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 onset\(^1\). 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 cases\(^1\). 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 ALS\(^5\), 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 GWAS 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 FTD\(^19\). Further expanding the ALS–FTD spectrum, a genetic correlation with progressive supranuclear palsy (PSP) has been described\(^20\). 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 resource\(^21\). We observed moderate inflation of the test statistics (\(\lambda_{GC} = 1.12\), \(\lambda_{HRC} = 1.003\), and linkage disequilibrium (LD) score regression yielded an intercept of 1.029 (s.e. = 0.0073), indicating that the majority of inflation was due to the polygenic signal in ALS (LD score regression (LDSC): \(h_I^2 = 0.028\), s.e. = 0.003, \(K = 350^{-1}\), \(P = 5.5 \times 10^{-21}\)). The European ancestry analysis identified 12 loci reaching genome-wide significance (\(P < 5 \times 10^{-8}\); 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)\(^3,16\), and all showed a consistent direction of effects (\(P_{\text{Asian}} = 2.0 \times 10^{-3}\)). 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 risk in European and Asian ancestries resulted in a trans-ancestry genetic correlation of 0.57 (s.e. = 0.28) for genetic effect and 0.58 (s.e. = 0.30) for genetic impact, which were not statistically significantly different from unity (\(P = 0.13\) and \(P = 0.16\), respectively).

---

A full list of affiliations appears at the end of the paper.
We therefore performed a cross-ancestry meta-analysis totaling 29,612 cases and 122,656 controls, which revealed three additional loci, totaling 15 genome-wide significant risk loci for ALS risk (Fig. 1, Table 1 and Supplementary Tables 4–18). Conditional and joint analysis did not identify secondary signals within these loci.

Of these findings, eight loci have been reported in previous GWASs (C9orf72, UNC13A, SCFD1, MOBP–RPSA, KIF5A, CFA4P10, GPX3–TNIP1 and TBK1)\(^{1,2,11,13}\). The rs80265967 variant corresponds to the p.D90A mutation in SODI previously identified in a Finnish ALS cohort enriched for familial ALS\(^{11}\). Interestingly, we observed a genome-wide significant common variant association signal within the NEK1 locus, which was previously shown to harbor rare variants associated with ALS\(^{8}\). The recently reported association at the ACSL5–ZDHHC6 locus\(^{11,12}\) did not reach the threshold for genome-wide significance (rs58854276, \(P_{\text{LR}} = 5.4 \times 10^{-4}, P_{\text{ES}} = 4.9 \times 10^{-4}, P_{\text{hub}} = 6.5 \times 10^{-4}\); Supplementary Table 19), despite the fact that our analysis includes all data from the original discovery studies.

**Rare variant gene-based association analyses in ALS.** To assess a general pattern of underlying architectures that link associated SNPs to causal genes, we first tested for annotation-specific enrichment using stratified LDSC. This revealed that 5′ UTR regions as well as coding regions in the genome and those annotated as conserved were most enriched for ALS-associated SNPs (Extended Data Fig. 2). Subsequently, we investigated how rare, coding variants contributed to ALS risk by generating a whole-genome sequencing (WGS) dataset of patients with ALS (n = 6,538) and control participants (n = 2,415), which is a subset of the common variant GWAS cohort. The exome-wide association analysis included transcript-level rare variant burden testing for different models of allele-frequency thresholds and variant annotations (Methods). This identified NEK1 as the strongest associated gene (minimal \(P = 4.9 \times 10^{-4}\) for disruptive and damaging variants at \(\text{MAF} < 0.005\)), which was the only gene to pass the exome-wide significance thresholds (0.05 \(\times 17,994 = 2.8 \times 10^{-4}\) and 0.05 \(\times 58,058 = 8.6 \times 10^{-7}\) for number of genes and protein-coding transcripts, respectively; Supplementary Table 20). This association was independent from the previously reported increased rare variant burden in selected patients with familial ALS (ref. \(^{1}\)) who were not included in this study. Polygenic risk score (PRS) analyses did not illustrate a difference in PRSs in patients carrying rare variants in ALS-risk genes (SOD1, C9orf72 repeat expansion, TARDBP, FUS, NEK1, TBK1 and CFA4P10) compared to all patients with ALS (Extended Data Fig. 3). Although power was limited, this is compatible with a scenario in which the genetic risk of ALS in these patients is a sum of rare variants in ALS genes and other (common) genetic variation.

**Gene prioritization shows locus-specific underlying architectures.** To assess whether rare variant associations could drive the common variant signals at the 15 genome-wide significant loci, we combined the common and rare variant analyses to prioritize genes within these loci. The SNP effects on gene expression were assessed using stratified LDSC. This revealed that 5′ UTR regions as well as coding regions in the genome and those annotated as conserved were most enriched for ALS-associated SNPs (Extended Data Fig. 2). Here, eQTL and mQTL analyses indicated that the risk-increasing effects of the common variants were mediated 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, \(P = 5.2 \times 10^{-14}\); false discovery rate (FDR) = 4.7 \(\times 10^{-14}\); 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 significant 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. 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\(^{26,30}\). Furthermore, we could not prioritize a specific gene in the HLA locus (rs9275477; Supplementary Fig. 5).

