP in the WGS data to reliably
determine its independent contribution to ALS risk.

Lastly, the fourth group contains seven remaining loci for which
there was no direct link to a causal gene through coding variants or
repeat expansions. Here, we investigated regulatory effects of the
associated SNPs on target genes acting as either eQTL or mQTL.
Single genes were prioritized by SMR using both mQTL and eQTL
for *rs2985994* (*COG3*; Supplementary Fig. 10), *rs229243* (*SCFD1*;
Supplementary Fig. 11) and *rs517339* (*ERGIC1*; Supplementary
Fig. 4). In other loci, both methods prioritized multiple genes,
such as *rs631312* (*MOBP* and *RPSA*; Supplementary Fig. 1) and
rs10463311 (*GPX3* and *TNIP1*; Supplementary Fig. 3). Aside from
the prioritized genes, each of these loci harbored multiple genes that
were not prioritized by any method and are therefore less likely to
contribute to ALS risk.

For two loci, no gene was prioritized with these approaches.
Within the *UNC13A* locus (*rs12608932*; Supplementary Fig. 12),
recent studies illustrate that the genome-wide significant SNPs act
as splicing quantitative trait loci conditional on dysfunction of TAR
DNA-binding protein (TDP)-43, resulting in inclusion of a cryptic
exon in *UNC13A*^{29,30}. Furthermore, we could not prioritize a specific
gene in the *HLA* locus (*rs9275477*; Supplementary Fig. 5).

Table 1 | Genome-wide significant loci

Chr	Position (bp)	ID	Prioritized gene	A ₁	A ₂	Freq	European ancestries			Asian ancestries			Cross-ancestry		
							Effect (s.e.m.)	P	P	Effect (s.e.m.)	P	Effect (s.e.m.)	P	P	
9	27,563,868	rs2453555	C9orf72	A	G	0.248	0.174 (0.013)	1.0 × 10 ⁻⁴³	0.017 (0.066)	0.80	0.168 (0.012)	1.5 × 10 ⁻⁴¹			
19	17,752,689	rs12608932	UNC13A	C	A	0.347	0.125 (0.012)	8.8 × 10 ⁻²⁵	0.074 (0.038)	0.053	0.120 (0.012)	3.0 × 10 ⁻²⁵			
21	33,039,603	rs80265967	SOD1	C	A	0.006	1.078 (0.124)	3.5 × 10 ⁻¹⁸	-	-	-	-			
14	31,045,596	rs229195	SCFD1	A	G	0.337	0.091 (0.012)	9.2 × 10 ⁻¹⁵	-	-	-	-			
14	31,045,181	rs229194*	SCFD1	A	G	0.337	0.091 (0.012)	9.2 × 10 ⁻¹⁵	0.002 (0.036)	0.97	0.083 (0.011)	1.5 × 10 ⁻¹³			
3	39,508,968	rs631312	MOBP, RPSA	G	A	0.291	0.079 (0.012)	5.2 × 10 ⁻¹¹	0.084 (0.036)	0.020	0.080 (0.011)	3.3 × 10 ⁻¹²			
6	32,672,641	rs9275477	HLA	C	A	0.096	-0.143 (0.021)	5.5 × 10 ⁻¹²	-0.110 (0.111)	0.32	-0.142 (0.02)	3.5 × 10 ⁻¹²			
12	57,975,700	rs113247976	KIF5A	T	A	0.016	0.332 (0.049)	1.4 × 10 ⁻¹¹	-	-	-	-			
21	45,753,117	rs75087725	CFAP410	A	C	0.012	0.418 (0.063)	2.7 × 10 ⁻¹¹	-	-	-	-			
5	150,410,835	rs10463311	GPX3, TNIP1	C	T	0.253	0.079 (0.013)	3.5 × 10 ⁻¹⁰	0.042 (0.036)	0.24	0.075 (0.012)	2.7 × 10 ⁻¹⁰			
20	48,438,761	rs17785991	SLC9A8, SPATA2	A	T	0.353	0.074 (0.012)	3.5 × 10 ⁻¹⁰	0.045 (0.076)	0.55	0.073 (0.012)	3.2 × 10 ⁻¹⁰			
12	64,877,053	rs4075094	TBK1	A	T	0.112	-0.098 (0.018)	1.7 × 10 ⁻⁸	-0.216 (0.090)	0.017	-0.103 (0.017)	2.1 × 10 ⁻⁹			
5	172,354,731	rs517339	ERGIC1	C	T	0.397	-0.065 (0.011)	8.5 × 10 ⁻⁹	-0.067 (0.074)	0.37	-0.065 (0.011)	5.6 × 10 ⁻⁹			
4	170,583,157	rs62333164	NEK1	A	G	0.335	0.063 (0.012)	7.0 × 10 ⁻⁸	0.203 (0.070)	3.8 × 10 ⁻³	0.067 (0.012)	6.9 × 10 ⁻⁹			
13	46,113,984	rs2985994	COG3	C	T	0.259	0.066 (0.013)	1.9 × 10 ⁻⁷	0.100 (0.041)	0.014	0.069 (0.012)	1.2 × 10 ⁻⁸			
7	157,481,780	rs10280711	PTPRN2	G	C	0.124	0.076 (0.017)	5.8 × 10 ⁻⁶	0.132 (0.037)	2.9 × 10 ⁻⁴	0.086 (0.015)	1.8 × 10 ⁻⁸			

Details of two-sided SAIGE logistic mixed model regression for the top associated SNPs within each genome-wide significant locus ($P < 5 \times 10^{-8}$). For the strongest associated SNP in the SCFD1 locus, rs229195 (MAF = 0.337), details of the LD proxy rs229194 are described (MAF = 0.337, $r^2 = 0.996$ in Asian ancestries), as only the LD proxy was present in the Asian ancestry GWAS. The low-frequency SNPs rs80265967, rs113247976 and rs75087725 were not present in the Asian ancestry GWAS, and no LD proxies ($r^2 > 0.8$) were found. Chr, chromosome; position, position in the reference genome GRCh37; A₁, effect allele; A₂, non-effect allele; freq, frequency of the effect allele in the European ancestry GWAS; s.e.m., standard error of the effect estimate.

Genetic modifiers of ALS disease progression. We investigated whether genetic risk factors for ALS also act as disease modifiers that affect disease onset and progression. Genotypes for the 15 genome-wide significant SNPs, PRSs and the rare variant burden for *SOD1*, *C9orf72* (repeat expansion status), *TARDBP*, *FUS*, *NEK1*, *TBK1* and *CFAP410* were obtained for all individuals with WGS for whom the complete core clinical data (sex, age at onset, site of onset, survival, time to censoring) were available ($n = 6,095$). Association analyses with survival and age at onset showed that common variants had a limited effect on survival (Fig. 2a) and age at onset (Fig. 2b) but confirmed the association between faster disease progression for the *UNC13A* risk allele (rs12608932, hazard ratio (HR) = 1.10, 95% confidence interval (CI) = 1.05–1.15, $P = 1.2 \times 10^{-4}$) and slower disease progression in patients with the *SOD1* p.D90A mutation (rs80265967, HR = 0.35, 95% CI = 0.16–0.77, $P = 8.4 \times 10^{-4}$). This limited effect of common genetic risk factors for ALS susceptibility on disease progression was reflected in the PRS analyses in which we found no effect of the full-genome PRS on survival (HR = 1.02, 95% CI = 0.98–1.06, $P = 0.28$) or age at onset ($b = 0.10$, s.e.m. = 0.21, $P = 0.64$). Analyses of rare variants confirmed faster disease progression in patients with the *C9orf72* repeat expansion (HR = 1.45, 95% CI = 1.28–1.65, $P = 1.2 \times 10^{-8}$) with an earlier age at onset ($b = -2.62$, s.e.m. = 0.77, $P = 6.4 \times 10^{-4}$).

Locus-specific sharing of risk loci between ALS and neurodegenerative diseases. To investigate the pleiotropic properties of ALS-associated variants and shared genetic risk with other brain diseases, we estimated genetic correlations between neurodegenerative diseases, psychiatric traits, cerebrovascular diseases and multiple sclerosis (Extended Data Fig. 5). This showed strong genetic correlations among neurodegenerative diseases. Bivariate LDSC confirmed a statistically significant genetic correlation between ALS and PSP ($r_g = 0.44$, s.e.m. = 0.11, $P = 1.0 \times 10^{-4}$) as previously reported²⁰ and also revealed a significant genetic correlation between ALS and AD ($r_g = 0.31$, s.e.m. = 0.12, $P = 9.6 \times 10^{-3}$) as well as between ALS and PD ($r_g = 0.16$, s.e.m. = 0.061, $P = 0.011$; Fig. 3a). The point estimate for the genetic correlation between ALS and FTD was high ($r_g = 0.59$, s.e.m. = 0.41, $P = 0.15$) but not statistically significant due to the limited size of the FTD GWAS (3,526 cases and 9,402 controls). Thus, power to detect a genetic correlation between ALS and FTD using LDSC was limited.

Patterns of sharing disease-associated genetic variants appeared to be locus specific (Fig. 3b and Supplementary Table 21). To assess whether two traits shared a common signal, indicating shared causal variants, we performed colocalization analyses for all loci meeting $P < 5 \times 10^{-5}$ in any of the GWASs of neurodegenerative diseases ($n = 161$ loci). This revealed a shared signal in the *MOBP-RPSA* locus between ALS, PSP and corticobasal degeneration (CBD) as well as a shared signal in the *UNC13A* locus between ALS and FTD (posterior probability, $PP_{H4} > 95\%$; Extended Data Fig. 6). For the *HLA* locus, there was evidence for a shared causal variant between ALS and PD ($PP_{H4} = 88\%$) but no conclusive evidence for ALS and AD ($PP_{H4} = 51\%$ for a shared causal variant and $PP_{H3} = 49\%$ for independent signals in both traits).

Furthermore, colocalization analyses identified two additional shared loci that were not genome-wide significant in the ALS GWAS: between ALS and PD at the *GAK* locus (rs34311866, $PP_{H4} = 99\%$) and between ALS and AD at the *BZRAP-ASI* locus (rs2632516, $PP_{H4} = 90\%$). Of note, the association at *BZRAP-ASI* was not genome-wide significant in the GWAS of clinically diagnosed AD ($P = 3.7 \times 10^{-7}$) either but was identified in the larger AD-by-proxy GWAS³¹. For FTD subtypes, *C9orf72* showed a colocalization signal for a shared causal variant between ALS and the motor neuron disease subtype of FTD (mndFTD, $PP_{H4} = 93\%$; Extended Data Figs. 6 and 7).

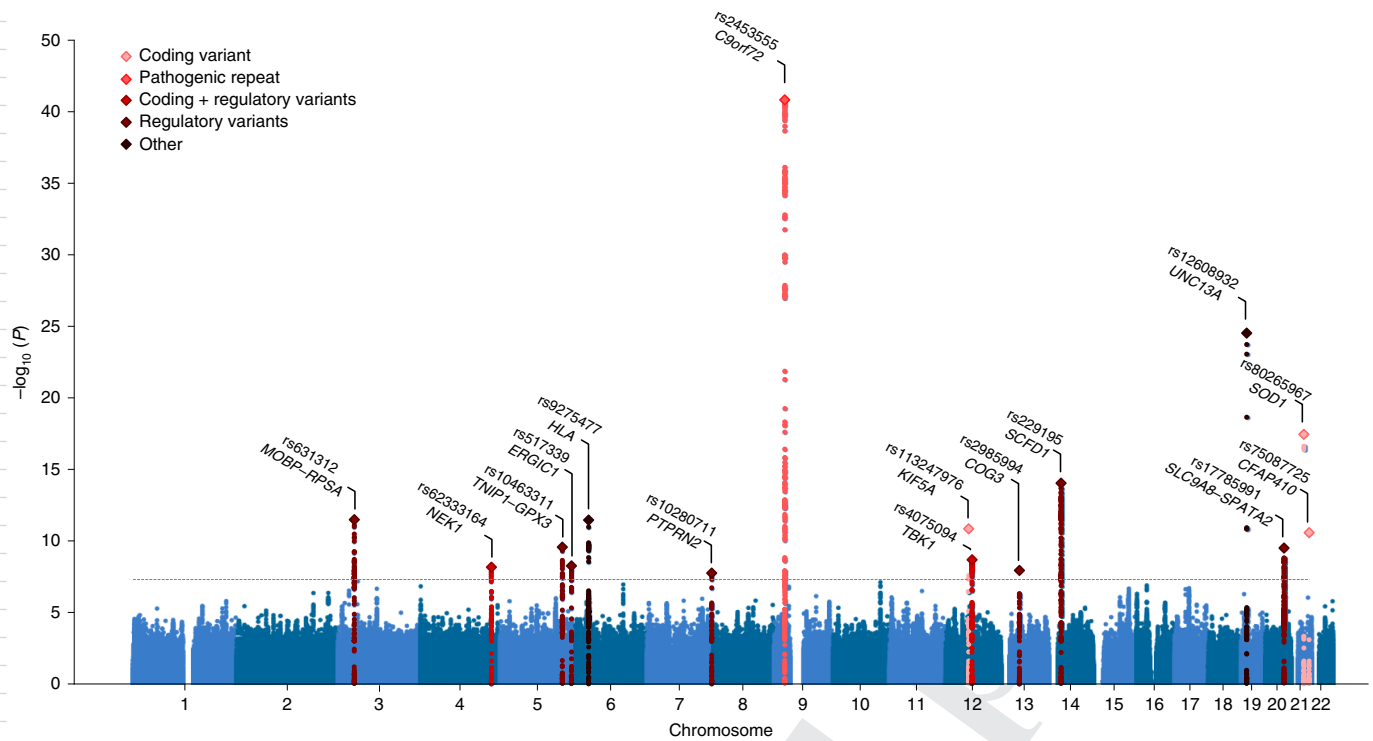


Fig. 1 | Manhattan plot of cross-ancestry meta-analysis. Genome-wide association statistics obtained by IVW meta-analysis of the stratified SAIGE logistic mixed model regression. The y axis corresponds to two-tailed $-\log_{10}(P)$ values; the x axis corresponds to genomic coordinates (GRCh37). The horizontal dashed line reflects the threshold for calling genome-wide significant SNPs ($P = 5 \times 10^{-8}$). Color coding and gene labels reflect those prioritized by the gene-prioritization analysis. Labels in bold indicate genes with known highly pathogenic mutations for ALS.

Enrichment of glutamatergic neurons indicates cell-autonomous processes in ALS susceptibility.

To find tissues and cell types for which gene expression profiles were enriched for genes within ALS-risk loci, we first combined gene-based association statistics calculated using MAGMA³² with gene expression patterns from the Genotype-Tissue Expression (GTEx) project (version 8) in a gene set enrichment analysis using FUMA³³. We observed a significant enrichment in genes expressed in brain tissues across multiple brain regions but not in peripheral nervous tissue or muscle. Whereas this pattern roughly resembled the enrichments observed in PD and psychiatric traits, it was strikingly different from that reported³¹ and observed in AD in which blood, lung and spleen were mostly enriched, resembling the pattern observed in multiple sclerosis, which is a typical immune-mediated brain disease (Fig. 4a and full results in Supplementary Fig. 16 and Extended Data Fig. 8a). We subsequently queried single-cell RNA-seq datasets of human-derived brain samples to further specify brain-specific enriched cell types using the cell type analysis module in FUMA³⁴. This showed significant enrichment for neurons but not for microglia or astrocytes (Fig. 4b). Further subtyping of these neurons illustrated that genes expressed in glutamatergic neurons were mostly enriched for genes within the ALS-associated risk loci. Again, this contrasted with AD, which showed specific enrichment of microglia, similar to multiple sclerosis (Extended Data Fig. 8b). In single-cell RNA-seq data obtained from brain tissues in mice, a similar pattern was observed showing neuron-specific enrichment in ALS and PD but microglia in AD (Extended Data Fig. 9). Together, this indicates that susceptibility to neurodegeneration in ALS is mainly driven by neuron-specific pathology and not by immune-related tissues and microglia.

Brain-specific coexpression networks improve detection of ALS-relevant pathways.

To determine which processes were

mostly enriched in ALS, we performed enrichment analyses that combined gene-based association statistics with gene coexpression patterns obtained from either multi-tissue transcriptome datasets³⁵ or RNA-seq data from brain cortex samples (MetaBrain²⁴). To validate this approach, we first tested for enrichment of human phenotype ontology (HPO) terms that are linked to well-established disease genes in the Online Mendelian Inheritance in Man (OMIM) and Orphanet catalogs. Using the multi-tissue coexpression matrix, we found no enriched HPO terms after Bonferroni correction for multiple testing. Using the brain-specific coexpression matrix, however, we found a strong enrichment of HPO terms that are related to ALS or neurodegenerative diseases in general, including ‘cerebral cortical atrophy’ ($P = 1.8 \times 10^{-8}$), ‘abnormal nervous system electrophysiology’ ($P = 4.1 \times 10^{-7}$) and ‘distal amyotrophy’ ($P = 8.6 \times 10^{-7}$; full list in Supplementary Table 22). In general, HPO terms in the neurological branch (‘abnormality of the nervous system’) showed an increase in enrichment statistics in ALS when using the brain-specific coexpression matrix compared to the multi-tissue dataset (Extended Data Fig. 10), which illustrates the benefit of the brain-specific coexpression matrix. Subsequently, we tested for enriched biological processes using reactome and gene ontology terms. Again, using the multi-tissue expression profiles, we found that no reactome annotations were enriched. Leveraging the brain-specific coexpression networks, we identified vesicle-mediated transport (‘membrane trafficking’, $P = 4.2 \times 10^{-6}$, ‘intra-Golgi and retrograde Golgi-to-endoplasmic reticulum (ER) trafficking’, $P = 1.4 \times 10^{-5}$) and autophagy (‘macroautophagy’, $P = 3.2 \times 10^{-5}$) as enriched processes after Bonferroni correction for multiple testing (Supplementary Table 23). The subsequently identified enriched gene ontology terms were all related to vesicle-mediated transport or autophagy (Supplementary Tables 24 and 25).

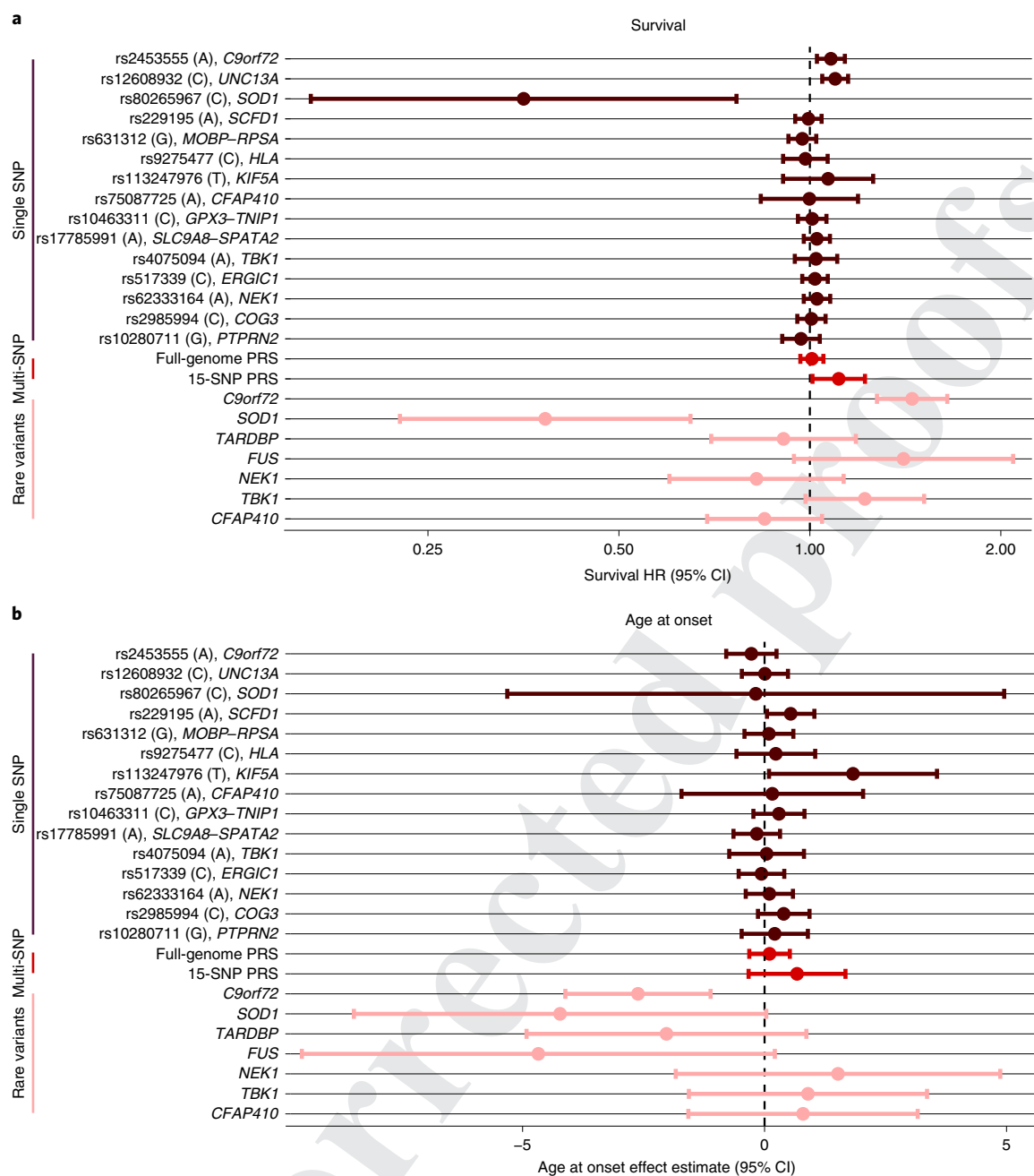


Fig. 2 | Genetic modifier analyses. **a**, Cox proportional HRs for genome-wide significant SNPs (brown, $n=15$), PRSs (red, $n=2$) and rare variant burden in ALS-risk genes (pink, $n=7$) on survival (months) tested in 6,095 patients with ALS. Estimated HRs are displayed with error bars corresponding to 95% CIs. Higher HRs correspond to shorter survival times. **b**, Effect estimates from a linear regression model of age at onset (years) in 6,095 patients with ALS. Lower effect estimates correspond to a younger age at onset. Effect estimates from linear regression are displayed with error bars corresponding to 95% CIs. The risk-increasing allele for ALS corresponds to the effect allele for both survival and age-at-onset analyses.