**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...
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 PD ($r = 0.44$, s.e. $= 0.11$, $P = 1.0 \times 10^{-5}$) as previously reported and also revealed a significant genetic correlation between ALS and AD ($r = 0.31$, s.e. $= 0.12$, $P = 9.6 \times 10^{-5}$) as well as between ALS and PD ($r = 0.16$, s.e. $= 0.061$, $P = 0.011$; Fig. 3a). The point estimate for the genetic correlation between ALS and FTD was high ($r = 0.59$, s.e. $= 0.41$, $P = 0.15$) 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, PD and corticobasal degeneration (CBD) as well as a shared signal in the UNCI3A locus between ALS and FTD (posterior probability, $PP_{\text{uni}} > 95$; Extended Data Fig. 6). For the HLA locus, there was evidence for a shared causal variant between ALS and PD ($PP_{\text{uni}} = 88$%) but no conclusive evidence for ALS and AD ($PP_{\text{uni}} = 51$% for a shared causal variant and $PP_{\text{uni}} = 49$% 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 ($rs43311866$, $PP_{\text{uni}} = 99$%) and between ALS and AD at the TSPOAP1–AS1 locus ($rs2635216$, $PP_{\text{uni}} = 90$%). Note, the association at TSPOAP1–AS1 was not genome-wide significant in the GWAS of clinically diagnosed AD ($P = 3.7 \times 10^{-5}$) either but was identified in the larger AD-by-proxy GWAS. For FTD subtypes, C9orf72 showed a colocalization signal for a shared causal variant between ALS and the motor neuron disease subtype of FTD (mndFTD, $PP_{\text{uni}} = 93$%; Extended Data Figs. 6 and 7).

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

| Chr | Position(bp) | ID | Prioritized gene | A1 | A2 | Freq | P | s.e. | Effect (s.e.) |
|-----|--------------|----|-----------------|----|----|------|---|-----|--------------|
| 9   | 27,563,868   | rs2453555 | A1 A2 | 0.248 | 0.174 | 0.013 | 1.0×25 | 0.074 | 0.038 | 0.053 | 0.120 | 0.012 |
| 19  | 17,752,689   | rs12608932 | SCFD1 | 0.337 | 0.091 | 0.012 | 9.2 | 10 | 0.002 | 0.036 | 0.97 | 0.083 | 0.011 |
| 11  | 31,045,181   | rs229194a | 11 | 0.084 | 0.036 | 0.020 | 0.080 | 0.011 | 3.4 | 0.110 | 0.111 | 0.32 | 0.142 | 0.02 |
| 3   | 39,508,968   | rs631312 | MOBP | 0.397 | 0.079 | 0.013 | 3.5 | 10 | 0.067 | 0.074 | 0.37 | 0.065 | 0.011 |
| 5   | 150,410,835  | rs10463311 | TNIP1 | 0.253 | 0.079 | 0.013 | 3.5 | 10 | 0.096 | 0.079 | 0.32 | 0.142 | 0.02 |
| 12  | 64,877,053   | rs4075094 | 12 | 0.098 | 0.018 | 0.098 | 1.7 | 10 | 0.098 | 0.018 | 0.098 | 0.018 | 3.5 |
| 13  | 46,113,984   | rs2985994 | COG3 | 0.259 | 0.066 | 0.013 | 1.9 | 10 | 0.098 | 0.018 | 0.098 | 0.018 | 3.5 |

Details of two indexed single nucleotide polymorphism (SNP) loci within each genome-wide significant locus (P < 5 × 10^{-5}) for the top-associated SNP in the respective GWAS. The low frequency (MAF ≤ 0.01) and major allele frequency < 1.5% for all SNPs. In each SNP table, P values are provided for each SNP, with a star indicating genome-wide significance (P ≤ 5 × 10^{-8}). The effect allele is indicated by an ‘A’; non-effect allele, by a ‘C’.
Brain-specific coexpression networks improve detection of ALSD-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 datasets or RNA-seq data from brain cortex samples. To validate this approach, we first tested for enrichment of human phenotype ontology (HPO) terms that are related to ALS or neurodegenerative diseases in general, including ‘cerebral cortical atrophy’ ($P = 1.8 \times 10^{-7}$), ‘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 no enriched HPO terms 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).

MR analyses are in line with a causal relationship between cholesterol levels and ALS. From previous observational case–control studies and our blood-based methylome-wide study, numerous calculated using MAGMA with gene expression patterns from the Genotype–Tissue Expression (GTEx) project (version 8) in a gene set enrichment analysis using FUMA. 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 reported.

Enrichment for neurons but not for microglia or astrocytes. 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 datasets or RNA-seq data from brain cortex samples. 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’ ($P = 1.8 \times 10^{-7}$), ‘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’, $P = 4.2 \times 10^{-4}$, ‘intra-Golgi and retrograde Golgi-to-endoplasmic reticulum (ER) trafficking’, $P = 1.4 \times 10^{-7}$) and autophagy (‘macrowautophagy’, $P = 3.2 \times 10^{-3}$) 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).

MR analyses are in line with a causal relationship between cholesterol levels and ALS. From previous observational case–control studies and our blood-based methylome-wide study, 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 framework. 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 (weighted median β = 0.15, s.e. = 0.04, P = 3.2 × 10^{-4}; Fig. 5a and full results in Supplementary Table 27). These results were robust to removal of outliers through radial MR analysis, 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 participants 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 ALS risk (inverse-variance-weighted β_{IVW} = 2.0 × 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

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.

ARTICLES

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 framework. 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 (weighted median β = 0.15, s.e. = 0.04, P = 3.2 × 10^{-4}; Fig. 5a and full results in Supplementary Table 27). These results were robust to removal of outliers through radial MR analysis, 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 participants 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 ALS risk (inverse-variance-weighted β_{IVW} = 2.0 × 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 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 as well as PD and PSP with locus-specific patterns of shared genetic risk across all neurodegenerative diseases. Co-localization analysis identified two additional loci, GAK and TSPOAP1-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. 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 TBKI, 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 TSPoA1-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.47–49. 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.50–52.

Based on observational studies and MR analyses, conflicting evidence exists for lipid levels including cholesterol as a risk factor for ALS.44–50. 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 study showing methylation changes associated with increased cholesterol levels in ALS.50. 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.48,51. Interestingly, recent work has shown that lipid metabolism and autophagy are closely related, 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 autophagy.52. 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 AD.55. 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; Published online: 6 December 2021

References

1. van Es, M. A. et al. Amyotrophic lateral sclerosis. Lancet 390, 2084–2098 (2017).
2. 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).
3. Trabjerg, B. B. et al. ALS in Danish registries: heritability and links to psychiatric and cardiovascular disorders. Neurol. Genet. 6, e389 (2020).
4. Ryan, M., Heverin, M., McLaughlin, R. L. & Hardiman, O. Lifetime risk and heritability of amyotrophic lateral sclerosis. JAMA Neurol. 76, 1367–1374 (2019).
5. 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).
6. Circulli, E. T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 347, 1436–1441 (2015).
7. Freischmidt, A. et al. Haplotypinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat. Neurosci. 18, 631–636 (2015).
8. Kenna, K. P. et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat. Genet. 48, 1037–1042 (2016).
9. Brenner, D. et al. NEK1 mutations in familial amyotrophic lateral sclerosis. Brain 139, e28 (2016).
10. 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).
11. Nicolas, A. et al. Genome-wide analyses identify KIF5A as a novel ALS gene. Neuron 97, 1268–1283 (2018).
12. 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).
13. Laaksovirta, H. et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol. 9, 978–985 (2010).
14. van Rheenen, W. et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. 48, 1043–1048 (2016).

NATURE GENETICS | VOL 53 | DECEMBER 2021 | 1636-1648 | www.nature.com/naturegenetics

1643
15. Benyamin, B. et al. Cross-ethnic meta-analysis identifies association of the GPX3–TNPD1 locus with amyotrophic lateral sclerosis. *Nat. Commun.* 8, 611 (2017).

16. Nakamura, R. et al. A multi-ethnic meta-analysis identifies novel genes, including ACRS5, associated with amyotrophic lateral sclerosis. *Commun. Biol.* 3, 526 (2020).

17. Dejesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9orf72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256 (2011).

18. Renton, A. E. et al. A hexanucleotide repeat expansion in C9orf72 is the cause of chromosome 9p21-linked ALS–FTD. *Neuron* 72, 257–268 (2011).

19. Diekstra, F. P. et al. C9orf72 and UC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide meta-analysis. *Ann. Neurol.* 76, 120–133 (2014).

20. Chen, J. A. et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* 13, 41 (2018).

21. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283 (2016).

22. Iaconangi, A. et al. Genome-wide meta-analysis finds the ACRS5–ZDHHC6 locus is associated with ALS and links weight loss to the disease genetics. *Cell Rep.* 33, 108323 (2020).

23. Võsa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53, 1300–1310 (2021).

24. de Klein, N. et al. Brain expression quantitative trait loci and network analysis reveals downstream effects and putative drivers for brain-related diseases. Preprint at bioRxiv https://doi.org/10.1101/2021.03.01.433439 (2021).

25. Pidsley, R. et al. Critical evaluation of the Illumina MethylationEPIC beadchip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 17, 208 (2016).

26. Shireby, G. L. et al. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain* 143, 3763–3775 (2020).

27. Hannon, E. et al. An integrated genetic–epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* 17, 176 (2016).

28. Fang, X. et al. The NEK1 interactor, C21ORF2, is required for efficient DNA damage repair. *Acta Biochim. Biophys. Sin.* 47, 834–841 (2015).