MR analyses are in line with a causal relationship between cholesterol levels and ALS. From previous observational case-control studies and our accompanying blood-based methylome-wide study³⁶, numerous non-genetic risk factors have been implicated in ALS. Here, we studied a selection of those putative risk factors through causal inference in an MR framework³⁷. We selected 22 risk factors for which robust genetic predictors were available including body mass index, smoking, alcohol consumption, physical activity, cholesterol-related traits, cardiovascular diseases and inflammatory markers (Supplementary Table 26). These analyses

provided the strongest evidence that cholesterol levels were causally related to ALS risk ($\beta_{\text{weighted median}}=0.15$, s.e.m. = 0.04, $P=3.2 \times 10^{-4}$; Fig. 5a and full results in Supplementary Table 27). These results were robust to removal of outliers through radial MR analysis³⁸, and we observed no evidence for reverse causality (Supplementary Tables 28 and 29). Importantly, ascertainment bias can lead to the selection of more highly educated control participants³⁹ compared to patients with ALS who are mostly ascertained through the clinic. In line with control participants having higher education, MR analyses indicated a negative effect for years of schooling on

262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278

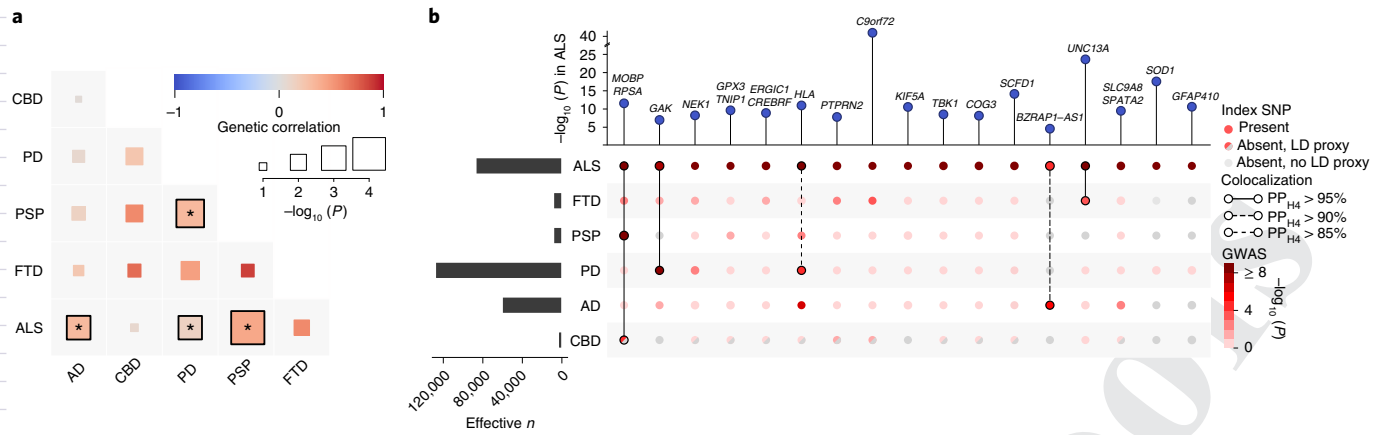


Fig. 3 | Shared genetic risk between ALS and neurodegenerative diseases. **a**, Genetic correlation analysis. Genetic correlation was estimated with LDSC between each pair of neurodegenerative diseases (ALS, AD, CBD, PD, PSP and FTD). Correlations marked with an asterisk reached nominal statistical significance ($P_{ALS,AD} = 0.01$, $P_{ALS,PD} = 0.01$, $P_{ALS,PSP} = 0.0001$, $P_{PSP,PD} = 0.002$). **b**, SNP associations of ALS lead SNPs or LD proxies in neurodegenerative diseases. The association with ALS is shown at the top. Effective sample size is shown on the left. Posterior probabilities of the same causal SNP affecting two diseases were estimated through colocalization analysis and are highlighted as connections.

279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313

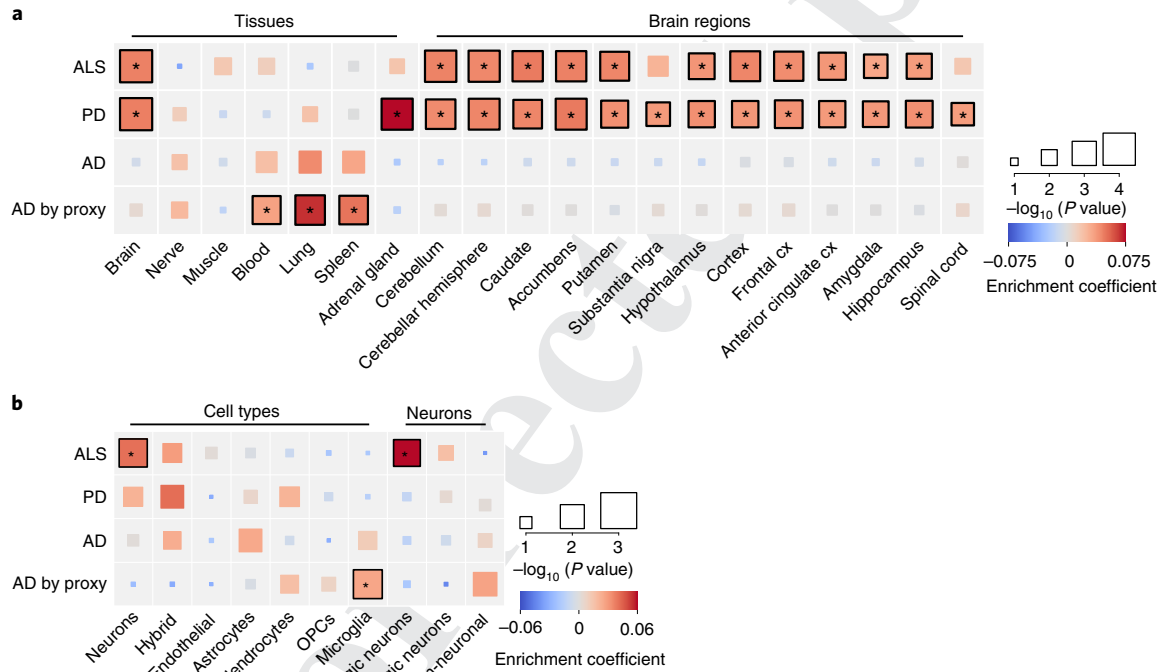


Fig. 4 | Tissue and cell type enrichment analysis. **a**, Enrichment of tissues and brain regions included in GTEx version 8 illustrates a brain-specific enrichment pattern in ALS, similar to that in PD but contrasting with that in AD. Tissues and brain regions displayed are those significantly enriched in ALS or PD, tissues previously reported in AD and tissues of specific interest for ALS (spinal cord, tibial nerve and muscle). Color represents the enrichment coefficient, and size indicates two-sided $-\log_{10}(P)$ values of enrichment obtained by the linear regression model in the MAGMA gene property analysis. **b**, Cell type enrichment analyses indicate neuron-specific enrichment for glutamatergic neurons. In ALS, no enrichment was found for microglia or other non-neuronal cell types, contrasting with the pattern observed in AD. Color represents the enrichment coefficient, and size indicates two-sided $-\log_{10}(P)$ values of enrichment obtained by the linear regression model in the MAGMA gene property analysis. Statistically significant enrichments after correction for multiple testing over all tissues ($n = 54$), cell types ($n = 7$) and neurons ($n = 3$) with $FDR < 0.05$ are marked with an asterisk. Cx, cortex; GABA, γ -aminobutyric acid; OPCs, oligodendrocyte progenitor cells.

314
315
316
317
318
319
320
321
322
323
324
325
326
327

ALS risk (inverse-variance-weighted $P_{IVW} = 2.0 \times 10^{-4}$; Fig. 5b). As a result, years of schooling can act as a confounder for the observed risk-increasing effect of higher total cholesterol levels through

ascertainment bias. To correct for this potential confounding, we applied multivariate MR analyses including both years of schooling and total cholesterol levels. The results for total cholesterol were

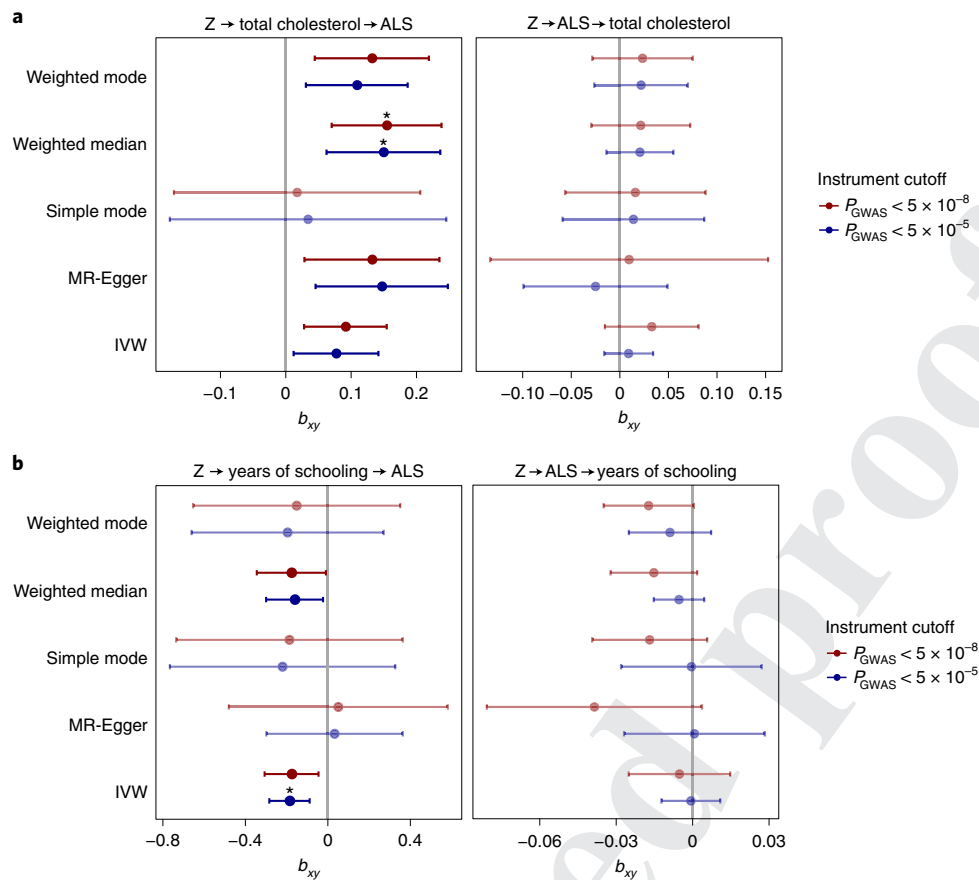


Fig. 5 | Causal inference of total cholesterol levels and years of schooling in ALS. a, MR results for ALS and total cholesterol levels. Results for the five different MR methods for two different P -value cutoffs for SNP instrument selection are presented. In total, 83 and 178 SNPs were used as instruments at cutoffs of $P < 5 \times 10^{-8}$ and $P < 5 \times 10^{-5}$, respectively. All methods show a consistent positive effect for an increased risk of ALS with higher total cholesterol levels. There is no evidence for reverse causality. Point estimates for MR are presented with error bars reflecting 95% CIs. **b**, MR results for ALS and years of schooling. In total, 306 and 681 SNPs were used as instruments at cutoffs of $P < 5 \times 10^{-8}$ and $P < 5 \times 10^{-5}$. Point estimates for MR are presented, with error bars reflecting 95% CIs. Statistically significant effects with a two-sided P value passing Bonferroni correction for multiple testing over all tested traits ($n = 22$), instrument P -value cutoffs ($n = 2$) and MR methods ($n = 5$) are marked with an asterisk (total cholesterol, $P_{\text{weighted median}} = 0.0003$ and $P_{\text{weighted median}} = 0.0007$ for cutoffs at $P < 5 \times 10^{-8}$ and $P < 5 \times 10^{-5}$, respectively; years of schooling, $P_{\text{IVW}} = 0.0002$ at the cutoff of $P < 5 \times 10^{-5}$). Here, SNP outliers were not removed for instrument selection. Z, genetic instrument; b_{xy} , estimated causal effect for an increase of 1 s.d. in genetically predicted exposure.

robust in the multivariate analyses, suggesting a causal role for total cholesterol levels on ALS susceptibility (Supplementary Table 30).

Discussion

In summary, in the largest GWAS on ALS to date including 29,612 patients with ALS and 122,656 control participants, we identified 15 risk loci contributing to ALS risk. Through in-depth analysis of these loci incorporating rare variant burden analyses and repeat expansion screens in WGS data and blood- and brain-specific eQTL and mQTL analyses, we prioritized genes in 13 of the loci. Across the spectrum of neurodegenerative diseases, we identified a genetic correlation between ALS and AD and between PD and PSP with locus-specific patterns of shared genetic risk across all neurodegenerative diseases. Colocalization analysis identified two additional loci, *GAK* and *BZRAP1-AS1*, with a high posterior probability of shared causal variants between ALS and PD and between ALS and AD, respectively. We found glutamatergic neurons as the most enriched cell type in the brain, and brain-specific coexpression network enrichment analyses indicated a role for vesicle-mediated transport and autophagy in ALS. Finally, causal inference of previously described risk factors provides evidence for high total cholesterol levels as a causal risk factor for ALS.

The cross-ancestry comparison illustrated similarities in the genetic risk factors for ALS in European and East Asian ancestries, providing an argument for cross-ancestry studies and to further expand ALS GWASs in non-European populations. It is important to note that three loci including those that harbor low-frequency variants (*KIF5A*, *SOD1* and *CFAP410*) were not included in the East Asian GWAS due to their low MAFs. Therefore, the shared genetic risk might not extend to rare genetic variation, for which population-specific frequencies have been observed even within Europe.

The multi-layered gene-prioritization analyses highlighted four different classes of genome-wide significant loci in ALS. First, the sample size of this GWAS combined with accurate imputation of low-frequency variants directly identified rare coding variants that increase ALS risk. These include the known p.D90A mutation in *SOD1* (MAF=0.006) as well as rare variants in *KIF5A* (MAF=0.016) and *CFAP410* (MAF=0.012) for which, after their identification through GWAS, experimental work confirmed their direct role in ALS pathophysiology^{11,28,40}. Second, we confirmed that the pathogenic *C9orf72* repeat expansion is tagged by genome-wide significant GWAS SNPs and that no residual signal is left by conditioning the SNP on the repeat expansion. Although more repeat

expansions are known to affect ALS risk, we found no similar loci for which the SNPs tag a highly pathogenic repeat expansion. This suggests that highly pathogenic repeat expansions on a stable haplotype are merely the exception rather than the rule in ALS. Third, common and rare variant association signals can converge on the same gene as observed for *NEK1* and *TBK1*, consistent with observations for other traits and diseases^{41–43}. We show that these signals are conditionally independent and that the common variants act on the same gene through regulatory effects as eQTL or mQTL. Fourth, we find evidence for regulatory effects of ALS-associated SNPs that act as eQTL or mQTL. These locus-specific architectures illustrate the complexity of ALS-associated GWAS loci for which not one solution fits all, but instead a multi-layered approach to prioritize genes is warranted.

In addition, we find locus-specific patterns of shared effects across neurodegenerative diseases. The *MOBP* locus has previously been identified in PSP and ALS, and here we show that indeed both diseases as well as CBD are likely to share the same causal variant in this locus. The same is true for *UNC13A* and *C9orf72* with FTD and mndFTD, respectively. The colocalization analysis with PD identified a shared causal variant in the *GAK* locus, which was not found in the ALS GWAS alone. Furthermore, the *BZRAP1-AS1* locus harbors SNPs associated with ALS and AD risk. Although this locus was not significant in either of the GWASs, a larger GWAS including AD-by-proxy cases confirmed this as a risk locus for AD. This illustrates the power of cross-disorder analyses to leverage the shared genetic risk of neurodegenerative diseases.

We aimed to clarify the role of neuron-specific pathology in ALS susceptibility as opposed to non-cell-autonomous pathology through detailed cell type enrichment analyses. Previous experiments have illustrated multiple lines of evidence for non-cell-autonomous pathology in microglia, astrocytes and oligodendrocytes, which ultimately leads to neurodegeneration in ALS^{44–46}. These experiments have shown that non-cell-autonomous processes, such as neuroinflammation, mainly act as modifiers of disease in *SOD1* models of ALS^{45,46}. Here, we show that genes within loci associated with ALS susceptibility are specifically expressed in (glutamatergic) neurons. This provides evidence for neuron-specific pathology as a driver of ALS susceptibility, which is in stark contrast to the signal of inflammation-associated tissues and cell types in AD and multiple sclerosis. It also shows that disease susceptibility and disease modification can be distinct processes, which is supported by our finding that most genetic susceptibility factors do not have a strong effect on survival. This motivates future large-scale genetic studies on modifiers of ALS progression, as these can be targets for potential new treatments for ALS as well.

The subsequent functional enrichment analyses identified that membrane trafficking, Golgi-to-ER trafficking and autophagy were enriched for genes within ALS-associated loci. These terms and their related gene ontology terms of biological processes are all related to autophagy and degradation of (misfolded) proteins. This corroborates the central hypothesis of impaired protein degradation leading to aberrant protein aggregation in neurons, which is the pathological hallmark of ALS. Our results suggest that this is a central mechanism in ALS even in the absence of rare known mutations in genes directly involved in these biological processes such as *TARDBP*, *FUS*, *UBQLN2* and *OPTN*⁴⁷.

Based on observational studies and MR analyses, conflicting evidence exists for lipid levels including cholesterol as a risk factor for ALS^{48–50}. Potential selection bias, reverse causality and the subtype of cholesterol studied challenge the interpretation of these results. Here, we provided support for a causal relationship between high total cholesterol levels and ALS independent of educational attainment and ruling out reverse orientation of the MR effect. The total cholesterol effects were consistent across the different MR methods tested, indicating that this finding is robust to violation of

the ‘no horizontal pleiotropy’ assumption. This is in line with our accompanying study showing methylation changes associated with increased cholesterol levels in ALS³⁶. We do not find a clear pattern for either low-density lipoprotein (LDL) or high-density lipoprotein (HDL) cholesterol subtypes in relation to ALS risk. While cholesterol levels are closely related to cardiovascular risk, the association between cardiovascular risk and ALS risk remains controversial with conflicting reports^{3,48,51}. Interestingly, recent work has shown that lipid metabolism and autophagy are closely related⁵², which brings the results of our pathway analyses and MR together. Both in vitro and in vivo experiments have shown that autophagy regulates lipid homeostasis through lipolysis and that impaired autophagy increases triglyceride and cholesterol levels. Conversely, high lipid levels were shown to impair autophagy⁵². Further studies on the effect of high cholesterol levels and protein degradation through autophagy illustrate that high cholesterol levels decrease the fusogenic ability of autophagic vesicles through decreased function of soluble *N*-ethylmaleimide-sensitive factor-attachment protein receptor (SNARE)^{53,54} and lead to increased protein aggregation due to impaired autophagy in mouse models of AD⁵⁵. Therefore, the risk-increasing effect of cholesterol on ALS might be mediated through impaired autophagy.

In conclusion, our GWAS identifies 15 risk loci in ALS and illustrates locus-specific interplay between common and rare genetic variation that helps to prioritize genes for future follow-up studies. We show a causal role for cholesterol, which can be linked to impaired autophagy as common denominators of neuron-specific pathology that drive ALS susceptibility and serve as potential targets for therapeutic strategies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-021-00973-1>.

Received: 12 March 2021; Accepted: 18 October 2021;

References

- van Es, M. A. et al. Amyotrophic lateral sclerosis. *Lancet* **390**, 2084–2098 (2017).
- Al-Chalabi, A., van den Berg, L. H. & Veldink, J. H. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat. Rev. Neurol.* **13**, 96–104 (2017).
- Trabjerg, B. B. et al. ALS in Danish registries: heritability and links to psychiatric and cardiovascular disorders. *Neurol. Genet.* **6**, e398 (2020).
- Ryan, M., Heverin, M., McLaughlin, R. L. & Hardiman, O. Lifetime risk and heritability of amyotrophic lateral sclerosis. *JAMA Neurol.* **76**, 1367–1374 (2019).
- Byrne, S., Elamin, M., Bede, P. & Hardiman, O. Absence of consensus in diagnostic criteria for familial neurodegenerative diseases. *J. Neurol. Neurosurg. Psychiatry* **83**, 365–367 (2012).
- Cirulli, E. T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441 (2015).
- Freischmidt, A. et al. Haploinsufficiency of *TBK1* causes familial ALS and fronto-temporal dementia. *Nat. Neurosci.* **18**, 631–636 (2015).
- Kenna, K. P. et al. *NEK1* variants confer susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.* **48**, 1037–1042 (2016).
- Brenner, D. et al. *NEK1* mutations in familial amyotrophic lateral sclerosis. *Brain* **139**, e28 (2016).
- Majounie, E. et al. Frequency of the *C9orf72* hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* **11**, 323–330 (2012).
- Nicolas, A. et al. Genome-wide analyses identify *KIF5A* as a novel ALS gene. *Neuron* **97**, 1268–1283 (2018).
- van Es, M. A. et al. Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41**, 1083–1087 (2009).