29. Brown, A.-L. et al. Common ALS/FTD risk variants in UC13A exacerbate its cryptic splicing and loss upon TDP-43 mislocalization. Preprint at bioRxiv https://doi.org/10.1101/2021.04.02.438213 (2021).

30. Ma, X. R. et al. TDP-43 represses cryptic exon inclusion in FTD/ALS gene UC13A. Preprint at bioRxiv https://doi.org/10.1101/2021.04.02.438213 (2021).

31. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51, 404–413 (2019).

32. Leeuw, C. A., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. MAGMA: predictive gene–phenotype associations using large-scale gene expression analysis. *Nat. Commun.* 10, 2837 (2019).

33. Hopp, 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).

34. Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 362, k601 (2018).

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

Ettore Beghi42, Elisabetta Pupillo42, Giancarlo Comi45,46,147, NilO Riva45, Christian Lunetta48, Francesca Gerardi48, Maria Sofia Cotelli49,150, Fabrizio Rinaldi149, Luca Chiveri151, Maria Cristina Gauita152, Patrizia Perrone152, Mauro Ceroni153, Luca Diamanti153, Carlo Ferrarese154, Lucio Tremolizzo154, Maria Luisa Delodovici155 and Giorgio Bono155

PARALS Consortium

Adriano Chiò40,41, Andrea Calvo40,41, Cristina Moglia40,41, Antonio Canosa40,41,156, Umberto Manera40, Rosario Vasta40, Alessandro Bombaci40, Maurizio Grassano40, Maura Brunetti40, Federico Casale40, Giuseppe Fuda40, Paolina Salamone40, Barbara Iazzolino40, Laura Peotta40, Paolo Cugnasco40, Giovanni De Marco41, Maria Claudia Torrieri40, Francesca Palummo40, Salvatore Gallone41, Marco Barberis157, Luca Sbaiz157, Salvatore Gentile158, Alessandro Mauro40,159, Letizia Mazzini160,161, Fabiola De Marchi160,161, Lucia Corrado161,162, Sandra D'Alfonso161,162, Antonio Bertolotto163,
Maurizio Gionco, Daniela Leotta, Enrico Oddodenino, Daniele Imperiale, Roberto Cavallo, Pietro Pignatta, Marco De Mattei, Claudio Geda, Diego Maria Papareullo, Graziano Gusmaroli, Cristoforo Comi, Carmelo Labate, Luigi Ruiz, Delfina Ferrandi, Eugenia Rota, Marco Aguglia, Nicoletta Di Vito, Piero Meineri, Paolo Ghiglione, Nicola Launaro, Michele Dotta, Alessia Di Sapio and Guido Giardini.

SLAGEN Consortium

Vincenzo Silani, Nicola Ticozzi, Antonia Ratti, Isabella Fogh, Cinzia Tiloca, Silvia Peverelli, Cinzia Gellera, Giuseppe Lauria, Franco Taroni, Viviana Pensato, Barbara Castellotti, Giacomo P. Comi, Stefania Corti, Roberto Del Bo, Cristina Cereda, Mauro Ceroni, Stella Gagliardi, Sandra D’Alfonso, Lucia Corrado, Letizia Mazzini, Gianni Sorari, Flavia Raggi, Gabriele Siciliano, Costanza Simoncini, Annalisa Lo Gerfo, Massimiliano Filosto, Maurizio Inghilleri and Alessandra Ferlini.

SLAP Consortium

Giancarlo Logroscino, Ettore Beghi, Isabella L. Simone, Bruno Passarella, Vito Guerra, Stefano Zoccoli, Cecilia Nozzoli, Ciro Mundi, Maurizio Leone, Michele Zarrelli, Filippo Tamma, Francesco Valluzzi, Gianluigi Calabrese, Giovanni Boero and Augusto Rini.
GWAS quality control and imputation. For each cohort, we first performed individual- and variant-level quality control, after which cohorts were merged into six data sets based on platform and genotyping platform. The platform-strain-variant-quality control was performed, and strata were imputed up to the HaploType Reference Consortium panel (r.1.1 2016) through the Michigan Imputation Server. Full quality-control details are described in the Supplementary Note and Supplementary Table 1. Numbers of individuals and variants passing each quality-control step are described in Supplementary Table 2.

Association testing and meta-analysis. After quality control, a null logistic mixed model was fitted using SAIGE (version 0.9.9) SNPs pruned with PLINK (version 1.9) (–indep-pairwise 50 25 0.1) in a leave-one-chromosome-out scheme. Subsequently, a SNP-wise logistic mixed model including the saddlepoint approximation test was performed using genotype dosages with SAIGE. Association statistics for all strata were combined in an IVW fixed-effects meta-analysis using METAL. Genomic inflation factors were calculated per stratum and for the full meta-analysis. To assess any residual confounding due to population stratification and artificial structure in the data, we calculated the LDS intercept using SNP LD scores calculated in the HapMap3 CEU population.