- 460 13. Laaksovirta, H. et al. Chromosome 9p21 in amyotrophic lateral sclerosis in
461 Finland: a genome-wide association study. *Lancet Neurol.* **9**, 978–985 (2010).
- 462 14. van Rheenen, W. et al. Genome-wide association analyses identify new risk
463 variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat.*
464 *Genet.* **48**, 1043–1048 (2016).
- 465 15. Benyamin, B. et al. Cross-ethnic meta-analysis identifies association of the
466 *GPX3-TNIP1* locus with amyotrophic lateral sclerosis. *Nat. Commun.* **8**,
467 611 (2017).
- 468 16. Nakamura, R. et al. A multi-ethnic meta-analysis identifies novel genes,
469 including *ACSL5*, associated with amyotrophic lateral sclerosis. *Commun.*
470 *Biol.* **3**, 526 (2020).
- 471 17. DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in
472 noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and
473 ALS. *Neuron* **72**, 245–256 (2011).
- 474 18. Renton, A. E. et al. A hexanucleotide repeat expansion in *C9ORF72* is the
475 cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
- 476 19. Diekstra, F. P. et al. *C9orf72* and *UNC13A* are shared risk loci for
477 amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide
478 meta-analysis. *Ann. Neurol.* **76**, 120–133 (2014).
- 479 20. Chen, J. A. et al. Joint genome-wide association study of progressive
480 supranuclear palsy identifies novel susceptibility loci and genetic correlation
481 to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 41 (2018).
- 482 21. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype
483 imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- 484 22. Iacoangeli, A. et al. Genome-wide meta-analysis finds the *ACSL5-ZDHHC6*
485 locus is associated with ALS and links weight loss to the disease genetics.
486 *Cell Rep.* **33**, 108323 (2020).
- 487 23. Vösa, U. et al. Large-scale *cis*- and *trans*-eQTL analyses identify thousands
488 of genetic loci and polygenic scores that regulate blood gene expression.
489 *Nat. Genet.* **53**, 1300–1310 (2021).
- 490 24. de Klein, N. et al. Brain expression quantitative trait locus and network
491 analysis reveals downstream effects and putative drivers for brain-related
492 diseases. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.01.433439>
493 (2021).
- 494 25. Pidsley, R. et al. Critical evaluation of the Illumina MethylationEPIC
495 BeadChip microarray for whole-genome DNA methylation profiling.
496 *Genome Biol.* **17**, 208 (2016).
- 497 26. Shireby, G. L. et al. Recalibrating the epigenetic clock: implications for
498 assessing biological age in the human cortex. *Brain* **143**, 3763–3775 (2020).
- 499 27. Hannon, E. et al. An integrated genetic–epigenetic analysis of
500 schizophrenia: evidence for co-localization of genetic associations and
501 differential DNA methylation. *Genome Biol.* **17**, 176 (2016).
- 502 28. Fang, X. et al. The NEK1 interactor, C21ORF2, is required for efficient
503 DNA damage repair. *Acta Biochim. Biophys. Sin.* **47**, 834–841 (2015).
- 504 29. Brown, A.-L. et al. Common ALS/FTD risk variants in *UNC13A* exacerbate
505 its cryptic splicing and loss upon TDP-43 mislocalization. Preprint at
506 *bioRxiv* <https://doi.org/10.1101/2021.04.02.438170> (2021).
- 507 30. Ma, X. R. et al. TDP-43 represses cryptic exon inclusion in FTD/ALS gene
508 *UNC13A*. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.02.438213>
509 (2021).
- 510 31. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and
511 functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**,
512 404–413 (2019).
- 513 32. Leeuw, C. A., de Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA:
514 generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**,
515 e1004219 (2015).
- 516 33. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional
517 mapping and annotation of genetic associations with FUMA. *Nat. Commun.*
518 **8**, 1826 (2017).
- 519 34. Watanabe, K., Umičević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P.
520 & Posthuma, D. Genetic mapping of cell type specificity for complex traits.
521 *Nat. Commun.* **10**, 3222 (2019).
- 522 35. Deelen, P. et al. Improving the diagnostic yield of exome-sequencing by
523 predicting gene–phenotype associations using large-scale gene expression
524 analysis. *Nat. Commun.* **10**, 2837 (2019).
- 525 36. Hop, P. J. et al. Genome-wide study of DNA methylation in amyotrophic
lateral sclerosis identifies differentially methylated loci and implicates
metabolic, inflammatory and cholesterol pathways. Preprint at *medRxiv*
<https://doi.org/10.1101/2021.03.12.21253115> (2021).
37. Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian
randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*
362, k601 (2018).
38. Bowden, J. et al. Improving the visualization, interpretation and analysis of
two-sample summary data Mendelian randomization via the radial plot and
radial regression. *Int. J. Epidemiol.* **47**, 1264–1278 (2018).
39. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G.
Collider scope: when selection bias can substantially influence observed
associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).
40. Watanabe, Y. et al. An amyotrophic lateral sclerosis-associated mutant of
C21ORF2 is stabilized by NEK1-mediated hyperphosphorylation and the
inability to bind FBXO3. *iScience* **23**, 101491 (2020).
41. Wood, A. R. et al. Defining the role of common variation in the genomic
and biological architecture of adult human height. *Nat. Genet.* **46**,
1173–1186 (2014).
42. Luo, Y. et al. Exploring the genetic architecture of inflammatory bowel
disease by whole-genome sequencing identifies association at *ADCY7*. *Nat.*
Genet. **49**, 186–192 (2017).
43. Kathiresan, S. et al. Six new loci associated with blood low-density
lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides
in humans. *Nat. Genet.* **40**, 189–197 (2008).
44. Saez-Atienzar, S. et al. Genetic analysis of amyotrophic lateral
sclerosis identifies contributing pathways and cell types. *Sci. Adv.* **7**,
eabd9036 (2021).
45. Yamanaka, K. et al. Mutant SOD1 in cell types other than motor neurons
and oligodendrocytes accelerates onset of disease in ALS mice. *Proc. Natl*
Acad. Sci. USA **105**, 7594–7599 (2008).
46. Ralph, G. S. et al. Silencing mutant SOD1 using RNAi protects against
neurodegeneration and extends survival in an ALS model. *Nat. Med.* **11**,
429–433 (2005).
47. Blokhuis, A. M., Groen, E. J. N., Koppers, M., van den Berg, L. H. &
Pasterkamp, R. J. Protein aggregation in amyotrophic lateral sclerosis. *Acta*
Neuropathol. **125**, 777–794 (2013).
48. Seelen, M. et al. Prior medical conditions and the risk of amyotrophic
lateral sclerosis. *J. Neurol.* **261**, 1949–1956 (2014).
49. Bandres-Ciga, S. et al. Shared polygenic risk and causal inferences in
amyotrophic lateral sclerosis. *Ann. Neurol.* **85**, 470–481 (2019).
50. Armon, C. Smoking is a cause of ALS. High LDL-cholesterol levels?
Unsure. *Ann. Neurol.* **85**, 465–469 (2019).
51. Turner, M. R., Wotton, C., Talbot, K. & Goldacre, M. J. Cardiovascular
fitness as a risk factor for amyotrophic lateral sclerosis: indirect
evidence from record linkage study. *J. Neurol. Neurosurg. Psychiatry* **83**,
395–398 (2012).
52. Singh, R. et al. Autophagy regulates lipid metabolism. *Nature* **458**,
1131–1135 (2009).
53. Koga, H., Kaushik, S. & Cuervo, A. M. Altered lipid content inhibits
autophagic vesicular fusion. *FASEB J.* **24**, 3052–3065 (2010).
54. Fraldi, A. et al. Lysosomal fusion and SNARE function are impaired by
cholesterol accumulation in lysosomal storage disorders. *EMBO J.* **29**,
3607–3620 (2010).
55. Barbero-Camps, E. et al. Cholesterol impairs autophagy-mediated
clearance of amyloid β while promoting its secretion. *Autophagy* **14**,
1129–1154 (2018).





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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Wouter **van Rheenen** ^{1,203} , Rick A. A. **van der Spek** ^{1,203}, Mark K. **Bakker** ^{1,203}, Joke J. F. A. **van Vugt** ¹, Paul J. **Hop** ¹, Ramona A. J. **Zwamborn** ¹, Niek **de Klein** ² , Harm-Jan **Westra** ² , Olivier B. **Bakker** ², Patrick **Deelen** ^{2,3} , Gemma **Shireby** ⁴, Eilis **Hannon** ⁴ , Matthieu **Moisse** ^{5,6,7}, Denis **Baird** ^{8,9}, Restuadi **Restuadi** ¹⁰, Egor **Dolzhenko** ¹¹, Annelot M. **Dekker** ¹, Klara **Gawor** ¹ , Henk-Jan **Westeneng** ¹, Gijs H. P. **Tazelaar** ¹, Kristel R. **van Eijk** ¹, Maarten **Kooyman** ¹ , Ross P. **Byrne** ¹² , Mark **Doherty** ¹², Mark **Heverin** ¹³, Ahmad Al **Khleifat** ¹⁴ , Alfredo **Iacoangeli** ^{14,15,16}, Aleksey **Shatunov** ¹⁴, Nicola **Ticozzi** ^{17,18} , Johnathan **Cooper-Knock** ¹⁹, Bradley N. **Smith** ¹⁴, Marta **Gromicho** ²⁰ , Siddharthan **Chandran** ^{21,22}, Suvankar **Pal** ^{21,22}, Karen E. **Morrison** ²³ , Pamela J. **Shaw** ¹⁹, John **Hardy** ²⁴, Richard W. **Orrell** ²⁵, Michael **Sendtner** ²⁶, Thomas **Meyer** ²⁷ , Nazli **Başak** ²⁸, Anneke J. **van der Kool** ²⁹, Antonia **Ratti** ^{17,30}, Isabella **Fogh** ¹⁴, Cinzia **Gellera** ³¹, Giuseppe Lauria **Pinter** ^{32,33}, Stefania **Corti** ^{18,34}, Cristina **Cereda** ³⁵ , Daisy **Sproviero** ³⁵, Sandra **D'Alfonso** ³⁶, Gianni **Soraru** ³⁷, Gabriele **Siciliano** ³⁸, Massimiliano **Filosto** ³⁹, Alessandro **Padovani** ³⁹, Adriano **Chiò** ^{40,41}, Andrea **Calvo** ^{40,41} , Cristina **Moglia** ^{40,41}, Maura **Brunetti** ⁴⁰, Antonio **Canosa** ^{40,41} , Maurizio **Grassano** ⁴⁰, Ettore **Beghi** ⁴², Elisabetta **Pupillo** ⁴², Giancarlo **Logroscino** ⁴³, Beatrice **Nefussy** ⁴⁴, Alma **Osmanovic** ^{45,46}, Angelica **Nordin** ⁴⁷ , Yossef **Lerner** ^{48,49}, Michal **Zabari** ^{48,49}, Marc **Gotkine** ^{48,49} , Robert H. **Baloh** ^{50,51}, Shaughn **Bell** ^{50,51}, Patrick **Vourc'h** ^{52,53}, Philippe **Corcia** ^{53,54}, Philippe **Couratier** ^{55,56}, Stéphanie **Millecamps** ⁵⁷, Vincent **Meininger** ⁵⁸, François **Salachas** ^{57,59}, Jesus S. Mora **Pardina** ⁶⁰, Abdelilah **Assialioui** ⁶¹, Ricardo **Rojas-García** ⁶² , Patrick **Dion** ⁶³, Jay P. **Ross** ^{63,64} , Albert C. **Ludolph** ⁶⁵, Jochen H. **Weishaupt** ⁶⁶, David **Brenner** ⁶⁶, Axel **Freischmidt** ^{65,67}, Gilbert **Bensimon** ^{68,69,70,71}, Alexis **Brice** ^{72,73,74}, Alexandra **Dürr** ⁷⁵, Christine A. M. **Payan** ⁶⁸, Safa **Saker-Delye** ⁷⁶, Nicholas **Wood** ⁷⁷ , Simon **Topp** ¹⁴ , Rosa **Rademakers** ⁷⁸, Lukas **Tittmann** ⁷⁹, Wolfgang **Lieb** ⁷⁹ , Andre **Franke** ⁸⁰ , Stephan **Ripke** ^{81,82,83}, Alice **Braun** ⁸³ , Julia **Kraft** ⁸³ , David C. **Whiteman** ⁸⁴ , Catherine M. **Olsen** ⁸⁴ , Andre G. **Uitterlinden** ^{85,86} , Albert **Hofman** ⁸⁶, Marcella **Rietschel** ^{87,88} , Sven **Cichon** ^{89,90,91,92}, Markus M. **Nöthen** ^{89,90} , Philippe **Amouyel** ⁹³ , SLALOM Consortium*, PARALS Consortium*, SLAGEN Consortium*, SLAP Consortium*, Bryan J. **Traynor** ^{94,95}, Andrew B. **Singleton** ⁹⁶, Miguel **Mitne Neto** ⁹⁷, Ruben J. **Cauchi** ⁹⁸ , Roel A. **Ophoff** ^{99,100,101}, Martina **Wiedau-Pazos** ¹⁰², Catherine **Lomen-Hoerth** ¹⁰³, Vivianna M. **van Deerlin** ¹⁰⁴ , Julian **Grosskreutz** ^{105,106} , Annekathrin **Rödiger** ¹⁰⁵, Nayana **Gaur** ¹⁰⁵, Alexander **Jörk** ¹⁰⁵, Tabea **Barthel** ¹⁰⁵, Erik **Theele** ¹⁰⁵, Benjamin **Ilse** ¹⁰⁵, Beatrice **Stubendorff** ¹⁰⁵, Otto W. **Witte** ¹⁰⁵, Robert **Steinbach** ¹⁰⁵ , Christian A. **Hübner** ¹⁰⁷, Caroline **Graff** ¹⁰⁸, Lev **Brylev** ^{109,110,111}, Vera **Fominykh** ^{109,111} , Vera **Demeshonok** ¹¹², Anastasia **Ataulina** ¹⁰⁹, Boris **Rogelj** ^{113,114,115} , Blaž **Koritnik** ¹¹⁶ , Janez **Zidar** ¹¹⁶, Metka **Ravnik-Glavac** ¹¹⁷, Damjan **Glavac** ¹¹⁸, Zorica **Stevic** ¹¹⁹, Vivian **Drory** ^{119,120}, Monica **Povedano** ⁶¹, Ian P. **Blair** ¹²¹, Matthew C. **Kiernan** ¹²², Beben **Benyamin** ^{10,123}, Robert D. **Henderson** ^{124,125}, Sarah **Furlong** ¹²¹, Susan **Mathers** ¹²⁶ , Pamela A. **McCombe** ^{125,127}, Merrilee **Needham** ^{128,129,130}, Shyuan T. **Ngo** ^{124,125,127} , Garth A. **Nicholson** ^{121,131,132}, Roger **Pamphlett** ¹³³ , Dominic B. **Rowe** ¹²¹ , Frederik J. **Steyn** ^{125,134}, Kelly L. **Williams** ¹²¹ , Karen A. **Mather** ^{135,136} , Perminder S. **Sachdev** ^{135,137} , Anjali K. **Henders** ¹⁰, Leanne **Wallace** ¹⁰, Mamede **de Carvalho** ²⁰, Susana **Pinto** ²⁰, Susanne **Petri** ⁴⁵, Markus **Weber** ¹³⁸, Guy A. **Rouleau** ⁶³ , Vincenzo **Silani** ^{17,18} , Charles **Curtis** ¹³⁹, Gerome **Breen** ^{140,141} , Jonathan **Glass** ¹⁴² , Robert H. **Brown** ¹⁴³ , John E. **Landers** ¹⁴³ , Christopher E. **Shaw** ¹⁴, Peter M. **Andersen** ⁴⁷ , Ewout J. N. **Groen** ¹ , Michael A. **van Es** ¹ , R. Jeroen **Pasterkamp** ¹⁴⁴ , Dongsheng **Fan** ¹⁴⁵, Fleur C. **Garton** ¹⁰, Allan F. **McRae** ¹⁰, George Davey **Smith** ^{9,146} , Tom R. **Gaunt** ^{9,146} , Michael A. **Eberle** ¹¹ , Jonathan **Mill** ⁴ , Russell L. **McLaughlin** ¹² , Orla **Hardiman** ¹³, Kevin P. **Kenna** ^{1,144}, Naomi R. **Wray** ^{10,127} , Ellen **Tsai** ⁸ , Heiko **Runz** ⁸, Lude **Franke** ² 

Ammar Al-Chalabi ^{14,147}, **Philip Van Damme** ^{5,6,7}, **Leonard H. van den Berg**^{1,204} and **Jan H. Veldink** ^{1,204} 

¹Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. ²Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. ³Department of Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. ⁴University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter, UK. ⁵Department of Neurosciences, Experimental Neurology and Leuven Brain Institute (LBI), KU Leuven—University of Leuven, Leuven, Belgium. ⁶Laboratory of Neurobiology, VIB, Center for Brain & Disease Research, Leuven, Belgium. ⁷Department of Neurology, University Hospitals Leuven, Leuven, Belgium. ⁸Translational Biology, Biogen, Boston, MA, USA. ⁹MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, University of Bristol, Bristol, UK. ¹⁰Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. ¹¹Illumina, San Diego, CA, USA. ¹²Complex Trait Genomics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. ¹³Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Ireland. ¹⁴Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ¹⁵Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ¹⁶National Institute for Health Research Biomedical Research Centre and Dementia Unit, South London and Maudsley NHS Foundation Trust and King's College London, London, UK. ¹⁷Department of Neurology, Stroke Unit and Laboratory of Neuroscience, Istituto Auxologico Italiano IRCCS, Milan, Italy. ¹⁸Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milan, Italy. ¹⁹Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK. ²⁰Instituto de Fisiologia, Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal. ²¹Euan MacDonald Centre for Motor Neurone Disease Research, Edinburgh, UK. ²²Centre for Neuroregeneration and Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK. ²³School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK. ²⁴Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK. ²⁵Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, UK. ²⁶Institute of Clinical Neurobiology, University Hospital Würzburg, Würzburg, Germany. ²⁷Charité University Hospital, Humboldt University, Berlin, Germany. ²⁸Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey. ²⁹Department of Neurology, Academic Medical Center, Amsterdam, the Netherlands. ³⁰Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, Milan, Italy. ³¹Unit of Medical Genetics and Neurogenetics, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ³²3rd Neurology Unit, Motor Neuron Diseases Center, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ³³'L. Sacco' Department of Biomedical and Clinical Sciences, Università degli Studi di Milano, Milan, Italy. ³⁴Neurology Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. ³⁵Genomic and Post-Genomic Center, IRCCS Mondino Foundation, Pavia, Italy. ³⁶Department of Health Sciences, University of Eastern Piedmont, Novara, Italy. ³⁷Department of Neurosciences, University of Padova, Padova, Italy. ³⁸Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. ³⁹Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy. ⁴⁰'Rita Levi Montalcini' Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy. ⁴¹Neurologia 1, Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Turin, Italy. ⁴²Laboratory of Neurological Diseases, Department of Neuroscience, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy. ⁴³Department of Clinical Research in Neurology, University of Bari at 'Pia Fondazione Card G. Panico' Hospital, Bari, Italy. ⁴⁴Neuromuscular Diseases Unit, Department of Neurology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. ⁴⁵Department of Neurology, Hannover Medical School, Hannover, Germany. ⁴⁶Essener Zentrum für Seltene Erkrankungen (EZSE), University Hospital Essen, Essen, Germany. ⁴⁷Department of Clinical Sciences, Neurosciences, Umeå University, Umeå, Sweden. ⁴⁸Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel. ⁴⁹Department of Neurology, the Agnes Ginges Center for Human Neurogenetics, Hadassah Medical Center, Jerusalem, Israel. ⁵⁰Center for Neural Science and Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ⁵¹Department of Neurology, Neuromuscular Division, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ⁵²Service de Biochimie et Biologie Moléculaire, CHU de Tours, Tours, France. ⁵³UMR 1253, Université de Tours, Inserm, Tours, France. ⁵⁴Centre de référence sur la SLA, CHU de Tours, Tours, France. ⁵⁵Centre de référence sur la SLA, CHRU de Limoges, Limoges, France. ⁵⁶UMR 1094, Université de Limoges, Inserm, Limoges, France. ⁵⁷ICM, Institut du Cerveau, Inserm, CNRS, Sorbonne Université, Hôpital Pitié-Salpêtrière, Paris, France. ⁵⁸Hôpital des Peupliers, Ramsay Générale de Santé, Paris, France. ⁵⁹Département de Neurologie, Centre de référence SLA Ile de France, Hôpital de la Pitié-Salpêtrière, AP-HP, Paris, France. ⁶⁰ALS Unit, Hospital San Rafael, Madrid, Spain. ⁶¹Functional Unit of Amyotrophic Lateral Sclerosis (UFELA), Service of Neurology, Bellvitge University Hospital, L'Hospitalet de Llobregat, Barcelona, Spain. ⁶²MND Clinic, Neurology Department, Hospital de la Santa Creu i Sant Pau de Barcelona, Universitat Autònoma de Barcelona, Barcelona, Spain. ⁶³Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada. ⁶⁴Department of Human Genetics, McGill University, Montreal, Quebec, Canada. ⁶⁵Department of Neurology, Ulm University, Ulm, Germany. ⁶⁶Division of Neurodegeneration, Department of Neurology, University Medicine Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ⁶⁷German Center for Neurodegenerative Diseases (DZNE) Ulm, Ulm, Germany. ⁶⁸Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, UPMC Pharmacologie, AP-HP, Paris, France. ⁶⁹Pharmacologie Sorbonne Université, Paris, France. ⁷⁰Institut du Cerveau, Paris Brain Institute ICM, Paris, France. ⁷¹Laboratoire de Biostatistique, Epidémiologie Clinique, Santé Publique Innovation et Méthodologie (BESPIIM), CHU-Nîmes, Nîmes, France. ⁷²INSERM U289, Hôpital Salpêtrière, AP-HP, Paris, France. ⁷³Département de Génétique, Cytogénétique et Embryologie, Hôpital Salpêtrière, AP-HP, Paris, France. ⁷⁴Fédération de Neurologie, Hôpital Salpêtrière, AP-HP, Paris, France. ⁷⁵Department of Medical Genetics, l'Institut du Cerveau et de la Moelle Épinrière, Hôpital Salpêtrière, Paris, France. ⁷⁶Genethon, CNRS UMR, Evry, France. ⁷⁷Department of Neurogenetics, UCL Institute of Neurology, Queen Square, London, UK. ⁷⁸Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, FL, USA. ⁷⁹Popgen Biobank and Institute of Epidemiology, Christian Albrechts-University Kiel, Kiel, Germany. ⁸⁰Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. ⁸¹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. ⁸²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁸³Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin, Berlin, Germany. ⁸⁴Cancer Control Group, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia. ⁸⁵Department of Internal Medicine, Genetics Laboratory, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands. ⁸⁶Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands. ⁸⁷Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany. ⁸⁸Central Institute of Mental Health, Mannheim, Germany. ⁸⁹Institute of Human Genetics, University of Bonn, Bonn, Germany. ⁹⁰Department of Genomics, Life and Brain Center, Bonn, Germany. ⁹¹Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel, Basel, Switzerland. ⁹²Institute of Neuroscience and Medicine INM-1, Research Center Juelich, Juelich, Germany. ⁹³UMR1167—RID-AGE LabEx DISTALZ—Risk Factors and Molecular Determinants of Aging-Related Diseases, University of Lille, Inserm, Centre Hospitalier of the University of Lille, Institut Pasteur de Lille, Lille, France. ⁹⁴Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD, USA. ⁹⁵Department of Neurology, Johns Hopkins University, Baltimore, MD, USA. ⁹⁶Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD, USA. ⁹⁷Universidade de São Paulo, São Paulo, Brazil. ⁹⁸Centre for Molecular Medicine and Biobanking and