Cross-ancestry analyses. GWAS summary statistics from two Asian ancestry studies were obtained (http://www.1000genomes.org). These summary statistics were meta-analyzed with all European ancestry data in a platform-strain-specific, variant-strain-specific meta-analysis framework. Cross-ancestry meta-analysis was performed using the Popcorn software. The regression model (–use_regression) was used to estimate genetic correlation. We calculated both the correlation of genetic effects (correlation of allelic effect sizes) and genetic impact (correlation of allelic effect size adjusted for difference in allele frequencies).

Conditional SNP analysis. Conditional joint SNP analysis (COJO, GCTA version 1.91.1b) was performed to identify potential secondary GWAS signals within a single locus. SNPs with association P ≤ 5 × 10^-8 were considered. Controls of European ancestry from the Health and Retirement Study (HRS, cohort 65, Supplementary Table 1), included in stratum 4 of this study, were used as the LD reference panel.

Gene prioritization. Whole-genome sequencing. Sample selection, sequencing and data preparation. Participants with ALS and control participants from Project MinE were recruited for WGS. The participating cohorts were not pre-screened for ALS-associated mutations and are described in the Supplementary Note. In total, 228 patients were known to have at least one first- or second-degree relative with ALS. A full description of Project MinE and the sequencing and quality-control pipeline were described previously. In summary, the first batch of 2,250 cases and control samples was sequenced on the Illumina HiSeq 2000 platform. All remaining 7,350 case and control samples were sequenced on the Illumina HiSeq X platform. All samples were sequenced to approximately 150× coverage with 100-bp reads and ~25× coverage with 150-bp reads for HiSeq 2000 and HiSeq X, respectively. Both sequencing sets used PCR-free library preparation. Samples were also genotyped on the Illumina 2.5M SNP platform, then aligned to GRCh37 using the UCSC Aligner, and variants were called using the UCSC variant caller; both the aligner and caller are standard to Illumina’s aligning and calling pipeline. Full details of individual- and variant-level quality control are described in the Supplementary Note.

Genetic burden association analyses. To analyze rare variants in a genetic burden test framework, we used a variety of variant filters to allow for different genetic architectures of ALS-associated variants per gene as we and others did previously. In summary, variants were annotated according to allele-frequency threshold (MAF < 0.01 or MAF < 0.005) and predicted variant impact (‘missense’, ‘damaging’, ‘disruptive’). ‘Disruptive’ variants were those variants classified as frameshift, splice site, exon loss, stop gain, start loss and transcriptional ablation by SnpEff. ‘Damaging’ variants were missense variants predicted to be damaging by seven prediction algorithms (SIFT, PolyPhen-2 (ref. 31), LRT, MutationTaster2 (ref. 32), MutationsAssessor and PROVEAN). ‘Missense’ variants were those missense variants that did not meet the ‘damaging’ criteria. All combinations of allele-frequency threshold and variant annotations were used to test the genetic burden on a transcript level in a Firth logistic regression framework in which burden was defined as the number of variants per individual. Sex and the first allele were included as covariates. All Ensembl-annotated transcripts for which at least five individuals had a non-zero burden were included in the analysis.

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

Short tandem repeat screen. For all individuals who had sequencing results in the HiSeq X dataset (5,392 cases, 1,795 controls), we screened all loci harboring SNPs associated with ALS meeting genome-wide significance for expansions of known and new STRs using ExpansionHunter and ExpansionHunter Denovo. For all individuals who used ExpansionHunter version 4.0 to screen for expansions of known STRs located within 1 Mb of the top ALS-associated SNP, we used the STRs identified from indels in 18 high-quality genomes and the GangSTR catalog based on STR annotations in the reference genome. We excluded all homopolymers from these catalogs. Repeat length was subsequently re-estimated on case–control status using Firth logistic regression including the first 20 PCs as covariates, recoding the STR size to a biallelic variant using a sliding window over all observed repeat lengths. To correct for multiple testing across all possible thresholds, we applied Benjamini–Hochberg correction per STR.

To screen for extremely long STR expansions (similar to the C9orf72 expansion at loci that were not included in the predefined STR catalog) we used ExpansionHunter Denovo. This method aims to only find STR expansions that exceed the sequencing read length (>150 bp) by identifying reads (mapped, mismapped and unmapped) that contain STR motifs, using their mate pairs for de novo mapping to the reference genome.

For all STRs, we calculated LD statistics (r² and |D'|) between recoded genotypes at the optimal threshold and the top locus as the associated GWAS SNP. Subsequently, we conditioned the SNP association on the repeat genotype in a Firth logistic regression.

Summary-based Mendelian randomization. We used multi-SNP SMR79–81 to infer the underlying gene expression variation on ALS using summary-based Mendelian randomization (a SNP with expression of a gene) on ALS risk. We chose to apply SMR because this method yielded very similar results when compared to S-PrediXcan and TWA55 (Supplementary Fig. 18) when applied using GTEx version 7 eQTL, and it can be applied to the large relevant eQTL datasets (MetaBrain and eQTLGen) without access to individual-level genotype and gene expression data. MetaBrain is a harmonized set of 8,727 RNA-seq samples from seven regions of the central nervous system from 15 datasets, and we selected eQTL derived from the cortex region of the brain in samples of European ancestry (MetaBrain Cortex-EUR eQTL, n = 2,970 individuals, n = 6,601 RNA-seq samples) as our instrument variable. European-only ALS summary statistics were used as the outcome.

To implement this analysis, we also used eQTL in blood from the eQTLGen Consortium, as this is a large available eQTL resource. Samples of European ancestry in the HRS (cohort 65 of this GWAS) were used as the LD reference panel. SNPs with MAF ≥ 1% in the HRS were included. Further SMR settings were left as default, meaning probes with at least one eQTL with P ≤ 5 × 10^-8 were included.

We subsequently performed SMR using DNA mQTL data and European-only ALS summary statistics. Human prefrontal cortex and whole-blood DNA mQTL were generated as part of ongoing analyses by the Complex Disease Epigenomics Group at the University of Exeter using the Illumina EPIC HumanMethylation array that quantifies DNAm at >850,000 sites across the genome. The prefrontal cortex mQTL dataset was generated using DNA methylation and SNP data from 522 individuals from the Dementia Research Cohort and includes 4,623,966 cis mQTL (distance between quantitative trait loci SNP and DNA site ≤500 kb) from 1,744,102 SNPs and 43,357 DNA methylation sites. The whole-blood mQTL dataset was generated using DNAm and SNP data from 2,082 individuals and included 30,432,023 cis mQTL between 4,030,902 SNPs and 167,854 DNA methylation sites. mQTL
Articles

Genetic risk factor 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 Analysis (GCTA) and obtained 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 ancestry. 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 and the number of rare coding mutations in ALS-risk genes (SOD1, TARDBP, FUS, NEK1, TBK1 and C9orf101) 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 PCI–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 including sex, site of onset, country, GWAS stratum and PCI–PC5 as covariates.

Cross-trait analyses. Datasets and data preparation. GWAS summary statistics for clinically diagnosed AD, PD, FTD, CBD and PSP 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-design, 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 high, 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-trait 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 Haploype 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 Consortium. For cerebrovascular diseases, GWAS summary statistics were obtained for ischemic stroke (any ischemic stroke)13, intracerebral hemorrhage14 and intracranial aneurysms15. For psychiatric traits, GWAS summary statistics were obtained from Psychiatric Genomics Consortium studies on anorexia nervosa16, obsessive–compulsive disorder17, anxiety disorders (anxiety score)18, post-traumatic stress disorder (all European ancestries)19, major depressive disorder19, bipolar disorder20, schizophrenia21, Tourette’s syndrome22, autism spectrum disorder23 and attention–deficit hyperactivity disorder (European ancestries)22.