526 Department of Physiology and Biochemistry, Faculty of Medicine and Surgery, University of Malta, Msida, Malta. ⁹⁹University Medical Center Utrecht,
 527 Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, Utrecht, the Netherlands. ¹⁰⁰Department of Human Genetics, David Geffen School of
 528 Medicine, University of California, Los Angeles, CA, USA. ¹⁰¹Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior,
 529 University of California, Los Angeles, CA, USA. ¹⁰²Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA,
 530 USA. ¹⁰³Department of Neurology, University of California, San Francisco, CA, USA. ¹⁰⁴Center for Neurodegenerative Disease Research, Perelman School of
 531 Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ¹⁰⁵Hans Berger Department of Neurology, Jena University Hospital, Jena, Germany.
 532 ¹⁰⁶Precision Neurology Unit, Department of Neurology, University Hospital Schleswig-Holstein, University of Luebeck, Luebeck, Germany. ¹⁰⁷Institute of
 533 Human Genetics, Jena University Hospital, Jena, Germany. ¹⁰⁸Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm,
 534 Sweden. ¹⁰⁹Department of Neurology, Bujanov Moscow Clinical Hospital, Moscow, Russia. ¹¹⁰Moscow Research and Clinical Center for Neuropsychiatry of
 535 the Healthcare Department, Moscow, Russia. ¹¹¹Department of Functional Biochemistry of the Nervous System, Institute of Higher Nervous Activity and
 536 Neurophysiology Russian Academy of Sciences, Moscow, Russia. ¹¹²ALS-Care Center, 'GAOORDI', Medical Clinic of the St. Petersburg, St. Petersburg,
 537 Russia. ¹¹³Department of Biotechnology, Jožef Stefan Institute, Ljubljana, Slovenia. ¹¹⁴Biomedical Research Institute BRIS, Ljubljana, Slovenia. ¹¹⁵Faculty of
 538 Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia. ¹¹⁶Ljubljana ALS Centre, Institute of Clinical Neurophysiology, University
 539 Medical Centre Ljubljana, Ljubljana, Slovenia. ¹¹⁷Institute of Biochemistry and Molecular Genetics, Faculty of Medicine, University of Ljubljana, Ljubljana,
 540 Slovenia. ¹¹⁸Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. ¹¹⁹Clinic of Neurology,
 541 Clinical Center of Serbia, School of Medicine, University of Belgrade, Belgrade, Serbia. ¹²⁰Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.
 542 ¹²¹Centre for Motor Neuron Disease Research, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, New South Wales,
 543 Australia. ¹²²Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. ¹²³Australian Centre for Precision Health and Allied Health
 544 and Human Performance, University of South Australia, Adelaide, South Australia, Australia. ¹²⁴Centre for Clinical Research, University of Queensland,
 545 Brisbane, Queensland, Australia. ¹²⁵Department of Neurology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia. ¹²⁶Calvary Health
 546 Care Bethlehem, Parkdale, Victoria, Australia. ¹²⁷Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. ¹²⁸Fiona Stanley
 547 Hospital, Perth, Western Australia, Australia. ¹²⁹Notre Dame University, Fremantle, Western Australia, Australia. ¹³⁰Institute for Immunology and Infectious
 548 Diseases, Murdoch University, Perth, Western Australia, Australia. ¹³¹Northcott Neuroscience Laboratory, ANZAC Research Institute, Concord, New South
 549 Wales, Australia. ¹³²Molecular Medicine Laboratory, Concord Repatriation General Hospital, Concord, New South Wales, Australia. ¹³³Discipline of
 550 Pathology and Department of Neuropathology, Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. ¹³⁴The School of
 551 Biomedical Sciences, Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia. ¹³⁵Centre for Healthy Brain Ageing, School of
 552 Psychiatry, University of New South Wales, Sydney, New South Wales, Australia. ¹³⁶Neuroscience Research Australia Institute, Randwick, New South
 553 Wales, Australia. ¹³⁷Neuropsychiatric Institute, the Prince of Wales Hospital, UNSW, Randwick, New South Wales, Australia. ¹³⁸Neuromuscular Diseases
 554 Unit/ALS Clinic, Kantonsspital St. Gallen, St. Gallen, Switzerland. ¹³⁹MRC Social, Genetic and Developmental Psychiatry Centre, King's College London,
 555 London, UK. ¹⁴⁰IoPPN Genomics and Biomarker Core, Translational Genetics Group, MRC Social, Genetic and Developmental Psychiatry Centre, King's
 556 College London, London, UK. ¹⁴¹NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology and
 557 Neuroscience, King's College London, London, UK. ¹⁴²Department Neurology, Emory University School of Medicine, Atlanta, GA, USA. ¹⁴³Department of
 558 Neurology, University of Massachusetts Medical School, Worcester, MA, USA. ¹⁴⁴Department of Translational Neuroscience, UMC Utrecht Brain Center,
 559 University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. ¹⁴⁵Department of Neurology, Third Hospital, Peking University, Beijing,
 560 China. ¹⁴⁶Population Health Science, Bristol Medical School, Bristol, UK. ¹⁴⁷King's College Hospital, London, UK. ²⁰³These authors contributed equally:
 561 Wouter van Rheenen, Rick A. A. van der Spek, Mark K. Bakker. ²⁰⁴These authors jointly supervised this work: Leonard H. van den Berg, Jan H. Veldink. *Lists
 562 of authors and their affiliations appear at the end of the paper. ^{83e}e-mail: w.vanrheenen-2@umcutrecht.nl; j.h.veldink@umcutrecht.nl

563 SLALOM Consortium

566 **Ettore Beghi⁴², Elisabetta Pupillo⁴², Giancarlo Comi¹⁴⁸, Nilo Riva¹⁴⁸, Christian Lunetta¹⁴⁹,**
 567 **Francesca Gerardi¹⁴⁹, Maria Sofia Cotelli¹⁵⁰, Fabrizio Rinaldi¹⁵⁰, Luca Chiveri¹⁵¹, Maria Cristina Guaita¹⁵²,**
 568 **Patrizia Perrone¹⁵², Mauro Ceroni¹⁵³, Luca Diamanti¹⁵³, Carlo Ferrarese¹⁵⁴, Lucio Tremolizzo¹⁵⁴,**
 569 **Maria Luisa Delodovici¹⁵⁵ and Giorgio Bono¹⁵⁵**

573 ¹⁴⁸IRCCS San Raffaele Hospital, Milan, Italy. ¹⁴⁹NEMO Clinical Center, Serena Onlus Foundation, Niguarda Ca' Granda Hospital, Milan, Italy. ¹⁵⁰Civil Hospital
 574 of Brescia, Brescia, Italy. ¹⁵¹Ospedale Valduce, Como, Italy. ¹⁵²AO Ospedale Civile di Legnano, Legnano, Italy. ¹⁵³IRCCS Istituto Neurologico Nazionale 'C.
 575 Mondino', Pavia, Italy. ¹⁵⁴AO 'San Gerardo' di Monza and University of Milano-Bicocca, Milano-Bicocca, Italy. ¹⁵⁵AO 'Ospedale di Circolo Fondazione
 576 Macchi' di Varese, Varese, Italy.

580 PARALS Consortium

582 **Adriano Chiò^{40,41}, Andrea Calvo^{40,41}, Cristina Moglia^{40,41}, Antonio Canosa^{40,41,156}, Umberto Manera⁴⁰,**
 583 **Rosario Vasta⁴⁰, Alessandro Bombaci⁴⁰, Maurizio Grassano⁴⁰, Maura Brunetti⁴⁰, Federico Casale⁴⁰,**
 584 **Giuseppe Fuda⁴⁰, Paolina Salamone⁴⁰, Barbara Iazzolino⁴⁰, Laura Peotta⁴⁰, Paolo Cugnasco⁴⁰,**
 585 **Giovanni De Marco⁴¹, Maria Claudia Torrieri⁴⁰, Francesca Palumbo⁴⁰, Salvatore Gallone⁴¹,**
 586 **Marco Barberis¹⁵⁷, Luca Sbaiz¹⁵⁷, Salvatore Gentile¹⁵⁸, Alessandro Mauro^{40,159}, Letizia Mazzini^{160,161},**
 587 **Fabiola De Marchi^{160,161}, Lucia Corrado^{161,162}, Sandra D'Alfonso^{161,162}, Antonio Bertolotto¹⁶³,**

Maurizio Gionco¹⁶⁴, Daniela Leotta¹⁶⁵, Enrico Odddenino¹⁶⁵, Daniele Imperiale¹⁶⁶, Roberto Cavallo¹⁶⁷, Pietro Pignatta¹⁶⁸, Marco De Mattei¹⁶⁹, Claudio Geda¹⁷⁰, Diego Maria Papurello¹⁷¹, Graziano Gusmaroli¹⁷², Cristoforo Comi^{173,174}, Carmelo Labate¹⁷⁵, Luigi Ruiz¹⁷⁶, Delfina Ferrandi¹⁷⁷, Eugenia Rota¹⁷⁸, Marco Aguggia¹⁷⁹, Nicoletta Di Vito¹⁷⁹, Piero Meineri¹⁸⁰, Paolo Ghiglione¹⁸¹, Nicola Launaro¹⁸², Michele Dotta¹⁸³, Alessia Di Sapio¹⁸⁴ and Guido Giardini¹⁸⁵

¹⁵⁶Neurology Unit 1U, Azienda Ospedaliero Universitaria Città della Salute e della Scienza di Torino, Turin, Italy. ¹⁵⁷Department of Medical Genetics, Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Turin, Italy. ¹⁵⁸Neurologia 3, Azienda Ospedaliero Universitaria Città della Salute e della Scienza di Torino, Turin, Italy. ¹⁵⁹Istituto Auxologico Italiano, IRCCS, Piancavallo, Italy. ¹⁶⁰Department of Neurology, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy. ¹⁶¹Azienda Ospedaliero Universitaria 'Maggiore della Carità', Novara, Italy. ¹⁶²Department of Health Sciences, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy. ¹⁶³Department of Neurology and Multiple Sclerosis Center, Azienda Ospedaliero Universitaria San Luigi, Orbassano, Italy. ¹⁶⁴Department of Neurology, Azienda Ospedaliera 'Ordine Mauriziano' di Torino, Turin, Italy. ¹⁶⁵Department of Neurology, Ospedale Martini, ASL Città di Torino, Turin, Italy. ¹⁶⁶Department of Neurology, Ospedale Maria Vittoria, ASL Città di Torino, Turin, Italy. ¹⁶⁷Department of Neurology, Ospedale San Giovanni Bosco, ASL Città di Torino, Turin, Italy. ¹⁶⁸Ospedale Humanitas Gradenigo, Turin, Italy. ¹⁶⁹Department of Neurology, Ospedale 'Santa Croce' di Moncalieri, ASL Torino 5, Moncalieri, Italy. ¹⁷⁰Department of Neurology, Ospedale Civile di Ivrea, ASL Torino 4, Ivrea, Italy. ¹⁷¹Department of Neurology, Presidio Ospedaliero di Ciriè, ASL Torino 4, Ciriè, Italy. ¹⁷²Department of Neurology, Ospedale 'Degli Infermi' di Biella, ASL Biella, Ponderano, Italy. ¹⁷³Department of Neurology, Ospedale 'Sant'Andrea' di Vercelli, ASL Vercelli, Vercelli, Italy. ¹⁷⁴Department of Clinical and Experimental Medicine, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy. ¹⁷⁵Department of Neurology, Ospedale Civile 'Edoardo Agnelli' di Pinerolo, ASL Torino 2, Pinerolo, Italy. ¹⁷⁶Department of Neurology, Azienda Ospedaliera 'Santi Antonio e Biagio' di Alessandria, Alessandria, Italy. ¹⁷⁷Department of Neurology, Ospedale 'Santo Spirito' di Casale Monferrato, ASL Alessandria, Casale Monferrato, Italy. ¹⁷⁸Department of Neurology, Ospedale 'San Giacomo' di Novi Ligure, ASL Alessandria, Novi Ligure, Italy. ¹⁷⁹Department of Neurology, Ospedale 'Cardinal Massia' di Asti, ASL Asti, Asti, Italy. ¹⁸⁰Department of Neurology, Azienda Ospedaliera 'Santa Croce e Carle' di Cuneo, Cuneo, Italy. ¹⁸¹Department of Neurology, Ospedale 'Maggiore Santissima Annunziata' di Savigliano, ASL Cuneo 1, Savigliano, Italy. ¹⁸²Department of Anesthesiology, Ospedale 'Maggiore Santissima Annunziata' di Savigliano, ASL Cuneo 1, Savigliano, Italy. ¹⁸³Department of Neurology, Ospedale 'Michele e Pietro Ferrero' di Verduno, ASL Cuneo 2, Verduno, Italy. ¹⁸⁴Department of Neurology, Ospedale 'Regina Montis Regalis' di Mondovì, ASL Cuneo 1, Aosta, Italy. ¹⁸⁵Department of Neurology, Ospedale Regionale 'Umberto Parini' di Aosta, Aosta, Italy.

SLAGEN Consortium

Vincenzo Silani^{17,18}, Nicola Ticozzi^{17,18}, Antonia Ratti^{17,30}, Isabella Fogh¹⁴, Cinzia Tiloca¹⁷, Silvia Peverelli¹⁷, Cinzia Gellera³¹, Giuseppe Lauria Pinter^{32,33}, Franco Taroni¹⁸⁶, Viviana Pensato¹⁸⁶, Barbara Castellotti¹⁸⁶, Giacomo P. Comi^{18,34}, Stefania Corti^{18,34}, Roberto Del Bo^{18,34}, Cristina Cereda³⁵, Mauro Ceroni^{187,188}, Stella Gagliardi³⁵, Sandra D'Alfonso³⁶, Lucia Corrado³⁶, Letizia Mazzini¹⁸⁹, Gianni Sorarù³⁷, Flavia Raggi³⁷, Gabriele Siciliano³⁸, Costanza Simoncini³⁸, Annalisa Lo Gerfo³⁸, Massimiliano Filosto³⁹, Maurizio Inghilleri¹⁹⁰ and Alessandra Ferlini¹⁹¹

¹⁸⁶Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ¹⁸⁷Unit of General Neurology, IRCCS Mondino Foundation, Pavia, Italy. ¹⁸⁸Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy. ¹⁸⁹ALS Center, Department of Neurology, Azienda Ospedaliero Universitaria Maggiore della Carità, Novara, Italy. ¹⁹⁰Rare Neuromuscular Diseases Centre, Department of Human Neuroscience, Sapienza University, Rome, Italy. ¹⁹¹Unit of Medical Genetics, Department of Medical Science, University of Ferrara, Ferrara, Italy.

SLAP Consortium

Giancarlo Logrosco⁴³, Ettore Beghi⁴², Isabella L. Simone¹⁹², Bruno Passarella¹⁹³, Vito Guerra¹⁹⁴, Stefano Zoccollella¹⁹⁵, Cecilia Nozzoli¹⁹³, Ciro Mundi¹⁹⁶, Maurizio Leone¹⁹⁷, Michele Zarrelli¹⁹⁸, Filippo Tamma¹⁹⁹, Francesco Valluzzi²⁰⁰, Gianluigi Calabrese²⁰¹, Giovanni Boero²⁰² and Augusto Rini¹⁹³

¹⁹²Department of Basic Medical Sciences, Neurosciences and Sense Organs, University of Bari, Bari, Italy. ¹⁹³Neurological Department, Antonio Perrino's Hospital, Brindisi, Italy. ¹⁹⁴National Institute of Digestive Diseases, IRCCS S. de Bellis Research Hospital, Castellana Grotte, Italy. ¹⁹⁵ASL Bari, San Paolo Hospital, Bari, Italy. ¹⁹⁶Department of Neuroscience, United Hospital of Foggia, Foggia, Italy. ¹⁹⁷Unit of Neurology, Department of Emergency and Critical Care, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy. ¹⁹⁸Unit of Neurology, Department of Medical Sciences, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy. ¹⁹⁹Neurology Unit, Miulli Hospital, Acquaviva delle Fonti, Italy. ²⁰⁰Unit of Neurology, 'S. Giacomo' Hospital, Bari, Italy. ²⁰¹Department of Neurology, ASL (Local Health Authority) at the 'V Fazzi' Hospital, Lecce, Italy. ²⁰²Department of Neurology, ASL (Local Health Authority) at the 'SS Annunziata' Hospital, Taranto, Italy.

592 **Methods**

593 **Genome-wide association study.** *Data description.* We obtained individual
594 genotype-level data for all individuals in the previously published GWAS of ALS in
595 European ancestries^{11,14} and publicly available control datasets including 120,971
596 controls genotyped on Illumina platforms. Additionally, 6,374 cases and 22,526
597 controls were genotyped on the Illumina OmniExpress and Illumina GSA arrays.
598 Details for each cohort are provided in Supplementary Table 1. All patients with
599 ALS were diagnosed and ascertained through specialized MND clinics where
600 they were diagnosed with ALS according to the (revised) El Escorial Criteria⁴⁶
601 by neurologists specialized in motor neuron diseases. Whole-blood samples
602 were drawn for DNA isolation, which were specifically collected for ongoing
603 case-control studies of ALS. Both cases with and without a family history for
604 ALS and/or dementia were included. Cases were not pre-screened for specific
605 ALS-related mutations. Given the late onset and relatively low lifetime risk of ALS,
606 controls were not screened for (subclinical) signs of ALS. A detailed description
607 of the ascertainment of newly genotyped cases and controls is provided in the
608 Supplementary Note. All participants gave written informed consent, and the
609 relevant local institutional review boards approved this study (Supplementary
610 Note). Cases and controls formed cohorts when they were processed in the same
611 laboratory and were genotyped in the same batch, resulting in 117 independent
612 cohorts. Summary statistics were obtained for the Asian ancestry GWAS of ALS^{15,16}
613 (Supplementary Note).

614 *GWAS quality control and imputation.* For each cohort, we first performed SNP-
615 and variant-level quality control, after which cohorts were merged into six strata
616 based on genotyping platform. Subsequent stratum-wise quality control was
617 performed, and strata were imputed up to the Haplotype Reference Consortium
618 panel (r.1.1 2016) through the Michigan Imputation Server²¹. Full quality-control
619 details are described in the Supplementary Note and Supplementary Fig. 17.
620 Numbers of individuals and variants passing each quality-control step are
621 described in Supplementary Table 2.

622 *Association testing and meta-analysis.* After quality control, a null logistic
623 mixed model was fitted using SAIGE³⁷ 0.29.1 for each stratum with principal
624 component (PC)1–PC20 as covariates. The model was fit on a set of high-quality
625 (INFO > 0.95) SNPs pruned with PLINK 1.9 ('-indep-pairwise 50 25 0.1') in a
626 leave-one-chromosome-out scheme. Subsequently, a SNP-wise logistic mixed
627 model including the saddlepoint approximation test was performed using genotype
628 dosages with SAIGE. Association statistics for all strata were combined in an IVW
629 fixed-effects meta-analysis using METAL³⁸.

630 Genomic inflation factors were calculated per stratum and for the full
631 meta-analysis. To assess any residual confounding due to population stratification
632 and artificial structure in the data, we calculated the LDSC³⁹ intercept using SNP
633 LD scores calculated in the HapMap3 CEU population.

634 *Cross-ancestry analyses.* GWAS summary statistics from two Asian ancestry
635 studies were obtained^{15,16}. These summary statistics were meta-analyzed with all
636 European ancestry data in strata as described above. To assess genetic correlation
637 for ALS in European and Asian ancestries, we used Popcorn⁶⁰ version 0.9.9. We
638 used population-specific LD scores for genetic impact and genetic effect provided
639 with the Popcorn software. The regression model ('-use_regression') was used to
640 estimate genetic correlation. We calculated both the correlation of genetic effects
641 (correlation of allelic effect sizes) and genetic impact (correlation of allelic effect
642 size adjusted for difference in allele frequencies).

643 *Conditional SNP analysis.* Conditional and joint SNP analysis (COJO, GCTA
644 version 1.91.1b)^{61,62} was performed to identify potential secondary GWAS signals
645 within a single locus. SNPs with association $P \leq 5 \times 10^{-8}$ were considered. Controls
646 of European ancestry from the Health and Retirement Study (HRS, cohort 65,
647 Supplementary Table 1), included in stratum 4 of this study, were used as the LD
648 reference panel.

649 **Gene prioritization.** *Whole-genome sequencing.* Sample selection, sequencing and
650 data preparation. Patients with ALS and control participants from Project MinE⁶³
651 were recruited for WGS. The participating cohorts were not pre-screened for
652 ALS-associated mutations and are described in the Supplementary Note. In total,
653 228 patients were known to have at least one first- or second-degree relative with
654 ALS. A full description of Project MinE and the sequencing and quality-control
655 pipeline were described previously⁶⁴. In summary, the first batch of 2,250 cases
656 and control samples was sequenced on the Illumina HiSeq 2000 platform. All
657 remaining 7,350 case and control samples were sequenced on the Illumina HiSeq
658 X platform. All samples were sequenced to ~35× coverage with 100-bp reads and
659 ~25× coverage with 150-bp reads for HiSeq 2000 and HiSeq X, respectively. Both
660 sequencing sets used PCR-free library preparation. Samples were also genotyped
661 on the Illumina 2.5M array. Sequencing data were then aligned to GRCh37 using
662 the Isaac Aligner, and variants were called using the Isaac variant caller; both
663 the aligner and caller are standard to Illumina's aligning and calling pipeline.
664 Full details of individual- and variant-level quality control are described in the
665 Supplementary Note.

666 *Genic burden association analyses.* To aggregate rare variants in a genic
667 burden test framework, we used a variety of variant filters to allow for different
668 genetic architectures of ALS-associated variants per gene as we and others did
669 previously^{64,65}. In summary, variants were annotated according to allele-frequency
670 threshold (MAF < 0.01 or MAF < 0.005) and predicted variant impact ('missense',
671 'damaging', 'disruptive'). 'Disruptive' variants were those variants classified
672 as frameshift, splice site, exon loss, stop gained, start loss and transcription
673 ablation by SnpEff⁶⁶. 'Damaging' variants were missense variants predicted to
674 be damaging by seven prediction algorithms (SIFT⁶⁷, PolyPhen-2 (ref. ⁶⁸), LRT⁶⁹,
675 MutationTaster2 (ref. ⁷⁰), Mutations Assessor⁷¹ and PROVEAN⁷²). 'Missense'
676 variants were those missense variants that did not meet the 'damaging' criteria. All
677 combinations of allele-frequency threshold and variant annotations were used to
678 test the genic burden on a transcript level in a Firth logistic regression framework
679 in which burden was defined as the number of variants per individual. Sex and the
680 first 20 PCs were included as covariates. All Ensembl protein-coding transcripts for
681 which at least five individuals had a non-zero burden were included in the analysis.

682 *Conditional genic burden analysis.* We selected for each gene the protein-coding
683 transcripts that were the most strongly associated with ALS across all different
684 combinations of MAF and variant-impact thresholds. For these transcripts and
685 variants, we applied Firth logistic regression on individuals with overlapping
686 GWAS and WGS datasets (5,158 cases and 2,167 controls). To assess whether the
687 rare variant burden association and the signal from the GWAS were conditionally
688 independent, we subsequently included the genotype of the top associated SNP
689 within that locus as a covariate.

690 *Short tandem repeat screen.* For all individuals who had sequencing results in the
691 HiSeq X dataset (5,392 cases, 1,795 controls), we screened all loci harboring SNPs
692 associated with ALS meeting genome-wide significance for expansions of known
693 and new STRs using ExpansionHunter⁷³ and ExpansionHunter Denovo⁷⁴.

694 First, we used ExpansionHunter (version 4.0) to screen for expansions of
695 known STRs located within 1 Mb of the top ALS-associated SNP. For this, we used
696 the STRs identified from indels in 18 high-quality genomes and the GangSTR
697 STR catalog based on STR annotations in the reference genome⁷⁵. We excluded
698 all homopolymers from these catalogs. Repeat length was subsequently regressed
699 on case-control status using Firth logistic regression including the first 20 PCs
700 as covariates, recoding the STR size to a biallelic variant using a sliding window
701 over all observed repeat lengths. To correct for multiple testing across all possible
702 thresholds, we applied Benjamini–Hochberg correction per STR.

703 To screen for extremely long STR expansions (similar to the *C9orf72* repeat
704 expansion) at loci that were not included in the predefined STR catalogs,
705 we applied ExpansionHunter Denovo⁷⁴. This method aims to only find STR
706 expansions that exceed the sequencing read length (>150 bp) by identifying reads
707 (mapped, mismatched and unmapped) that contain STR motifs, using their mate
708 pairs for de novo mapping to the reference genome.

709 For all STRs, we calculated LD statistics (r^2 and $|D'|$) between recoded
710 repeat genotypes at the optimal threshold and the top associated GWAS SNP.
711 Subsequently, we conditioned the SNP association on the repeat genotype in a
712 Firth logistic regression.