Genetic correlation. Genome-wide genetic correlation between neurodegenerative traits was calculated using LDSC (version 1.0.0)24. Precomputed LD scores of European individuals in the 1000 Genomes Project for high-quality HapMap3 SNPs were used (http://ping.igc.cns.tw). 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., 0.0075); AD, 1.03 (s.e., 0.013); PD, 0.98 (s.e., 0.0065); PSP, 1.05 (s.e., 0.0076); CBD, 0.98 (s.e., 0.0073); FTD, 1.00 (s.e., 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 × 10−5, we performed colocalization analysis using the ‘coloc’ package in R25. We set the prior probabilities to π1 = 1 × 10−4, π2 = 1 × 10−4 and π12 = 1 × 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)24 to calculate SNP-based heritability, the LDSC intercept and SNP-based heritability enrichment for psychiatric traits in the genome. We used genomic control to correct for population stratification. For colocalization analysis across 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 FUMA26 software version 1.3.6a. FUMA performs a genomic association analysis of GWAS summary statistics 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 analyses, we used human-derived single-cell RNA-seq data on major brain cell types (GSE78335 without status samples27). Allen Brain Atlas data were used for the human-derived major neuronal subtypes and the DropViz28 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 software29 to identify enriched biological pathways and processes. First, gene–based association statistics were obtained with the Pascall method30, 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. 31) 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 resource32. 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 statistics32. 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 ‘pypo’ module in Python, all HPO terms were assigned to their parent term(s) in the ‘phenotypic anatomy’ (HP:0001118) 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 behavior-related traits: body mass index (anthropometric)33, years of schooling (educational attainment)34, age of smoking initiation and cigarettes per week of vigorous activity from the UK Biobank35; four blood pressure traits (coronary artery disease36, stroke37, diastolic blood pressure and systolic blood pressure37); seven immune system traits from Vuckovic et al.38 (basophil, eosinophil, lymphocyte, monocyte, neutrophil and white blood cell counts) and C-reactive protein39; and four lipid traits from Willer et al.40 (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 were used to serve as instruments for MR analyses were selected at two different P-value cutoffs (P < 5 × 10−5 and P < 5 × 10−8) 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 (r2 > 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 (P-statistic31 × 10−10;
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 study: the Wellcome Trust Case Control Consortium (https://www.wtcc.org.uk/) and dbGaP datasets (phs001011.v3.p1, NIH Genome-Wide Association Studies of Amyotrophic Lateral Sclerosis; phs00126.v1.p1, CERU: Genome Wide Association Study in Familial Parkinson Disease (PD); phs00196.v1.p1, Genome-Wide Association Study of Parkinson Disease: Genes and Environment; phs00344.v1.p1, Genome-Wide Association Study of Amyotrophic Lateral Sclerosis in Finland; phs0036.v1, a Genome-Wide Association Study of Lung Cancer Risk; phs00346, Genome-Wide Association Study of Bladder Cancer Risk; phs0078, Collaborative Study of Genes, Nutrients and Metabolites (CSGNM); phs00206, Whole Genome Scan for Pancreatic Cancer Risk in the Pancreatic Cancer Cohort Consortium and Pancreatic Cancer Case–Control Consortium (PanScan); phs00297, eMERGE Network Study of the Genetic Determinants of Resistant Hypertension; phs00367, a Cohort-Based Genome-Wide Association Study of Glomia (GlomiaScan); phs00869, Barrett’s and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGESS); phs00812, the Breast and Prostate Cancer Cohort Consortium (BPC3) GWAS of Aggressive Prostate Cancer and ER- Breast Cancer; phs00428, Genetics Resource with the HHS; phs00306.v3, eMERGE Network Genome-Wide Association Study of Red Cell Indices and 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; phs00893.v1, Genome-Wide Association Study of Endometrial Cancer in the Epidemiology of Endometrial Cancer Consortium (EEC2); phs00168.v2, National Institute on Aging—Late Onset Alzheimer’s Disease Family Study: Genome-Wide Association Study for Susceptibility Loci; phs00092.v1, Study of Addiction: Genetics and Environment (SAGE); phs00864.v1, Genetic Predictors of Combat Stress Vulnerability and Resilience; phs00170, a Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project: Cohort b2, IGA Nephropathy GWAS on Individuals of European Ancestry (IGANGWAS); phs00257.v1, Northwestern NUgene Project; Type 2 Diabetes; phs00169.v1, Whole Genome Association Study of Visceral Adiposity in the Health Aging and Body Composition (Health ABC) Study; phs00982.v1, Genetic Analysis of Psoriasis and Psoriatic Arthritis: GWAS of Psoriatic Arthritis; phs00208.v2, National Human Genome Research Institute (NHGRI) Genomic Medicine of Type 2 Diabetes: The Type 2 Diabetes (T2D)-GWAS (T2D-GWAS) of VTE; phs00634.v1, National Cancer Institute (NCI) Genome Wide Association Study (GWAS) of Lung Cancer in Never Smokers: phs00724.v1, Genome-Wide Association Study of Celiac Disease; phs001172.v1, National Institute of Neurological Disorders and Stroke (NINDS) Parkinson’s Disease; phs00257.v1, Genetic Nephropathy—An International Effort (GENIE) GWAS of Diabetic Nephropathy in the UK GoKinD and All-Ireland Cohorts; phs00460.v1, Genetics of 24 Hour Urine Composition; phs00138.v2, GWAS for Genetic Determinants of Bone Fragility in European–American Prenomenopausal Women; phs00394.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; phs00078.v1, a Family-Based Study of Genes and Environment in Young Onset Breast Cancer (YOBaC); phs000314.v1, Genetic Association of Renal Cell Carcinoma; phs003014.v1, Genetic Associations in Idiopathic Telipos Equinovarosus (Clubbfoot)—GAIT; phs00147.v3, Cancer Genetic Markers of Susceptibility (CGEMS) Breast Cancer Genome-wide Association Study (GWAS)—Primary Scan: Nurses’ Health Study—Additional Case-Only Analysis: phs00142.v1, National Prostate Cancer Genome-Wide Association Study for Uncommon Susceptibility Loci (PAGESUS); phs00238.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); phs00421.v1, a Genome-Wide Association Study of Fuchs’ Endothelial Corneal Dystrophy (FECD); phs00142.v1, a Whole Genome Association Scan for Myopia and Glaucoma Endophenotypes using Twin Studies; phs00303.v1, Genetic Epidemiology of Refractive Error in the KORA (Kooperatieve Gesundheitsforschung in der Region Augsburg) Study; phs00125.v1, CDDR Collaborative Study on the Genetics of Alcoholism Case Control Study; phs001039.v1, International Age-Related Macular Degeneration Genomics Consortium—Exome Chip Experiment; phs00187.v1, High Density SNP Association Analysis of Melanoma: Case–Control and Outcomes Investigation; phs001011.v5, Genome-Wide Association Study of Amyotrophic Lateral Sclerosis; phs00268.v1, 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, MASON version 7.3.53, SNPRelate version 1.2.60, logitSIF version 1.24, coloc version 5.1.0, twoSampleMR version 0.5.6, RadialMR version 1.0, MMR version 0.5.3, survival version 2.16 and survminer version 0.4.9 (https://cran.r-project.org/web/packages/survminer/index.html), Python version 3.7 with additional modules pandas version 1.1.3, numpy version 1.1.18, scipy version 1.4.1, CpGtools version 1.0.9, matplotlib version 3.1.3, pyrophost version 0.4 and pyplot version 2.5.0 (https://ancanoda.org/), GenomeStudio version 2.0 (https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html), OCTA version 1.93.2beta (https://csgenomics.com/software/OCTAViewer), EIG (https://github.com/max-trentini/DreichLab/EIG), SNPTEST version 2.5.4-beta3 (https://www.well.ox.ac.uk/~gav.snpTEST/), PLINK version 1.9 (https://www.cog-genomics.org/plink2), the Michigan Imputation Server (https://imputationserver.sph.umich.edu), EAGLE version 2.3 through the Michigan Imputation Server (https://imputationserver.sph.umich.edu), METAL 11.03-25 (https://geneph.scripps.edu/wiki/METAL), SNPdiff 4.3p (https://pcgingola.github.io/SnpEff/), ANNOVAR version 2017-07-17 for LRT, Polyphen-2, MutationTaster2, Mutation Assessor, PROVEAN and SIFT (https://us.apache.org/snpEff/), polyPhen2 (https://genetics.bwh.harvard.edu/ph2/). MutationTaster2 (https://www.mutationtaster.org), Mutation Assessor release 3 (https://mutationassessor.org/r3/), PROVEAN version 1.1 (http://provean.jcvi.org/index.php), SIFT version 6.21.1 (https:// sift.bia-a-star.edu.sg/), SnpEff 4.3p (https://pcgingola.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://nsgenomics.com/software/smrdsgma)/ MAGMA version 1.6 (https://ctcg.cnrc.nf/software/magma), FUMA (https://fuma.ctaglab.nl), FUMA Cell-type (https://fuma.ctaglab.nl/celltype), summary-BayesR (https://nsgenomics.com/software/gcb#SummaryBayesianAlp habet), S-PrediXcan (https://github.com/hakimylab/MetaXcan) and TWAS (https://gusevlab.org/projects/fusion/).