713 *Summary-based Mendelian randomization.* We used multi-SNP SMR^{76,77} to infer
714 the effect of gene expression variation on ALS using eQTL (the association of
715 a SNP with expression of a gene) on ALS risk. We chose to apply SMR because
716 this method yielded very similar results when compared to S-PrediXcan⁷⁸ and
717 TWAS⁷⁹ (Supplementary Fig. 18) when applied using GTEx version 7 eQTL, and
718 it can be applied to the large relevant eQTL datasets (MetaBrain and eQTLGen)
719 without access to individual-level genotype and gene expression data. MetaBrain
720 is a harmonized set of 8,727 RNA-seq samples from seven regions of the central
721 nervous system from 15 datasets, and we selected eQTL derived from the cortex
722 region of the brain in samples of European ancestry (MetaBrain Cortex-EUR
723 eQTL, $n = 2,970$ individuals, $n = 6,601$ RNA-seq samples) as our instrument
724 variable²⁴. European-only ALS summary statistics were used as the outcome.
725 To supplement this analysis, we also used eQTL in blood from the eQTLGen
726 Consortium, as this is a large available eQTL resource. Samples of European
727 ancestry in the HRS (cohort 65 of this GWAS) were used as the LD reference panel.
728 SNPs with MAF $\geq 1\%$ in the HRS were included. Further SMR settings were left as
729 default, meaning probes with at least one eQTL with $P \leq 5 \times 10^{-8}$ were included.

730 We subsequently performed SMR using DNA mQTL data and European-only
731 ALS summary statistics. Human prefrontal cortex and whole-blood DNA mQTL
732 were generated as part of ongoing analyses by the Complex Disease Epigenomics
733 Group at the University of Exeter (<https://www.epigenomicslab.com/>) using the
734 Illumina EPIC HumanMethylation array that quantifies DNAm at >850,000
735 sites across the genome²⁵. The prefrontal cortex mQTL dataset was generated
736 using DNA-methylation and SNP data from 522 individuals from the Brains for
737 Dementia Research cohort²⁶ and includes 4,623,966 *cis* mQTL (distance between
738 quantitative trait locus SNP and DNAm site ≤ 500 kb) between 1,744,102 SNPs and
739 43,337 DNA-methylation sites. The whole-blood mQTL dataset was generated
740 using DNAm and SNP data from 2,082 individuals⁸⁰ and included 30,432,023
741 *cis* mQTL between 4,030,902 SNPs and 167,854 DNA-methylation sites. mQTL

reaching the significance threshold $P \leq 1 \times 10^{-10}$ were taken forward for SMR analysis as described by Hannon and colleagues⁸⁰. To map CpG sites to their putative target genes, we used the expression quantitative trait methylation results from a paired methylation and gene expression (RNA-seq) study in blood⁸¹. For CpG sites where no expression quantitative trait methylation was present in this dataset, we used positional mapping based on the basal regulatory domains and extended regulatory domains as defined in the Genomic Regions Enrichment of Annotations Tool (GREAT)⁸², which is applied in the 'cpg_to_gene' function in the CpGtools toolkit⁸³.

Polygenic risk score calculation. PRSs were constructed based on the 15 lead SNPs of genome-wide significant loci (15-SNP PRS) or a full-genome-wide model (full-genome PRS). For the 15-SNP PRS, the SNP weights were defined as the meta-analyzed effect estimates. We used the summary-BayesR framework from the Genome-wide Complex Trait Bayesian analysis (GCTB) toolkit^{84,85} to obtain SNP weights for the full-genome PRS based on the European ancestry meta-analysis excluding stratum 6. We used the default model parameters and the precalculated sparse LD matrix of imputed HapMap3 SNPs in 50,000 random individuals included in the UK Biobank of European ancestries. Summary-BayesR SNP effects were plotted against marginal SNP effects to rule out potential biased estimates due to non-convergence of the MCMC algorithm. Finally, the PRSs for all individuals in stratum 6 were calculated using the '-score' function in PLINK and normalized to zero mean and unit variance.

Modifier analyses. For 6,095 of the patients with WGS and ALS, core clinical data were obtained including sex, site of onset (spinal or bulbar), age at onset (years), country of origin and survival, defined as time from disease onset to death, 23 h of continuous non-invasive ventilation per day or tracheostomy. Patients who were still alive were censored at the last date of follow-up.

The genetic risk factors included SNP genotypes, PRSs, *C9orf72* repeat expansion status and the number of rare coding mutations in ALS-risk genes (*SOD1*, *TARDBP*, *FUS*, *NEK1*, *TBK1* and *CFAP410*) as obtained from WGS as described above.

For survival analyses, the Cox proportional hazards mixed model from the 'coxme' package in R was used, modeling country of origin as a random effect. Fixed-effect covariates included sex, age at onset, site of onset, GWAS stratum and PC1-PC5. Violation of the proportional hazards assumption for genotype on survival was assessed by inspecting Schoenfeld residuals. For age-at-onset analyses, we applied linear regression of age at onset on genotype including sex, site of onset, country, GWAS stratum and PC1-PC5 as covariates.

Cross-trait analyses. Datasets and data preparation. GWAS summary statistics for clinically diagnosed AD⁸⁶, PD⁸⁷, FTD⁸⁸, CBD⁸⁹ and PSP²⁰ in individuals of European ancestry were obtained. For AD, we used the clinical diagnosis as the case definition to avoid spurious genetic correlations that could have been introduced through the by-proxy design³¹, in which by-proxy cases are defined as having a parent with AD. Although this is a powerful design for gene discovery and the genetic correlation with clinically diagnosed AD is high⁹⁰, mislabeling by-proxy cases when parents suffer from other types of dementia (for example, Lewy body dementia, Parkinson's dementia, FTD or vascular dementia) can lead to spurious genetic correlations with ALS and other neurodegenerative diseases. For FTD, we primarily used the results of the cross-subtype meta-analysis, which includes behavioral variant FTD, semantic dementia FTD, progressive non-fluent aphasia FTD and mndFTD. For CBD, allele coding was unavailable, and effect alleles were inferred by matching allele frequencies to those observed in the Haplotype Reference Consortium. SNPs with MAF > 0.4 were excluded. Because downstream methods rely on LD scores or population-specific LD patterns, the European ancestry summary statistics from the present study were used for ALS. For sample size parameters, effective sample size was calculated as described previously.

Multiple sclerosis summary statistics were obtained from the International Multiple Sclerosis Genetics Consortium⁹¹. For cerebrovascular diseases, GWAS summary statistics were obtained for ischemic stroke (any ischemic stroke)⁹², intracerebral hemorrhage⁹³ and intracranial aneurysm⁹⁴. For psychiatric traits, GWAS summary statistics were obtained from Psychiatric Genomics Consortium studies on anorexia nervosa⁹⁵, obsessive-compulsive disorder⁹⁶, anxiety disorders (anxiety score)⁹⁷, post-traumatic stress disorder (all European ancestries)⁹⁸, major depressive disorder⁹⁹, bipolar disorder¹⁰⁰, schizophrenia¹⁰¹, Tourette's syndrome¹⁰², autism spectrum disorder¹⁰³ and attention-deficit hyperactivity disorder (European ancestries)¹⁰⁴.

Genetic correlation. Genome-wide genetic correlation between neurodegenerative traits was calculated using LDSC (version 1.0.0)⁵⁹. Precomputed LD scores of European individuals in the 1000 Genomes project for high-quality HapMap3 SNPs were used ('eur_w_ld_chr'). A free intercept was modeled to allow for potential sample overlap.

Colocalization. Before the colocalization analysis of neurodegenerative diseases, we first assessed residual confounding by estimating the LDSC intercept using LDSC (version 1.0.0) (ALS, 1.03 (s.e.m., 0.0073); AD, 1.03 (s.e.m., 0.013); PD, 0.98 (s.e.m.,

0.0065); PSP, 1.05 (s.e.m., 0.0076); CBD, 0.98 (s.e.m., 0.0073); FTD, 1.00 (s.e.m., 0.0071)), showing limited inflation of test statistics due to confounding across these studies. For each locus (top SNP ± 100 kb) harboring SNPs with an association with any of the neurodegenerative diseases (ALS, AD, PD, PSP, CBD, FTD) at $P < 1 \times 10^{-5}$, we performed colocalization analysis using the 'coloc' package in R¹⁰⁵. We set the prior probabilities to $\pi_1 = 1 \times 10^{-4}$, $\pi_2 = 1 \times 10^{-4}$ and $\pi_{12} = 1 \times 10^{-5}$ for a causal variant in trait 1 or trait 2 and a shared causal variant between traits 1 and 2, respectively. Using the same parameters, we performed colocalization analysis for ALS and each of the FTD subtypes (behavioral variant FTD, semantic dementia FTD, progressive non-fluent aphasia FTD and mndFTD).

Enrichment analyses. Linkage disequilibrium score regression annotation-specific enrichment analysis. We used LDSC (version 1.0.0)⁵⁹ to calculate SNP-based heritability, the LDSC intercept and SNP-based heritability enrichment for partitions of the genome. In all LDSC analyses, summary statistics excluding the HLA region of only samples of European ancestry were included. LD scores and partitioned LD scores provided by LDSC were used for genome-wide and genic region-based heritability analyses. The option '-overlap-annot' was used in the partitioned heritability analysis to allow for overlapping SNPs between MAF bins. SNPs with MAF > 5% were included.

Tissue and cell type enrichment analysis. Tissue and cell type enrichment analyses were performed using the GWAS summary statistics of the European ancestry meta-analysis and FUMA³³ software version 1.3.6a. FUMA performs a genic aggregation analysis of GWAS association signals to calculate gene-wise association signals using MAGMA version 1.6 and subsequently tests whether tissues and cell types are enriched for expression of these genes. For tissue enrichment analysis, we used the GTEx version 8 reference set. FDR-corrected P values < 0.05 across all tissues ($n = 54$) were considered statistically significant. For cell type enrichment analyses³⁴, we used human-derived single-cell RNA-seq data on major brain cell types (*GSE67835* without fetal samples¹⁰⁶), Allen Brain Atlas cell types¹⁰⁷ for the human-derived major neuronal subtypes and the DropViz¹⁰⁸ dataset for mouse-derived brain cell types across all brain regions. We applied FDR correction for multiple testing within each expression dataset, and FDR-corrected P values < 0.05 were considered statistically significant.

Pathway enrichment analysis. We used Downstreamer software²⁴ to identify enriched biological pathways and processes. First, gene-based association statistics were obtained with the Pascal method¹⁰⁹, which aggregates SNP association statistics including SNPs up to 10 kb upstream and downstream of a gene, accounting for LD using the non-Finnish European individuals from the 1000 Genomes Project phase 3 (ref. ¹¹⁰) as a reference. In the Downstreamer method, putative core genes are defined as those that are coexpressed with disease-associated genes and can therefore be implicated in disease. Coexpression networks are based on either a large, multi-tissue transcriptome dataset including 56,435 genes and 31,499 individuals or brain-specific RNA-seq data obtained from the MetaBrain resource. The gene-based association statistics, coexpression matrix and gene Z scores per pathway or HPO term are then combined in a generalized least-squares regression model to obtain enrichment statistics²⁴. Enrichment analyses were performed for reactome, gene ontology and HPO terms using multi-tissue or brain-specific transcriptome datasets to calculate the coexpression matrix.

The distribution of enrichment Z -score statistics was compared between analyses using multi-tissue or brain-specific coexpression matrices. Using the 'pyhpo' module in Python, all HPO terms were assigned to their parent term(s) in the 'phenotypic abnormality' (HP:0000118) branch, which includes phenotypic abnormalities grouped per organ system.

Mendelian randomization. Causal inference through MR analysis was performed for 22 exposures for which large-scale GWASs are available and for which there is prior evidence for an association with ALS. These include seven behavioral-related traits: body mass index (anthropometric)¹¹¹, years of schooling (educational attainment)¹¹², alcoholic drinks per week, age of smoking initiation and cigarettes per day from Liu et al.¹¹³, days per week of moderate physical activity and days per week of vigorous activity from the UK Biobank¹¹⁴; four blood pressure traits (coronary artery disease¹¹⁵, stroke⁹², diastolic blood pressure and systolic blood pressure¹¹⁶); seven immune system traits from Vuckovic et al.¹¹⁷ (basophil, eosinophil, lymphocyte, monocyte, neutrophil and white blood cell counts) and C-reactive protein¹¹⁸; and four lipid traits from Willer et al.¹¹⁹ (HDL cholesterol, LDL cholesterol, total cholesterol and triglyceride levels). A full description of the included studies is provided in Supplementary Table 26. From these GWASs, SNPs to serve as instruments for MR analyses were selected at two different P -value cutoffs ($P < 5 \times 10^{-8}$ and $P < 5 \times 10^{-5}$) and then LD clumped to obtain independent SNPs. SNP effect estimates on ALS risk were obtained from the European ancestry-only GWAS and, if needed, an LD proxy was selected ($r^2 > 0.8$).

After harmonizing effect alleles and excluding palindromic SNPs, we performed a series of quality-control steps to avoid biased estimates of causal effects, checking for each exposure (1) instrument coverage (>85% overlapping SNPs; Supplementary Table 31), (2) instrument strength (F -statistic^{37,120,121} > 10;

Supplementary Table 32), (3) distribution and significance of the Wald ratios (visual inspection of volcano plots; Supplementary Table 33) and (4) heterogeneity across the instrument-exposure effects (Q -statistic at $P < 0.05$ indicated heterogeneity; Supplementary Table 34).

We applied five different MR methods: IVW using the random-effects model, MR-Egger and simple mode, weighted median and weighted mode methods. When only a single SNP was available, the Wald ratio test was conducted. MR analysis was conducted in R using the 'mr()' function in the 'TwoSampleMR' package¹²².

Subsequently, radial MR analysis was conducted to determine whether Wald ratio outliers needed to be removed from the IVW or MR-Egger MR estimates³⁸. In addition, we conducted a Q -test to identify outlier SNPs ($P < 0.05$). These outliers were then removed from the original MR analyses (across all five MR methods). The radial MR analysis was conducted using the RadialMR R package (<https://github.com/WSpiller/RadialMR>). To determine whether MR effects were orientated in the correct direction (from exposure to ALS), we conducted both reverse MR¹²³ and Steiger filtering¹²⁴ on our top MR findings.

Finally, we explored whether the MR effects of our total and LDL cholesterol and systolic blood pressure exposures may be confounded by the effect we observed for years of schooling by conducting multivariate MR analysis¹²⁵. Conditional F - and Q -statistics were calculated using the 'MVMR' package¹²⁶ in R.

Statistical analyses. All presented P values correspond to two-sided P values uncorrected for multiple testing unless explicitly stated otherwise.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The GWAS summary statistics generated in this study are publicly available in the NHGRI-EBI GWAS Catalog at <https://www.ebi.ac.uk/gwas/> (accession IDs GCST90027163 and GCST90027164 for cross-ancestry and European ancestry meta-analyses, respectively) and through the Project MinE website (<https://www.projectmine.com/research/download-data/>). Summary statistics of the rare variant burden analyses and eQTL and mQTL SMR analyses are available through the Project MinE website. The following publicly available datasets were used in this project: the Wellcome Trust Case Control Consortium (<https://www.wtccc.org.uk/>) and dbGaP datasets (phs000101.v3.p1, NIH Genome-Wide Association Studies of Amyotrophic Lateral Sclerosis; phs000126.v1.p1, CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD); phs000196.v1.p1, Genome-Wide Association Study of Parkinson Disease: Genes and Environment; phs000344.v1.p1, Genome-Wide Association Study of Amyotrophic Lateral Sclerosis in Finland; phs000336, a Genome-Wide Association Study of Lung Cancer Risk; phs000346, Genome-Wide Association Study for Bladder Cancer Risk; phs000789, Collaborative Study of Genes, Nutrients and Metabolites (CSGNM); phs000206, Whole Genome Scan for Pancreatic Cancer Risk in the Pancreatic Cancer Cohort Consortium and Pancreatic Cancer Case-Control Consortium (PanScan); phs000297, eMERGE Network Study of the Genetic Determinants of Resistant Hypertension; phs000652, Cohort-Based Genome-Wide Association Study of Glioma (GliomaScan); phs000869, Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGESS); phs000812, the Breast and Prostate Cancer Cohort Consortium (BPC3) GWAS of Aggressive Prostate Cancer and ER- Breast Cancer; phs000428, Genetics Resource with the HRS; phs000360.v3, eMERGE Network Genome-Wide Association Study of Red Cell Indices, White Blood Count (WBC) Differential, Diabetic Retinopathy, Height, Serum Lipid Levels, Specifically Total Cholesterol, HDL (High Density Lipoprotein), LDL (Low Density Lipoprotein), and Triglycerides, and Autoimmune Hypothyroidism; phs000893.v1, Genome-Wide Association Study of Endometrial Cancer in the Epidemiology of Endometrial Cancer Consortium (E2C2); phs000168.v2, National Institute on Aging—Late Onset Alzheimer's Disease Family Study: Genome-Wide Association Study for Susceptibility Loci; phs000092.v1, Study of Addiction: Genetics and Environment (SAGE); phs000864.v1, Genomic Predictors of Combat Stress Vulnerability and Resilience; phs000170.v2, a Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project Cohort; phs000431.v2, IgA Nephropathy GWAS on Individuals of European Ancestry (IGANGWAS2); phs000237.v1, Northwestern NUGene Project: Type 2 Diabetes; phs000169.v1, Whole Genome Association Study of Visceral Adiposity in the Health Aging and Body Composition (Health ABC) Study; phs000982.v1, Genetic Analysis of Psoriasis and Psoriatic Arthritis: GWAS of Psoriatic Arthritis; phs000289.v2, National Human Genome Research Institute (NHGRI) GENEVA Genome-Wide Association Study of Venous Thrombosis (GWAS of VTE); phs000634.v1, National Cancer Institute (NCI) Genome Wide Association Study (GWAS) of Lung Cancer in Never Smokers; phs000274.v1, Genome-Wide Association Study of Celiac Disease; phs001172.v1, National Institute of Neurological Disorders and Stroke (NINDS) Parkinson's Disease; phs000389.v1, GEnetics of Nephropathy—an International Effort (GENIE) GWAS of Diabetic Nephropathy in the UK GoKinD and All-Ireland Cohorts; phs000460.v1, Genetics of 24 Hour Urine Composition; phs000138.v2, GWAS for Genetic Determinants of Bone Fragility in European-American

Pre-menopausal Women; phs000394.v1, Autopsy-Confirmed Parkinson Disease GWAS Consortium (APDGC); phs000948.v1, Genetic Discovery and Application in a Clinical Setting: Continuing a Partnership (eMERGE Phase II); phs000630.v1, Exome Chip Study of NIMH Controls; phs000678.v1, a Family-Based Study of Genes and Environment in Young-Onset Breast Cancer; phs000351.v1, National Cancer Institute Genome-Wide Association Study of Renal Cell Carcinoma; phs000314.v1, Genetic Associations in Idiopathic Talipes Equinovarus (Clubfoot)—GAIT; phs000147.v3, Cancer Genetic Markers of Susceptibility (CGEMS) Breast Cancer Genome-wide Association Study (GWAS)—Primary Scan: Nurses' Health Study—Additional Cases: Nurses' Health Study 2; phs000882.v1, National Cancer Institute (NCI) Prostate Cancer Genome-Wide Association Study for Uncommon Susceptibility Loci (PEGASUS); phs000238.v1, National Eye Institute Glaucoma Human Genetics Collaboration (NEIGHBOR) Consortium Glaucoma Genome-Wide Association Study; phs000397.v1, National Institute on Aging (NIA) Long Life Family Study (LLFS); phs000421.v1, a Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy (FECD); phs000142.v1, a Whole Genome Association Scan for Myopia and Glaucoma Endophenotypes using Twin Studies; phs000303.v1, Genetic Epidemiology of Refractive Error in the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) Study; phs000125.v1, CIDR: Collaborative Study on the Genetics of Alcoholism Case Control Study; phs001039.v1, International Age-Related Macular Degeneration Genomics Consortium—Exome Chip Experiment; phs000187.v1, High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation; phs000101.v5, Genome-Wide Association Study of Amyotrophic Lateral Sclerosis; phs002068.v1.p1, Sporadic ALS Australia Systems Genomics Consortium (SALSA-SGC)). Source data are provided with this paper.

Code availability

The following software packages were used for data analyses: R version 3.6.3 with additional packages tidyverse version 1.3.0, data.table version 1.14.0, ggplot2 version 3.3.3, MASS version 7.3.53, SNPRelate version 1.26.0, logistf version 1.24, coloco version 5.1.0, twoSampleMR version 0.5.6, RadialMR version 1.0, MVMR version 0.3, survival version 3.1.8, coxme version 2.2.16 and survminer version 0.4.9 (<https://www.r-project.org/>), Python version 3.7 with additional modules pandas version 1.1.3, numpy version 1.18.1, scipy version 1.4.1, CpGtools version 1.0.9, matplotlib version 3.1.3, pyliftover version 0.4 and pypho version 2.5.0 (<https://anaconda.org/>), GenomeStudio version 2.0 (<https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html>), GCTA version 1.93.2beta (<https://cns.genomics.com/software/gcta/#Overview>), EIGENSOFT version 6.1.4 (<https://github.com/DreichLab/EIG>), SNPTEST version 2.5.4-beta3 (<https://www.well.ox.ac.uk/~gav/snptest/>), PLINK version 1.9 (<http://www.cog-genomics.org/plink2>), the Michigan Imputation Server (<https://imputationserver.sph.umich.edu>), SAIGE version 0.29.1 (<https://github.com/weizhouUMICH/SAIGE>), METAL 2011-03-25 (<https://genome.sph.umich.edu/wiki/METAL>), SnpSift 4.3p (<https://pcingola.github.io/SnpSift>), ANNOVAR version 2017-07-17 for LRT, Polyphen-2, MutationTaster2, Mutation Assessor, PROVEAN and SIFT (<https://annovar.openbioinformatics.org/>), Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), MutationTaster2 (<http://www.mutationtaster.org/>), Mutation Assessor release 3 (<http://mutationassessor.org/r3/>), PROVEAN version 1.1 (<http://provean.jcvi.org/index.php>), SIFT version 6.2.1 (<https://sift.bii.a-star.edu.sg/>), SnpEff 4.3p (<https://pcingola.github.io/SnpEff>), LDSC version 1.0.1 (<https://github.com/bulik/ldsc>), ExpansionHunter version 4 (<https://github.com/illumina/ExpansionHunter>), ExpansionHunter Denovo (<https://github.com/illumina/ExpansionHunterDenovo>), SMR (<https://cns.genomics.com/software/smr/>), MAGMA version 1.6 (<https://ctg.cncr.nl/software/magma>), FUMA (<https://fuma.ctglab.nl/>), FUMA Cell-type (<https://fuma.ctglab.nl/celltype>), summary-BayesR (<https://cns.genomics.com/software/gctb/#SummaryBayesianAlp>), S-PrediXcan (<https://github.com/hakyimlab/MetaXcan>) and TWAS (<http://gusevlab.org/projects/fusion/>).


References

- Brooks, B. R., Miller, R. G., Swash, M. & Munsat, T. L. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **1**, 293–299 (2000).
- Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Brown, B. C. et al. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).


- 790 61. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS
791 summary statistics identifies additional variants influencing complex traits.
792 *Nat. Genet.* **44**, 369–375 (2012).
- 793 62. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for
794 genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- 795 63. Project MinE ALS Sequencing Consortium. Project MinE: study design and
796 pilot analyses of a large-scale whole-genome sequencing study in
797 amyotrophic lateral sclerosis. *Eur. J. Hum. Genet.* **26**, 1537–1546 (2018).
- 798 64. Spek, R. A. Avander et al. The Project MinE databrowser: bringing
799 large-scale whole-genome sequencing in ALS to researchers and the public.
800 *Amyotroph. Lateral Scler. Frontotemporal Degener.* **20**, 432–440 (2019).
- 801 65. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants
802 among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441
803 (2016).
- 804 66. Cingolani, P. et al. A program for annotating and predicting the effects of
805 single nucleotide polymorphisms. *SnEff. Fly* **6**, 80–92 (2012).
- 806 67. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense
807 predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
- 808 68. Adzhubei, I. A. et al. A method and server for predicting damaging
809 missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- 810 69. Chun, S. & Fay, J. C. Identification of deleterious mutations within three
811 human genomes. *Genome Res.* **19**, 1553–1561 (2009).
- 812 70. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelove, D. MutationTaster2:
813 mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362
814 (2014).
- 815 71. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of
816 protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**,
817 e118 (2011).
- 818 72. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the
819 functional effect of amino acid substitutions and indels. *Bioinformatics* **31**,
820 2745–2747 (2015).
- 821 73. Dolzhenko, E. et al. Detection of long repeat expansions from PCR-free
822 whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).
- 823 74. Dolzhenko, E. et al. ExpansionHunter Denovo: a computational method for
824 locating known and novel repeat expansions in short-read sequencing data.
825 *Genome Biol.* **21**, 102 (2020).
- 826 75. Mousavi, N., Shleizer-Burko, S., Yanicky, R. & Gymrek, M. Profiling the
827 genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* **47**,
828 e90 (2019).
- 829 76. Wu, Y. et al. Integrative analysis of omics summary data reveals putative
830 mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
- 831 77. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies
832 predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
- 833 78. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue
834 specific gene expression variation inferred from GWAS summary statistics.
835 *Nat. Commun.* **9**, 1825 (2018).
- 836 79. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide
837 association studies. *Nat. Genet.* **48**, 245–252 (2016).
- 838 80. Hannon, E. et al. Leveraging DNA-methylation quantitative-trait loci to
839 characterize the relationship between methylomic variation, gene
840 expression, and complex traits. *Am. J. Hum. Genet.* **103**, 654–665 (2018).
- 841 81. Hop, P. J. et al. Genome-wide identification of genes regulating DNA
842 methylation using genetic anchors for causal inference. *Genome Biol.* **21**,
843 220 (2020).
- 844 82. McLean, C. Y. et al. GREAT improves functional interpretation of
845 *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
- 846 83. Wei, T. et al. CpGtools: a Python package for DNA methylation analysis.
847 *Bioinformatics* **37**, 1598–1599 (2021).
- 848 84. Zeng, J. et al. Signatures of negative selection in the genetic architecture of
849 human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
- 850 85. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple
851 regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
- 852 86. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease
853 identifies new risk loci and implicates A β , tau, immunity and lipid
854 processing. *Nat. Genet.* **51**, 414–430 (2019).
- 855 87. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and
heritable risk for Parkinson's disease: a meta-analysis of genome-wide
association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
88. Ferrari, R., Hernandez, D. G., Nalls, M. A. & Rohrer, J. D. Frontotemporal
dementia and its subtypes: a genome-wide association study. *Lancet Neurol.*
13, 686–699 (2014).
89. Kouri, N. et al. Genome-wide association study of corticobasal degeneration
identifies risk variants shared with progressive supranuclear palsy. *Nat.*
Commun. **6**, 7247 (2015).
90. Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. *Transl.*
Psychiatry **8**, 99 (2018).
91. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis
genomic map implicates peripheral immune cells and microglia in
susceptibility. *Science* **365**, eaav7188 (2019).
92. Malik, R. et al. Multiancestry genome-wide association study of 520,000
subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat.*
Genet. **50**, 524–537 (2018).
93. Woo, D. et al. Meta-analysis of genome-wide association studies identifies
1q22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum.*
Genet. **94**, 511–521 (2014).
94. Bakker, M. K. et al. Genome-wide association study of intracranial
aneurysms identifies 17 risk loci and genetic overlap with clinical risk
factors. *Nat. Genet.* **52**, 1303–1313 (2020).
95. Watson, H. J. et al. Genome-wide association study identifies eight risk loci
and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.*
51, 1207–1214 (2019).
96. International Obsessive Compulsive Disorder Foundation Genetics
Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association
Studies (OC GAS). Revealing the complex genetic architecture of
obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**,
1181–1188 (2018).
97. Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety
disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
98. Nievergelt, C. M. et al. International meta-analysis of PTSD genome-wide
association studies identifies sex- and ancestry-specific genetic risk loci.
Nat. Commun. **10**, 4558 (2019).
99. Wray, N. R. et al. Genome-wide association analyses identify 44 risk
variants and refine the genetic architecture of major depression. *Nat. Genet.*
50, 668–681 (2018).
100. Stahl, E. A. et al. Genome-wide association study identifies 30 loci
associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
101. Schizophrenia Working Group of the Psychiatric Genomics Consortium.
Biological insights from 108 schizophrenia-associated genetic loci. *Nature*
511, 421–427 (2014).
102. Yu, D. et al. Interrogating the genetic determinants of Tourette's syndrome
and other tic disorders through genome-wide association studies. *Am. J.*
Psychiatry **176**, 217–227 (2019).
103. Grove, J. et al. Identification of common genetic risk variants for autism
spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
104. Demontis, D. et al. Discovery of the first genome-wide significant risk loci
for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
105. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of
genetic association studies using summary statistics. *PLoS Genet.* **10**,
e1004383 (2014).
106. Darmanis, S. et al. A survey of human brain transcriptome diversity at the
single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
107. Hodge, R. D. et al. Conserved cell types with divergent features in human
versus mouse cortex. *Nature* **573**, 61–68 (2019).
108. Saunders, A. et al. Molecular diversity and specializations among the cells
of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
109. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast
and rigorous computation of gene and pathway scores from SNP-based
summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
110. 1000 Genomes Project Consortium et al. A global reference for human
genetic variation. *Nature* **526**, 68–74 (2015).
111. Yengo, L. et al. Meta-analysis of genome-wide association studies for height
and body mass index in ~700,000 individuals of European ancestry. *Hum.*
Mol. Genet. **27**, 3641–3649 (2018).
112. Lee, J. J. et al. Gene discovery and polygenic prediction from a
genome-wide association study of educational attainment in 1.1 million
individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
113. Liu, M. et al. Association studies of up to 1.2 million individuals yield new
insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**,
237–244 (2019).
114. Sudlow, C. et al. UK Biobank: an open access resource for identifying the
causes of a wide range of complex diseases of middle and old age. *PLoS*
Med. **12**, e1001779 (2015).
115. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci
provides an expanded view on the genetic architecture of coronary artery
disease. *Circ. Res.* **122**, 433–443 (2018).
116. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535
new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425
(2018).
117. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and
diseases. *Cell* **182**, 1214–1231 (2020).
118. Ligthart, S. et al. Genome analyses of >200,000 individuals identify 58 loci
for chronic inflammation and highlight pathways that link inflammation
and complex disorders. *Am. J. Hum. Genet.* **103**, 691–706 (2018).
119. Willer, C. J. et al. Discovery and refinement of loci associated with lipid
levels. *Nat. Genet.* **45**, 1274–1283 (2013).
120. Zeng, P., Wang, T., Zheng, J. & Zhou, X. Causal association of type 2
diabetes with amyotrophic lateral sclerosis: new evidence from Mendelian
randomization using GWAS summary statistics. *BMC Med.* **17**, 225 (2019).

- 856 121. Cragg, J. G. & Donald, S. G. Testing identifiability and specification in
857 instrumental variable models. *Econ. Theory* **9**, 222–240 (1993).
- 858 122. Hemani, G. et al. The MR-Base platform supports systematic causal
859 inference across the human genome. *eLife* **7**, e34408 (2018).
- 860 123. Smith, G. D., Davey Smith, G. & Hemani, G. Mendelian randomization:
861 genetic anchors for causal inference in epidemiological studies. *Hum. Mol.
862 Genet.* **23**, R89–R98 (2014).
- 863 124. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship
864 between imprecisely measured traits using GWAS summary data. *PLoS
865 Genet.* **13**, e1007081 (2017).
- 866 125. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the
867 use of pleiotropic genetic variants to estimate causal effects. *Am. J.
868 Epidemiol.* **181**, 251–260 (2015).
- 869 126. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An
870 examination of multivariable Mendelian randomization in the single-sample
871 and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727
(2019).

871 Acknowledgements

872 Acknowledgements and relevant funding details are provided in the Supplementary
873  Note.

874 Author contributions

875  Sample ascertainment: W.v.R., R.A.A.v.d.S., M.M., A.M.D., H.-J. Westenberg, G.H.P.T.,
876 N.T., J.C.-K., B.N.S., M.G., S.C., S.P., K.E.M., P.J.S., J.H., R.W.O., M.S., T.M., N.B.,
877 A.J.v.d.K., A.R., C.G., G.L.P., G.P.C., C.C., D.S., S.D.A., G. Sorarù, G. Siciliano, M.F., A.P.,
878 A.C., A. Calvo, C.M., M.B., A. Canosa, M. Grassano, E.B., E.P., G.L., B.N., A.O., A.N.,
879 Y.L., M.Z., M. Gotkine, R.H.B., S.B., P.V., P.C., P. Couratier, S.M., V.M., F.S., J.S.M.P.,
880 A.A., R.R.-G., P. Dion, J.P.R., A.C.L., J.H.W., D. Brenner, A.F., G.B., A.B., A.D., C.A.M.P.,
881 S.S.-D., N.W., S.T., R. Rademakers, A. Braun, J.K., D.C.W., C.M.O., A.G.U., A.H., M.R.,
882 S. Cichon, M.M.N., P.A., B.J.T., A.B.S., M. Mitne Neto, R.J.C., R.A.O., M.W.-P., C.L.-H.,
883 V.M.v.D., J.G., A. Rödiger, N.G., A.J., T.B., E.T., B.I., B.S., O.W.W., R.S., C.A.H., C. Graff,
884 L.B., V.F., V.D., A. Ataulina, B.R., B.K., J.Z., M.R.-G., D.G., Z.S., V. Drory, M.P., I.P.B.,
885 M.C.K., R.D.H., S. Mathers, P.A.M., M.N., G.A.N., R.P., D.B.R., K.A.M., P.S.S., M.d.C.,
886 S. Pinto, S. Petri, M.W., G.A.R., V.S., J. Glass, R.H. Brown, J.E.L., C.E.S., P.M.A., D.F.,
887 F.C.G., A.F.M., R.L.M., O.H., A.A.-C., P.V.D., L.H.v.d.B., J.H.V., SLALOM Consortium,
888 PARALS Consortium, SLAGEN Consortium and SLAP Consortium. SNP array
889 genotyping: W.v.R., R.A.A.v.d.S., A.M.D., A.S., I.F., G.B., A.B., A.D., C.A.M.P., S.S.-D.,
890 N.W., L.T., W.L., A. Franke, S.R., A. Braun, J.K., D.C.W., C.M.O., A.G.U., A.H., M.R.,

S. Cichon, M.M.N., P.A., B.J.T., A.B.S., B.B., S.F., S.T.N., F.J.S., K.L.W., A.K.H., L.W.,
C. Curtis, G. Breen, D.F., F.C.G., A.F.M., N.R.W., A.A.-C., P.V.D., L.H.v.d.B. and J.H.V.
GWAS quality control: W.v.R., R.A.A.v.d.S., M.K.B., R.R., R.L.M., N.R.W. and J.H.V.
GWAS data analysis: W.v.R., R.A.A.v.d.S., M.K.B., R.R., R.P.B., M.D., M.H., A.A.K., A.I.,
A.S., N.T., B.N.S., B.B., D.F., A.F.M., R.L.M., N.R.W. and J.H.V. WGS: W.v.R., R.A.A.v.d.S.,
P.J.H., R.A.J.Z., M.M., A.M.D., G.H.P.T., K.R.v.E., M.K., J.C.-K., B.N.S., K.P.K., A.A.-C.,
P.V.D., L.H.v.d.B. and J.H.V. WGS quality control: W.v.R., R.A.A.v.d.S., J.J.F.A.v.V., P.J.H.,
R.A.J.Z., M.M., K.P.K., P.V.D. and J.H.V. WGS rare variant burden analyses: W.v.R.,
R.A.A.v.d.S., P.J.H., R.A.J.Z., K.R.v.E., K.P.K., P.V.D. and J.H.V. WGS STR analyses: W.v.R.,
J.J.F.A.v.V., R.A.J.Z., E.D., M.A.E. and J.H.V. eQTL analyses: W.v.R., R.A.A.v.d.S., M.K.B.,
N.d.K., H.-J.W., O.B.B., P.D., J.M., L.F. and J.H.V. mQTL analyses: W.v.R., M.K.B., P.J.H.,
R.A.J.Z., G.S., E.H., A.M.D. and J.H.V. Cross-disorder analyses: W.v.R., R.A.A.v.d.S.,
M.K.B., N.d.K., H.-J.W., O.B.B., P.D., E.J.N.G., M.A.v.E., R.J.P., A.F.M., N.R.W., E. Tsai,
H.R., L.F. and J.H.V. MR analyses: W.v.R., R.A.A.v.d.S., M.K.B., D.B., H.-J.W., G.D.S.,
T.R.G., E. Tsai, H.R. and J.H.V. Writing the manuscript: W.v.R., M.K.B., D.B., J.M.,
E. Tsai and J.H.V. Revising the manuscript: W.v.R., R.A.A.v.d.S., M.K.B., J.J.F.A.v.V., G.S.,
E.H., D.B., R.R., E.D., H.-J.W., G.H.P.T., K.R.v.E., E.J.N.G., M.A.v.E., R.J.P., G.D.S., T.R.G.,
R.L.M., K.P.K., N.R.W., E. Tsai, H.R., L.F., L.H.v.d.B. and J.H.V. Funding acquisition and
study supervision: L.H.v.d.B. and J.H.V.

891 Competing interests

892 J.H.V. has sponsored research agreements with Biogen Idec. L.H.v.d.B. receives personal
893 fees from Cytokinetics outside of the submitted work. A.A.-C. has served on scientific
894 advisory boards for Mitsubishi Tanabe Pharma, Orion Pharma, Biogen Idec, Lilly, GSK,
895 Apellis, Amylyx and Wave Therapeutics. A.C. serves on scientific advisory boards for
896 Mitsubishi Tanabe, Roche, Biogen, Denali and Cytokinetics. The remaining authors
897 declare no competing interests related to this work.

898 Additional information

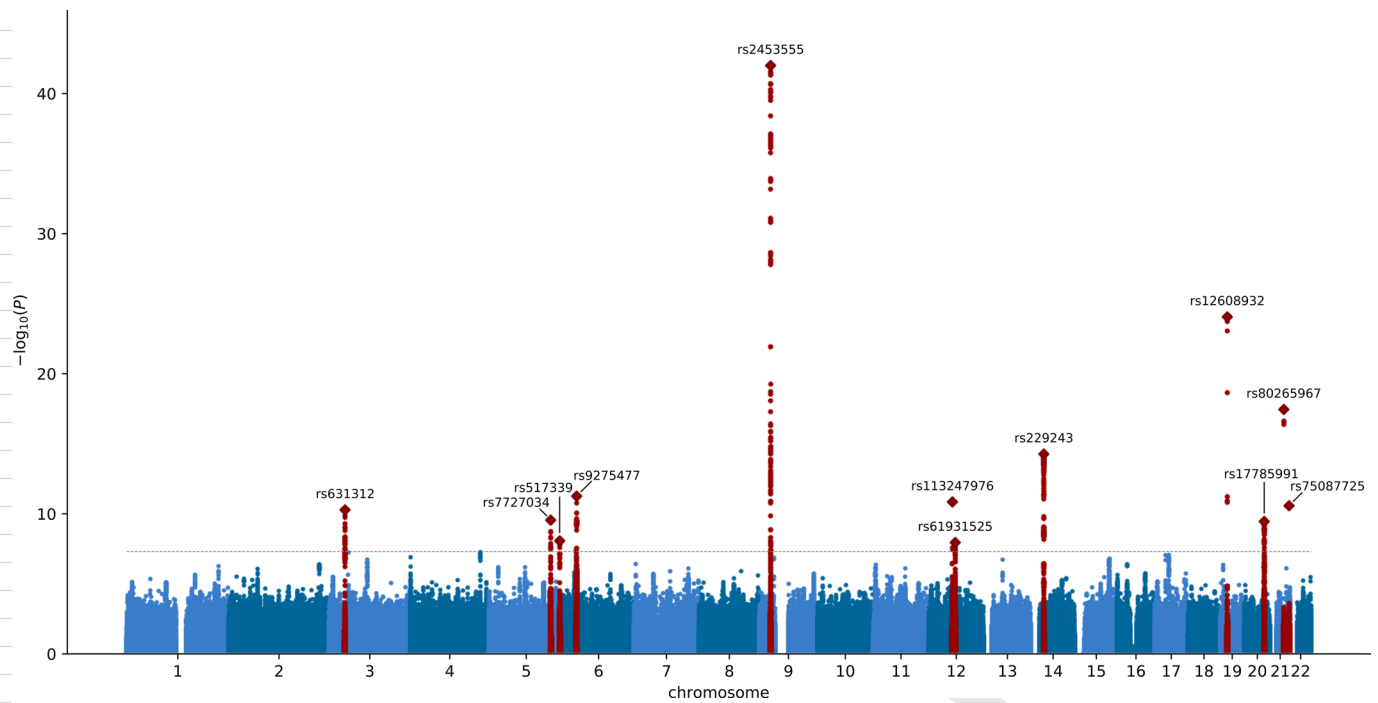
899 **Extended data** is available for this paper at <https://doi.org/10.1038/s41588-021-00973-1>.

900 **Supplementary information** The online version contains supplementary material
901 available at <https://doi.org/10.1038/s41588-021-00973-1>.

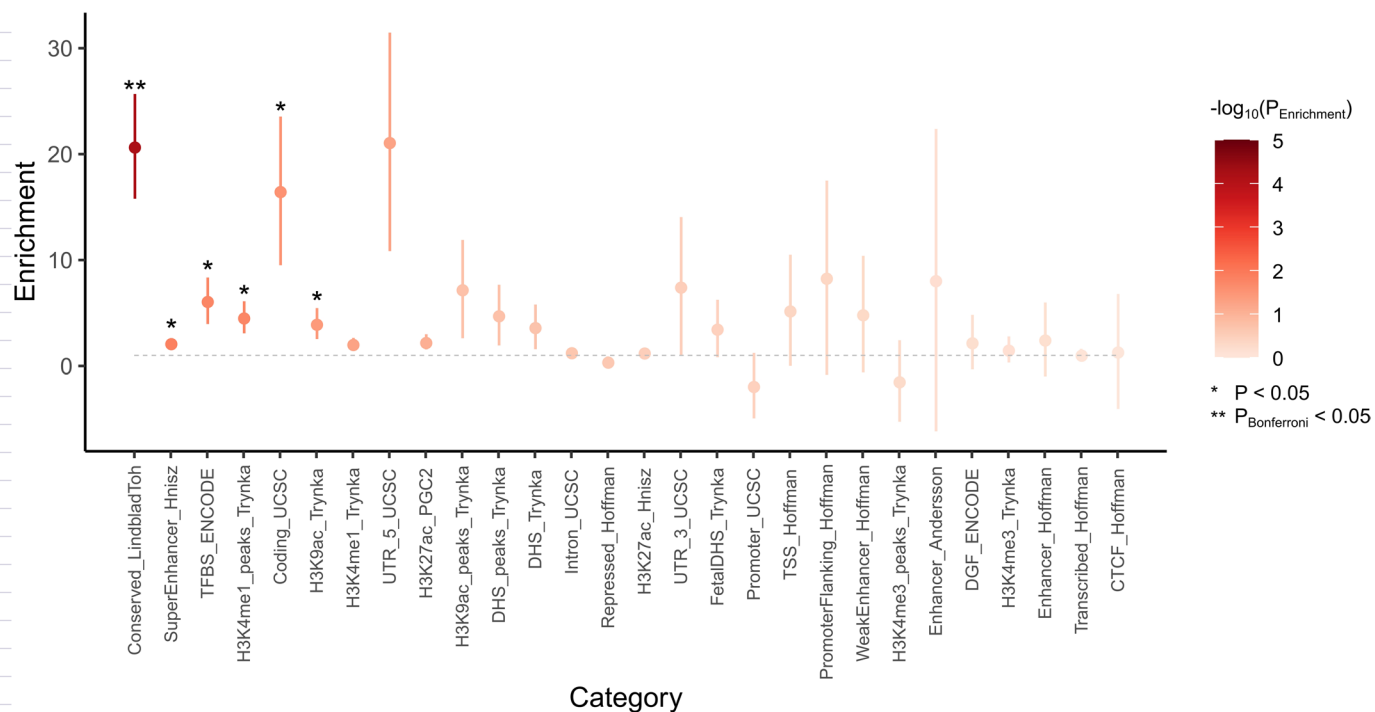
902 **Correspondence and requests for materials** should be addressed to
903 Wouter van Rheenen or Jan H. Veldink.

904 **Peer review information** *Nature Genetics* thanks David Goldstein and the other,
905 anonymous, reviewer(s) for their contribution to the peer review of this work.

906 **Reprints and permissions information** is available at www.nature.com/reprints.

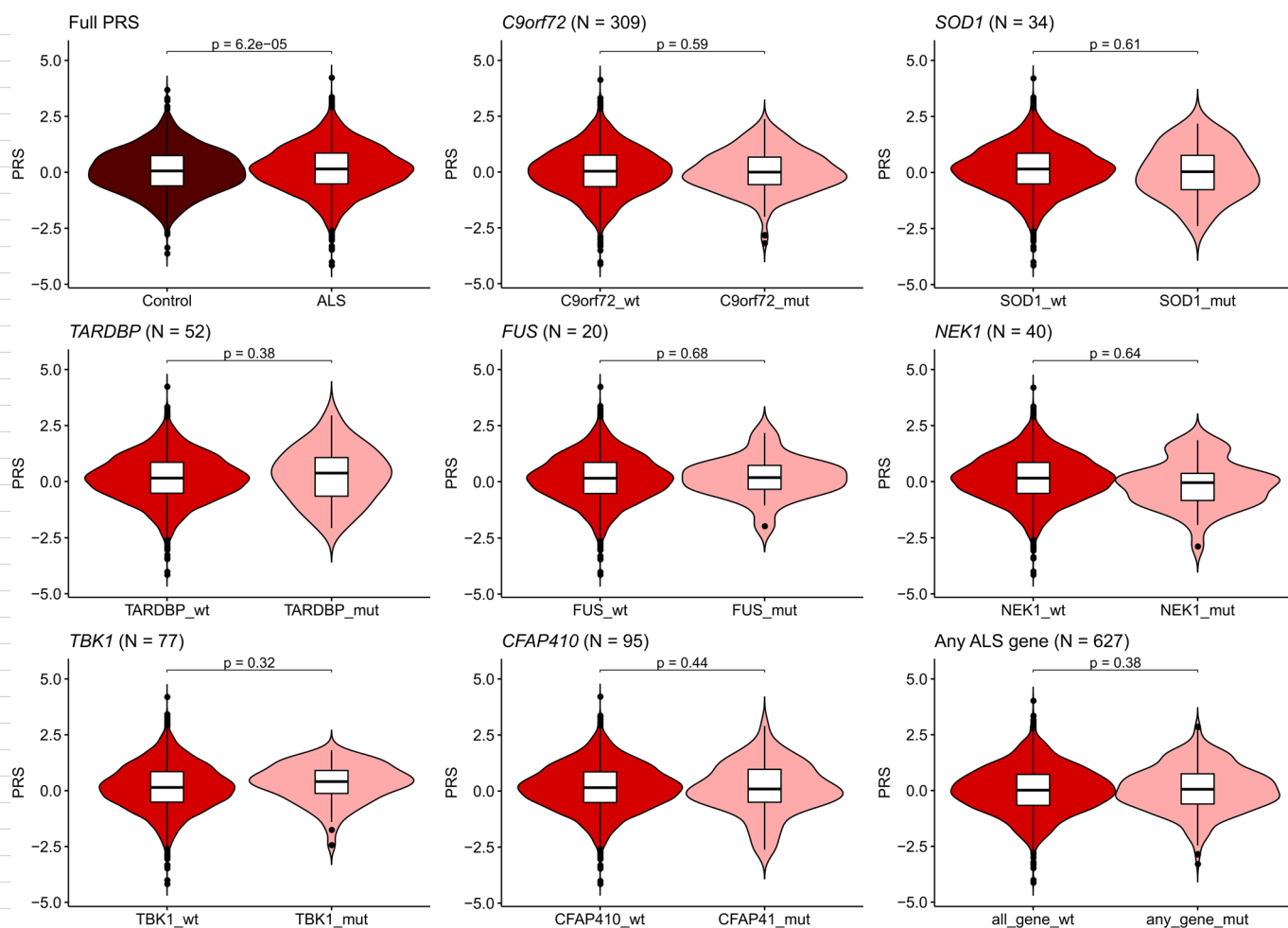


Extended Data Fig. 1 | Manhattan plot in European ancestries GWAS. Genome-wide association statistics obtained by inverse-variance weighted meta-analysis of the stratified SAIGE logistic mixed model regression in European ancestry cohorts. Y-axis corresponds to the two-tailed $-\log_{10}(P)$ -value, x-axis corresponds to the genomic coordinates (GRCh37). Loci containing a genome-wide significant SNP are highlighted in red. SNP IDs are the top associated SNPs in each locus. The dotted horizontal line reflects the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

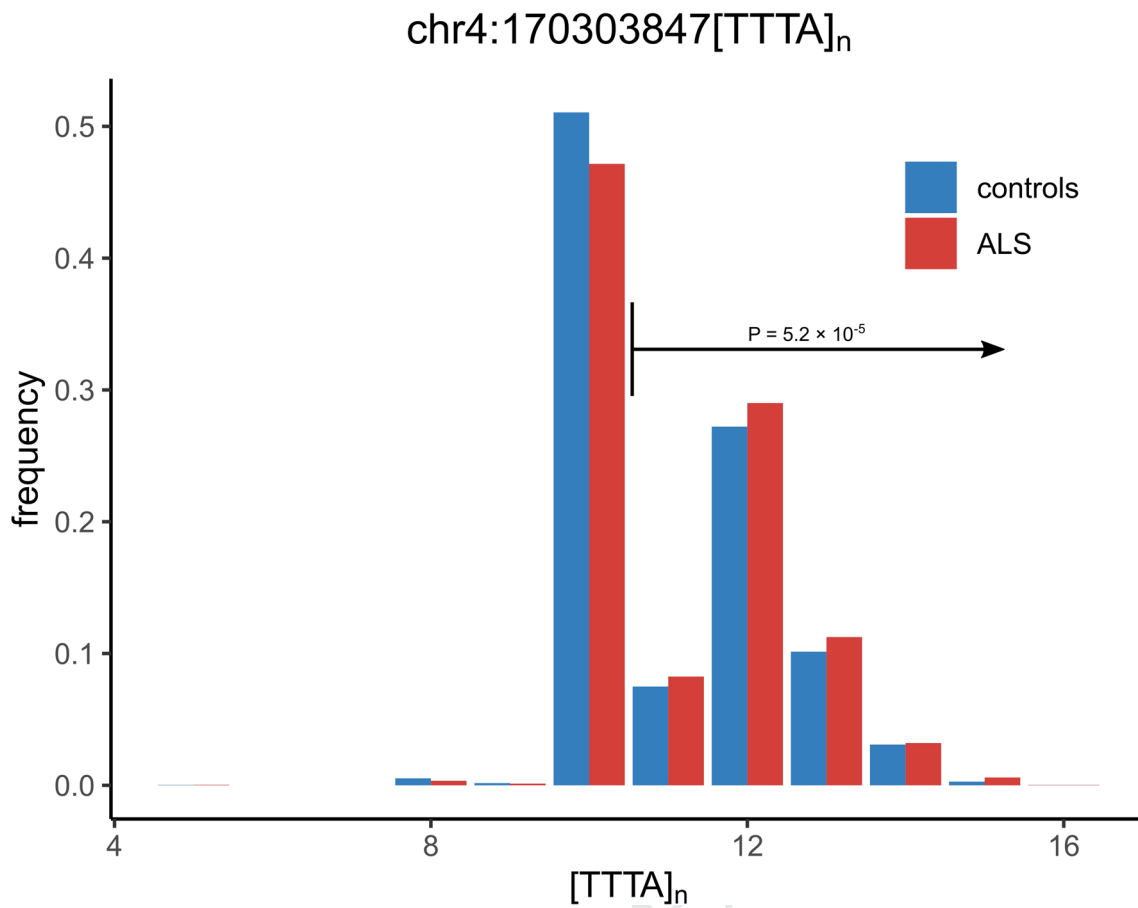


Extended Data Fig. 2 | Annotation specific heritability enrichment. Enrichment of SNP-based heritability was calculated with LD-score regression. Grey dashed line represents no enrichment (enrichment = 1). Error bars denote standard error of enrichment estimate. Nominal statistically significant enrichment estimates (two-sided $P < 0.05$) are marked with an asterisk (*Conserved_LindbladToh* $P = 6.5 \times 10^{-5}$, *SuperEnhancer_Hnisz* $P = 0.014$, *TFBS_ENCODE* $P = 0.017$, *H3K4me1_peaks_Trynka* $P = 0.018$, *Coding_UCSC* $P = 0.028$, *H3K9ac_Trynka* $P = 0.037$). The category *Conserved_LindbladToh* was significant after Bonferroni correction for multiple testing across all categories ($N = 28$). Due to the regression framework in LDSC, enrichment estimates < 0 are possible (with large standard errors).

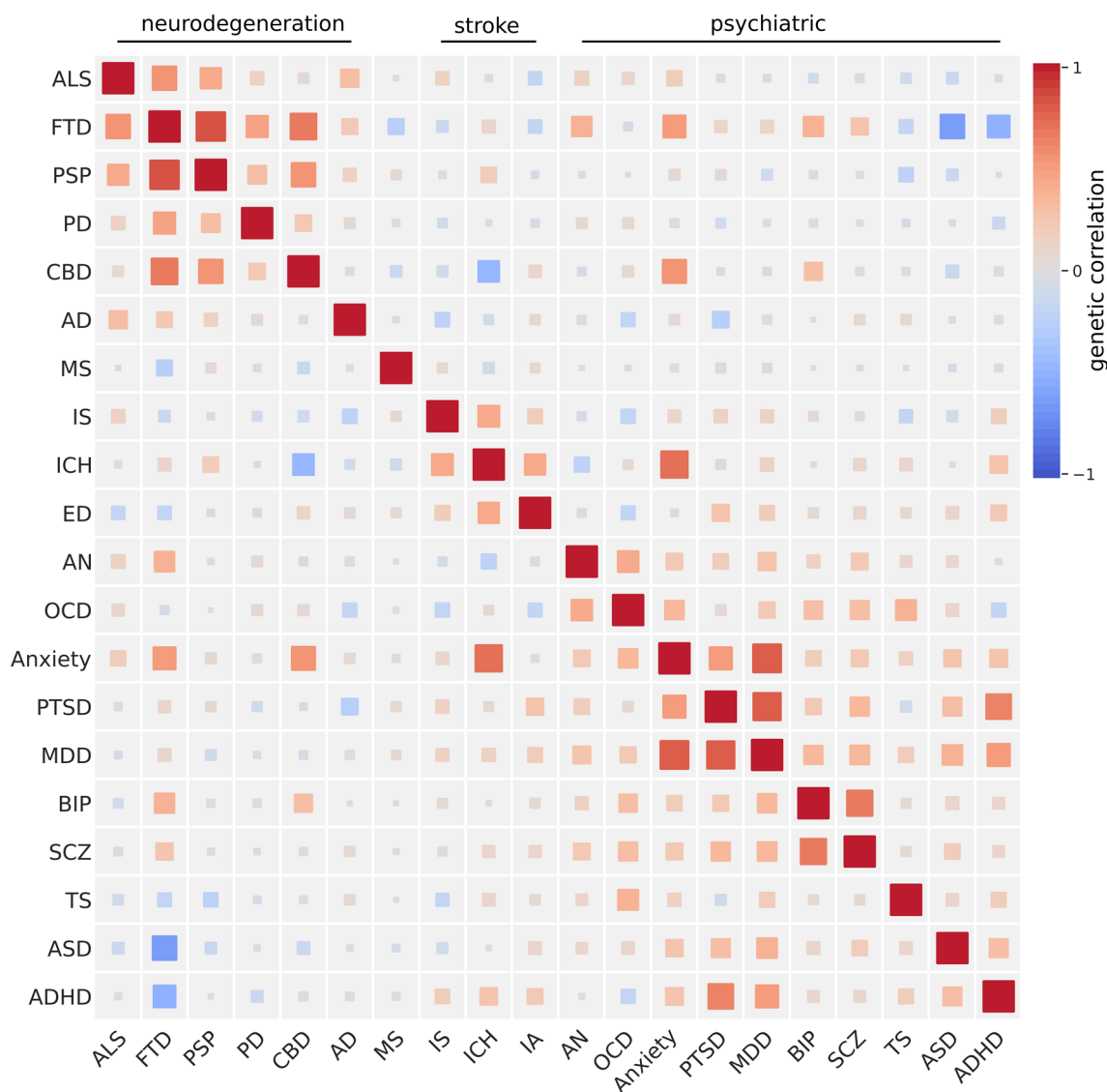
Uncorrected



Extended Data Fig. 3 | Polygenic risk scores stratified by rare variant carrier status. Distribution of polygenic risk scores in controls and ALS patients with or without one or more rare variants in ALS risk genes. There was no statistically significant difference in PRS between ALS patients with and without rare variants in ALS risk genes (labeled as *gene_mut* or *gene_wt* respectively). In total, 5,112 ALS patients and 2,132 controls from stratum 6 with whole-genome sequencing data available were included. For *SOD1*, *TARDBP*, *FUS*, *NEK1*, *TBK1*, and *CFAP410*, rare variants were included according to the model that yielded the strongest association in the rare variant burden association analyses. For *C9orf72*, patients with the pathogenic hexanucleotide repeat expansion were compared to those without the expansion. The 'any ALS gene' groups all patients together with a rare variant in any of the ALS risk genes. P-values for difference in PRS were derived by two-tailed logistic regression. The number of ALS patients carrying a rare variant per gene is denoted in the corresponding panel. Intervals for boxplots: center = median, box = lower and upper quartile, hinges = median $\pm 2 \times$ IQR, IQR = interquartile range.

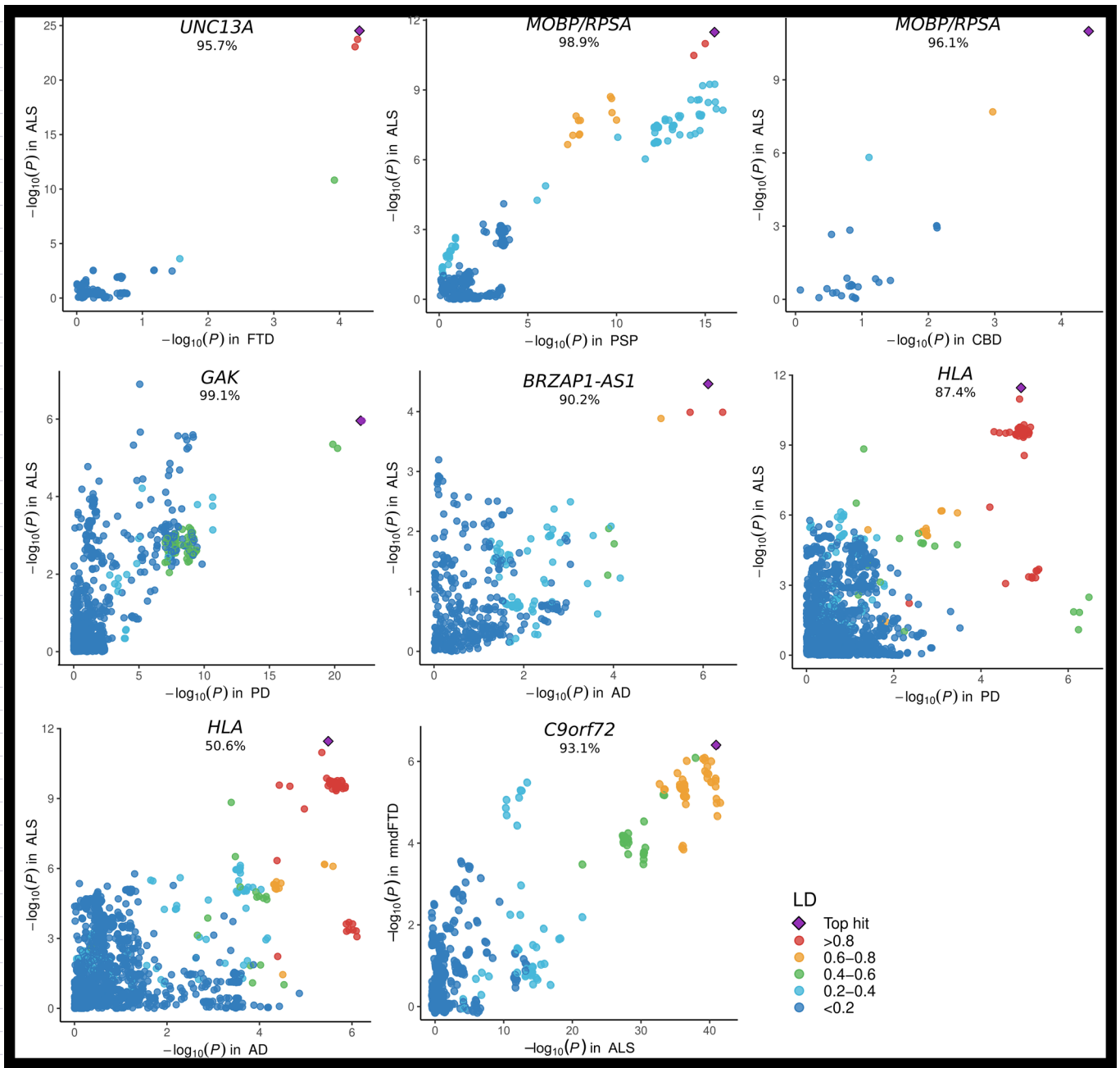


Extended Data Fig. 4 | NEK1 repeat distribution. The frequency of STR alleles in ALS cases and controls are shown. A repeat length of 11 and longer was used as the optimal threshold for disease-associated genotype. The P-value was calculated by Firth logistic regression and FDR correction over all possible thresholds. Y-axis shows the allele frequency of repeat lengths. Repeat position on GRCh37, and repeat motif are shown.



Extended Data Fig. 5 | Genetic correlations between brain diseases. Correlation matrix for genetic correlation estimates obtained from bivariate LD score regression. Colors correspond to genetic correlation estimates. Strongest clusters appear between neurodegenerative diseases and within the psychiatric traits. ALS = amyotrophic lateral sclerosis, FTD = frontotemporal dementia, PSP progressive supranuclear palsy, PD = Parkinson's disease, CBD = corticobasal degeneration, AD = (clinically diagnosed) Alzheimer's disease, MS = multiple sclerosis, IS = ischemic stroke (any), ICH = intracerebral hemorrhage, IA = intracranial aneurysm (any), AN = anorexia nervosa, OCD = obsessive compulsive disorder, Anxiety = anxiety disorder (score), PTSD = post-traumatic stress disorder, MDD = major depressive disorder, BIP = bipolar disorder, SCZ = schizophrenia, TS = Tourette's syndrome, ASD = autism spectrum disorder, ADHD = attention-deficit hyperactivity disorder.

1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317

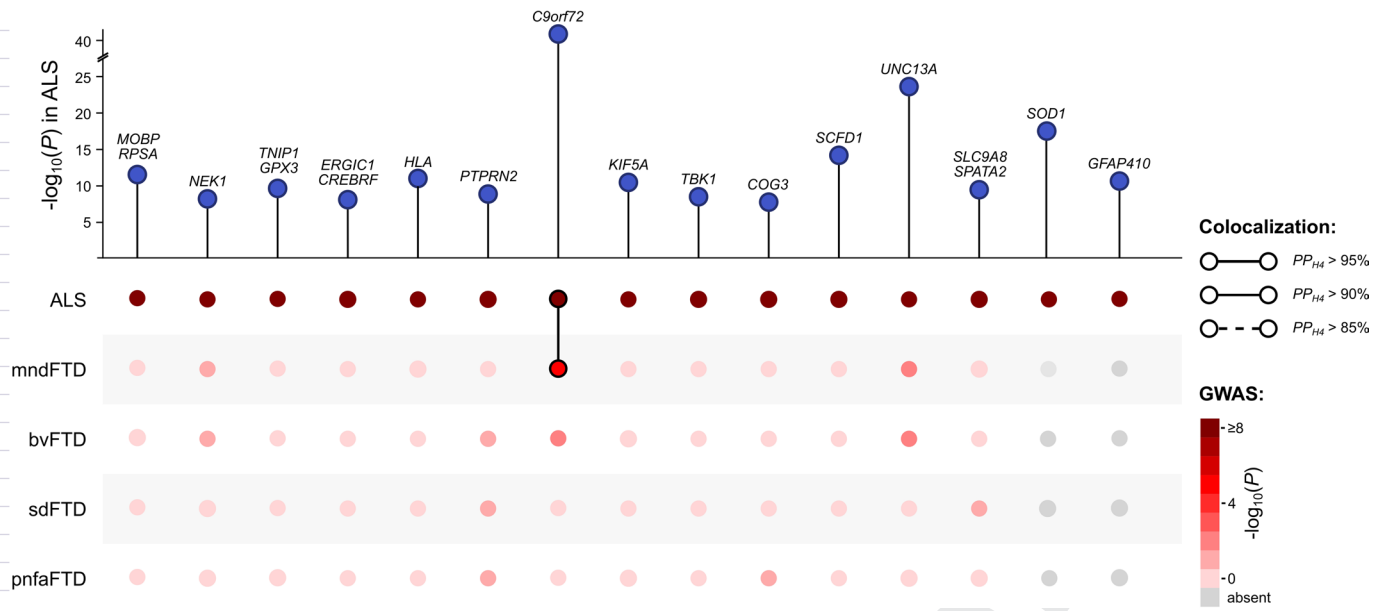


Extended Data Fig. 6 | Colocalization signals. Loci were selected for colocalization analysis when the top associated SNP was associated with any neurodegenerative disease at 5×10^{-5} . For ALS, the European ancestries meta-analysis was used. Bayesian posterior probabilities for a shared variant driving risk of both traits (PP_{H4}) are reported below locus names. Colors reflect LD between the variant and top associated SNP.

A

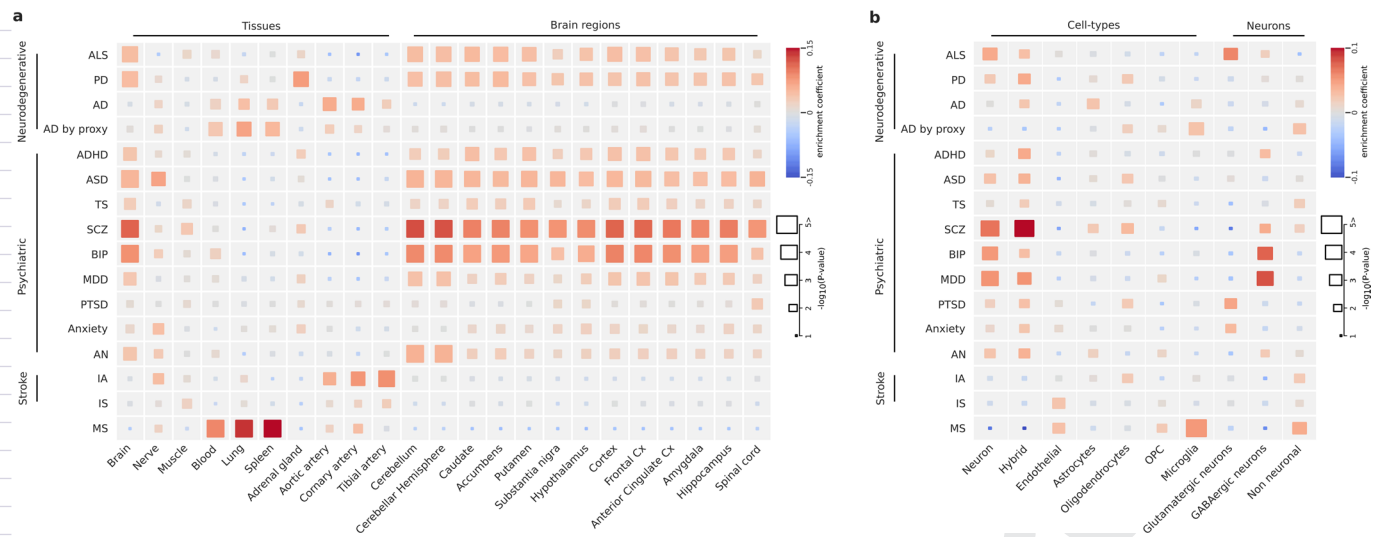
B

1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383



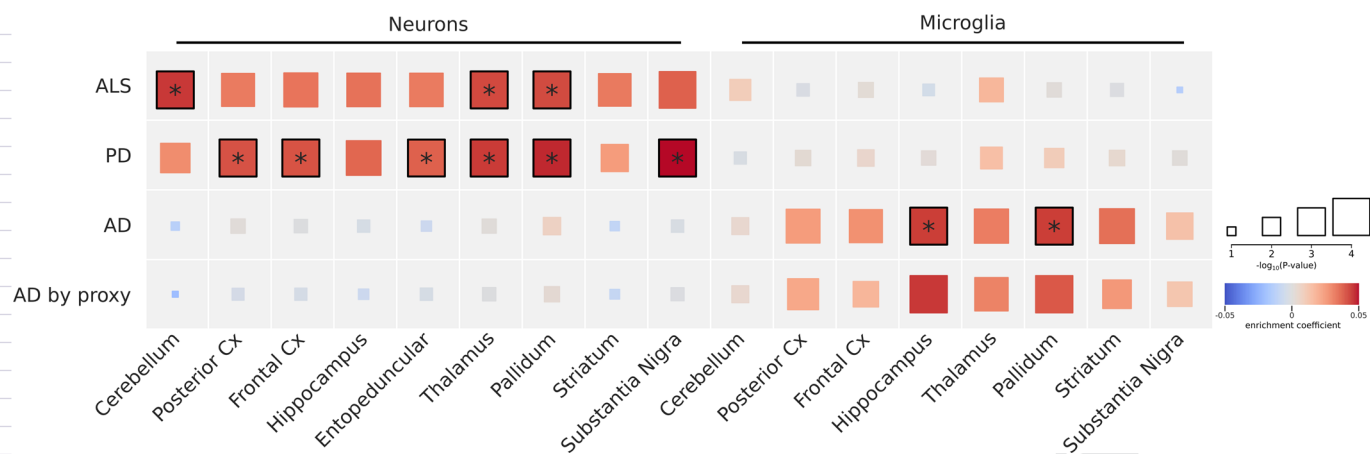
Extended Data Fig. 7 | Colocalization analysis with FTD subtypes. Top associated SNPs in the ALS GWAS were selected for colocalization analysis between ALS and FTD subtypes using COLOC. In the top panel, point height is the two-sided $-\log_{10}$ (P-value) of the top-associated SNP in the ALS GWAS. In the bottom panel, association P-values of these SNPs with FTD subtypes are shown by color. The Bayesian posterior probability for a shared causal variant between traits (PP_{H4}) is depicted by a connection between points.

1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449

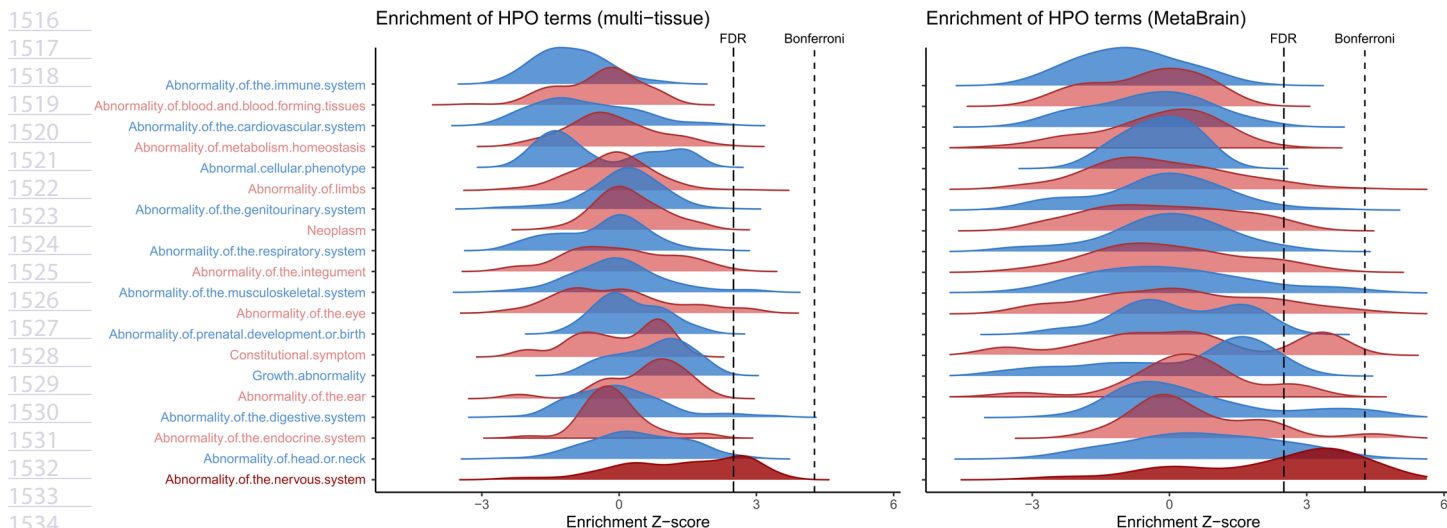


Extended Data Fig. 8 | Tissue and cell-type enrichment analyses for all brain diseases. Tissue **(a)** and cell-type **(b)** enrichment for all included brain diseases obtained from two-sided MAGMA linear regression. Only brain diseases with exome-wide significant gene-based MAGMA associations ($P < 2.7 \times 10^{-6}$) were suitable for tissue and cell-type enrichment analyses. The color represents enrichment coefficient and size indicates two-sided $-\log_{10}(P\text{-value})$ of enrichment obtained by the linear regression model in the MAGMA gene-property analysis. Due to the large number of significant genes in the gene-based MAGMA analyses for schizophrenia, bipolar disorder and multiple sclerosis the enrichment p-values were truncated at $P < 1.0 \times 10^{-5}$. ALS = amyotrophic lateral sclerosis, PD = Parkinson's disease, AD = Alzheimer's disease, ADHD = attention-deficit hyperactivity disorder, ASD = autism spectrum disorder, TS = Tourette's syndrome, SCZ = schizophrenia, BIP = bipolar disorder, MDD = major depressive disorder, PTSD = post-traumatic stress disorder, Anxiety = anxiety disorder (score), AN = anorexia nervosa, IA intracranial aneurysm (any), IS = ischemic stroke, MS = multiple sclerosis, Cx = cortex, OPC = oligodendrocyte progenitor cells.

Uncorrected



Extended Data Fig. 9 | Cell-type enrichment analysis in mice. Cell-type enrichment analysis using the DropViz single-cell RNA sequencing dataset obtained from mice. Similar to the cell-type enrichment analyses there is neuron-specific enrichment in ALS and Parkinson’s disease. In Alzheimer’s disease microglia are the most enriched cell-types. The color represents enrichment coefficient and size indicates two-sided $-\log_{10}(P\text{-value})$ of enrichment obtained by the linear regression model in the MAGMA gene-property analysis. Statistically significant enrichments after correction for multiple testing with a false discovery rate (FDR) < 0.05 are marked with an asterisk. ALS = amyotrophic lateral sclerosis, PD = Parkinson’s disease, AD = Alzheimer’s disease, Cx = cortex.



Extended Data Fig. 10 | Human phenotype ontology term enrichment. Downstreamer enrichment analyses were performed using the multi-tissue and brain-specific co-expression matrix to identify co-regulated ALS-genes. The distribution of enrichment statistics (Z-scores) for all Human phenotype ontology (HPO) terms are plotted per HPO parent branch. The multi-tissue analysis indicates enrichment for the neurology parent branch 'abnormality of the nervous system' (dark-red), although no term passes the Bonferroni threshold for multiple testing. The brain-specific analysis illustrates stronger enrichment for the neurology parent branch. In total, 58 HPO terms pass the threshold for multiple testing of which 42 are defined within the 'abnormality of the nervous system' branch.

Uncorrected Proof

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 973]
Author	Wouter van Rheenen

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Nature of Query
Q1:	Please confirm or correct details for affiliation 93.
Q2:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q3:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q4:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).
Q5:	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.
Q6:	If applicable, please ensure that any accession codes and datasets whose DOIs or other identifiers are mentioned in the paper are scheduled for public release as soon as possible, we recommend within a few days of submitting your proof, and update the database record with publication details from this article once available.
Q7:	Your paper has been copy edited. Please review every sentence to ensure that it conveys your intended meaning; if changes are required, please provide further clarification rather than reverting to the original text. Please note that formatting (including hyphenation, Latin words, and any reference citations that might be mistaken for exponents) has been made consistent with our house style.
Q8:	In the sentence beginning 'We observed moderate inflation of the test' and elsewhere in the text, please confirm whether SE can be replaced with s.e.m. (standard error of the mean).
Q9:	In the legend for Fig. 1, please provide a definition for SAIGE.
Q10:	The sentence beginning 'Interestingly, we observed a genome-wide significant' has been edited to remove the priority claim, according to style. Please check and confirm.
Q11:	In the sentence beginning 'The HEIDI outlier test', please include a definition for HEIDI.
Q12:	In the sentence beginning 'Polygenic risk score (PRS) analyses did not illustrate', please confirm whether polygenic risk score can be abbreviated as PRS throughout the text.

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 973]
Author	Wouter van Rheenen

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Nature of Query
Q13:	In the sentence beginning 'Furthermore, colocalization analyses identified', please include official gene symbols for <i>BZRAP</i> and <i>AS1</i> .
Q14:	In the sentence beginning 'We selected for each gene the protein-coding transcripts', please confirm whether it is ok to delete 'that exhibited the strongest association with ALS' as this is already stated once in the sentence. In the next sentence, please confirm whether it is correct to say 'individuals with overlapping'.
Q15:	Please check that all funders have been appropriately acknowledged and that all grant numbers are correct.
Q16:	In the author list, there are multiple authors with the initials A.A., A.B., A.C., A.F., A.R., B.I., C.C., C.G., C.M., D.B., D.F., E.T., G.B., H.-J.W., J.G., L.T., M.B., M.D., M.G., M.M.N., M.Z., P.C., P.D., R.H.B., R.R., S.C., S.M., S.P. and V.D. Please clarify in the Author contributions and Competing interests at each instance which individual is meant by including the full last name.
Q17:	Please check that the Competing Interests declaration is correct as stated. If you declare competing interests, please check the full text of the declaration for accuracy and completeness.
Q18:	If ref. 24,29,30,36 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

The following software packages have been used for data analyses: R v3.6.3 with additional packages tidyverse v1.3.0, data.table v1.14.0, ggplot2 v3.3.3, MASS v7.3.53, SNPRelate v1.26.0, logistf v1.24, coloc v5.1.0, twoSampleMR v0.5.6, RadialMR v1.0, MVMR v0.3, survival v3.1.8, coxme v2.2.16, survminer v0.4.9 (www.r-project.org), Python v3.7 with additional modules pandas v1.1.3, numpy v1.18.1, scipy v1.4.1, CpGtools v1.0.9, matplotlib v3.1.3, pyliftover v0.4, pyhpo v2.5.0 (www.anaconda.org), GenomeStudio v2.0 (<https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html>), GCTA v1.93.2beta (cns.genomics.com/software/gcta), EIGENSOFT v6.1.4 (www.github.com/DreichLab/EIG), SNPTEST v2.5.4-beta3 (<https://www.well.ox.ac.uk/~gav/snpctest/>), PLINK v1.9 (www.cog-genomics.org/plink2), Michigan Imputation Server (<https://imputationserver.sph.umich.edu>), EAGLE v2.3 through Michigan Imputation Server (<https://imputationserver.sph.umich.edu>), SAIGE v0.29.1 (www.github.com/weizhouUMICH/SAIGE), METAL 2011-03-25 (<https://genome.sph.umich.edu/wiki/METAL>), SnpSift 4.3p, (<https://pcingola.github.io/SnpEff>), ANNOVAR version 2017-07-17 for LRT, Polyphen-2, MutationTaster2, Mutation Assessor, PROVEAN and SIFT (<https://annovar.openbioinformatics.org/>), Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), MutationTaster2, (<http://www.mutationtaster.org/>), Mutation Assessor release 3 (<http://mutationassessor.org/r3/>), PROVEAN v1.1 (<http://provean.jcvi.org/index.php>), SIFT v6.2.1 (<https://sift.bii.a-star.edu.sg/>), SnpEff 4.3p (<https://pcingola.github.io/SnpEff>), LDSC v1.0.1 (www.github.com/bulik/ldsc), ExpansionHunter v4 (www.github.com/Illumina/ExpansionHunter), ExpansionHunter denovo, (www.github.com/Illumina/ExpansionHunterDenovo), SMR (www.cns.genomics.com/software/smr), MAGMA v1.6 (www.ctg.cncr.nl/software/magma), FUMA (<https://fuma.ctglab.nl/>), FUMA Cell-type (<https://fuma.ctglab.nl/celltype>), summary-BayesR (<https://cns.genomics.com/software/gctb/#SummaryBayesianAlphabet>), S-PrediXcan (<https://github.com/hakyimlab/MetaXcan>), TWAS (<http://gusevlab.org/projects/fusion/>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GWAS summary statistics generated in this study are publicly available in the NHGRI-EBI GWAS Catalog (accession IDs: GCST90027163 and GCST90027164 for cross-ancestry and European ancestries meta-analyses respectively) and through the Project MinE website (<https://www.projectmine.com/research/download-data/>). Summary statistics of the rare variant burden analyses and eQTL/mQTL summary-based Mendelian randomization analyses are available through the Project MinE website.

The following publicly available datasets were used in this project:

WellcomeTrust case-control consortium: www.wtccc.org.uk

dbGaP datasets:

phs000101.v3.p1: NIH Genome-Wide Association Studies of Amyotrophic Lateral Sclerosis
 phs000126.v1.p1: CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)
 phs000196.v1.p1: Genome-Wide Association Study of Parkinson Disease: Genes and Environment
 phs000344.v1.p1: Genome-Wide Association Study of Amyotrophic Lateral Sclerosis in Finland
 phs000336: A Genome-Wide Association Study of Lung Cancer Risk
 phs000346: Genome-wide association study for Bladder Cancer Risk
 phs000789: Collaborative Study of Genes, Nutrients and Metabolites (CSGNM)
 phs000206: Whole Genome Scan for Pancreatic Cancer Risk in the Pancreatic Cancer Cohort Consortium and Pancreatic Cancer Case-Control Consortium (PanScan)
 phs000297: eMERGE Network Study of the Genetic Determinants of Resistant Hypertension
 phs000652: Cohort-Based Genome-Wide Association Study of Glioma (GliomaScan)
 phs000869: Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGES)
 phs000812: The Breast and Prostate Cancer Cohort Consortium (BPC3) GWAS of Aggressive Prostate Cancer and ER- Breast Cancer
 phs000428: Genetics Resource with the Health and Retirement Study
 phs000360.v3: eMERGE Network Genome-Wide Association Study of Red Cell Indices, White Blood Count (WBC) Differential, Diabetic Retinopathy, Height, Serum Lipid Levels, Specifically Total Cholesterol, HDL (High Density Lipoprotein), LDL (Low Density Lipoprotein), and Triglycerides, and Autoimmune Hypothyroidism.
 phs000893.v1: Genome-Wide Association Study of Endometrial Cancer in the Epidemiology of Endometrial Cancer Consortium (E2C2)
 phs000168.v2: National Institute on Aging - Late Onset Alzheimer's Disease Family Study: Genome-Wide Association Study for Susceptibility Loci
 phs000092.v1: Study of Addiction: Genetics and Environment (SAGE)
 phs000864.v1: Genomic Predictors of Combat Stress Vulnerability and Resilience
 phs000170.v2: A Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project Cohort
 phs000431.v2: IgA Nephropathy GWAS on Individuals of European Ancestry (IGANGWAS2)
 phs000237.v1: Northwestern NUGene Project: Type 2 Diabetes
 phs000169.v1: Whole Genome Association Study of Visceral Adiposity in the Health Aging and Body Composition (Health ABC) Study
 phs000982.v1: Genetic Analysis of Psoriasis and Psoriatic Arthritis: GWAS of Psoriatic Arthritis
 phs000289.v2: National Human Genome Research Institute (NHGRI) GENEVA Genome-Wide Association Study of Venous Thrombosis (GWAS of VTE)
 phs000634.v1: National Cancer Institute (NCI) Genome Wide Association Study (GWAS) of Lung Cancer in Never Smokers
 phs000274.v1: Genome-Wide Association Study of Celiac Disease
 phs001172.v1: National Institute of Neurological Disorders and Stroke (NINDS) Parkinson's Disease
 phs000389.v1: GENetics of Nephropathy - an International Effort (GENIE) GWAS of Diabetic Nephropathy in the UK GoKinD and All-Ireland Cohorts
 phs000460.v1: Genetics of 24 hour urine composition
 phs000138.v2: GWAS for Genetic Determinants of Bone Fragility in European-American Premenopausal Women
 phs000394.v1: Autopsy-Confirmed Parkinson Disease GWAS Consortium (APDGC)
 phs000948.v1: Genetic Discovery and Application in a Clinical Setting: Continuing a Partnership (eMERGE Phase II)
 phs000630.v1: Exome Chip Study of NIMH Controls
 phs000678.v1: A Family-Based Study of Genes and Environment in Young-Onset Breast Cancer
 phs000351.v1: National Cancer Institute Genome-Wide Association Study of Renal Cell Carcinoma
 phs000314.v1: Genetic Associations in Idiopathic Talipes Equinovarus (Clubfoot) - GAIT
 phs000147.v3: Cancer Genetic Markers of Susceptibility (CGEMS) Breast Cancer Genome-wide Association Study (GWAS) - Primary Scan: Nurses' Health Study - Additional Cases: Nurses' Health Study 2
 phs000882.v1: National Cancer Institute (NCI) Prostate Cancer Genome-Wide Association Study for Uncommon Susceptibility Loci (PEGASUS)
 phs000238.v1: National Eye Institute Glaucoma Human Genetics Collaboration (NEIGHBOR) Consortium Glaucoma Genome-Wide Association Study
 phs000397.v1: National Institute on Aging (NIA) Long Life Family Study (LLFS)
 phs000421.v1: A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy (FECD)
 phs000142.v1: A Whole Genome Association Scan for Myopia and Glaucoma Endophenotypes using Twin Studies
 phs000303.v1: Genetic Epidemiology of Refractive Error in the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) Study
 phs000125.v1: CIDR: Collaborative Study on the Genetics of Alcoholism Case Control Study
 phs001039.v1: International Age-Related Macular Degeneration Genomics Consortium - Exome Chip Experiment
 phs000187.v1: High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation
 phs000101.v5: Genome-Wide Association Study of Amyotrophic Lateral Sclerosis
 phs002068.v1.p1: Sporadic ALS Australia Systems Genomics Consortium (SALSA-SGC)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not pre-specify a sample size given the wide distribution of allele-frequencies and effect-sizes typically seen in GWAS. We have included all largest available control cohorts of European ancestries matched for genotyping platform that were available through dbGaP to achieve a ~1:10 case:control ratio per stratum at maximum. This ratio was roughly determined based on power calculations (https://zdz.bwh.harvard.edu/gpc/cc2.html) that indicated that including even more controls would yield a limited increase in power and we expected increasing challenges introduced by batch effects when more smaller control cohorts were included.
Data exclusions	Individuals and genotypes were excluded from the analysis following rigorous quality control as described in the methods section.
Replication	We replicated our SNP associations in 2 independent GWAS in ALS patients and control subjects from Asian ancestries. All genome-wide significant SNPs showed an identical direction of effect. Given our effort to design a large-scale GWAS including all available individual level genotype data in ALS globally (including the newly genotyped ALS patients), there are no more independent datasets for replication in European ancestries.
Randomization	For newly genotyped case-control cohorts, samples were randomized by case-control status before hybridization on SNP genotyping arrays. For case-only and control-only cohorts (Supplementary table 1) samples could not be randomized before hybridization. We therefore matched these cohorts based on genotyping platform and included Illumina genotyping arrays only. We subsequently corrected for genotyping platform as confounder by the stratified analyses creating 6 separate strata. Furthermore, principal components and a genetic relationship matrix were included as covariates in the statistical analyses to correct for structure in the data due to technical artifacts and population stratification. We assessed residual confounding of test-statistics by LD-Score Regression.
Blinding	Individuals involved in sample ascertainment were blinded for genotypes, individuals involved in genotyping individuals were blinded for phenotypes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- Antibodies
 - Eukaryotic cell lines
 - Palaeontology and archaeology
 - Animals and other organisms
 - Human research participants
 - Clinical data
 - Dual use research of concern

Methods

- n/a Involved in the study
- ChIP-seq
 - Flow cytometry
 - MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Included were 8 strata of ALS patients and controls in the GWAS. The analyses were stratified for genotyping platform and reported ancestries: stratum 1: 2,254 ALS patients, 11,155 controls, IlluminaCoreExome, European ancestries stratum 2: 1,458 ALS patients, 2,043 controls, Illumina 317K, European ancestries stratum 3: 1,701 ALS patients, 2,555 controls, Illumina 370K, European ancestries stratum 4: 3,394 ALS patients, 42,402 controls, Illumina550K/Illumina610K/Illumina660K, European ancestries stratum 6: 14,402 ALS patients, 32,094 controls, IlluminaOmniExpress/Illumina2M, European ancestries stratum 7: 3,996 ALS patients, 20,632 controls, IlluminaGSA, European ancestries stratum 8: 1,234 ALS patients, 2850 controls, IlluminaHumanOmniZhongHua, Chinese ancestries stratum 9: 1,173 ALS patients, 8,925 controls, IlluminaHumanOmniExpressExome, Japanese ancestries A more detailed description is provided in Supplementary Table 1 (numbers stratified by cohort) and the Supplementary Text (for newly genotyped individuals). The European ancestries for individuals in stratum 1-7 were inferred from PCA (Supplementary Figure 18).
Recruitment	The Supplementary Text describes recruitment for each newly genotyped or sequenced cohort.
Ethics oversight	Local ethics committee of the Medical Faculty of Friedrich Schiller University Jena, Jena, Germany. Ethical Committee of Città della Salute Hospital, Torino, Italy.

Institutional Review Board of the Azienda Sanitaria Locale, Lecce, Italy.
Ethics Committee, Stockholm, Sweden
University Medical Center Utrecht Medical Ethics Committee, Utrecht, The Netherlands.
Review Ethics Board Office at McGill University Health Center, Montreal, Canada
Medical Research Ethics Committee of "Assistance Publique-Hôpitaux de Paris", Paris, France.
Ethics committee of Tours Hospital, Tours, France.
Ethics committee of Limoges University Hospital, Limoges, France.
National Medical Ethics Committee of Republic of Slovenia, Slovenia.
Sydney South West Area Health Service Human Research Ethics Committee, Australia.
Human Research Ethics Committee at the QIMR Berghofer Medical Research Institute, Australia.
Charité Universitätsmedizin, Berlin Medical Ethics Committee, Berlin, Germany.
Medical Ethics Committee of Hannover Medical School, Hannover, Germany.
Beaumont Hospital Research & Ethics Committee, Dublin, Ireland.
Ethics Committee of the IRCCS Istituto Auxologico Italiano, Milan, Italy.
Local ethical committee of Buyanov city hospital, Moscow, Russia.
Ethics Committee of the School of Medicine, University of Belgrade.
Yorkshire and the Humber - Sheffield Research Ethics Committee, UK.
Institutional review board of Cedars-Sinai, Los Angeles, USA.
Institutional review board of the University of California at Los Angeles, USA.
Trent University Medical Ethics Committee, UK.
Ethics Committee on Research with Human Participants (INAREK) at Bogazici University, Istanbul, Turkey.
Ethical Committee of University Hospital Leuven, Leuven, Belgium.
"Comité de Ética de la Investigación del Hospital Carlos III", Madrid, Spain.
Bellvitge University Hospital Ethics Committee, Barcelona, Spain.
Committee for the Protection of Human Subjects in Research of the University of Massachusetts Medical School, Worcester, USA.
Regional Ethical Review Board in Umeå, Sweden.
Hadassah University Hospital IRB board, Hadassah, Israel.
The Institutional Review Board of Tel Aviv Sourasky Medical Center, Tel Aviv, Israel.
The Local Research Ethics Committee at the Faculty of Medicine, University of Lisbon, Lisbon, Portugal.
Kantonale Ethikkommission des Kantons St. Gallen, Switzerland

Note that full information on the approval of the study protocol must also be provided in the manuscript.