References

56. Brey, R. R., Miller, R. G., Swash, M. & Munsat, T. L. El Escorial revised: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Motor Nervous Disord. 1, 293–299 (2000).
57. 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).
58. Winkler, C. J., Li, Y. & Abecasis, G. R. MET: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
59. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
60. Benjamin, E. J. et al. Transethnic genetic-correlation estimates from summary statistics. Am. J. Hum. Genet. 99, 76–88 (2016).
61. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).
89. Kouri, N. et al. Genome-wide association study of corticobasal degeneration. *Eur. J. Neurol.* 26, 1357–1364 (2018).
90. Spek, R. A. Avander et al. Project MinE database: bringing large-scale whole-genome sequencing in ALS to researchers and the public. *Amyotroph. Lateral Scler. Frontotemporal Degener.* 20, 432–440 (2019).
91. Genovesi, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* 15, 1433–1441 (2016).
92. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92 (2012).
93. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for new coding regions. *Nat. Protoc.* 11, 1–9 (2016).
94. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249 (2010).
95. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561 (2009).
96. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11, 361–362 (2014).
97. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39, e118 (2011).
98. Chot, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2547–2547 (2015).
99. Doolzhenko, E. et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* 27, 1895–1903 (2017).
100. Doolzhenko, E. & ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* 21, 102 (2020).
101. Mousavi, N., Shleizer-Burko, S., Yanicky, R. & Gymrek, M. Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* 47, e90 (2019).
102. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* 9, 918 (2018).
103. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487 (2016).
104. Bartholome, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* 9, 1825 (2018).
105. Gussev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–252 (2016).
106. Hannon, E. et al. Leveraging DNA methylation quantitative-trait loci to characterize the relationship between methylation variation, gene expression, and complex traits. *Am. J. Hum. Genet.* 103, 654–665 (2018).
107. Hop, P. J. et al. Genome-wide identification of genes regulating DNA methylation using genetic anchors for causal inference. *Genome Biol.* 21, 220 (2020).
108. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* 28, 495–501 (2010).
109. Wei, T. et al. CpGtools: a Python package for DNA methylation analysis. *Bioinformatics* 37, 1598–1599 (2021).
110. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* 50, 764–773 (2018).
111. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* 10, 5086 (2019).
112. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer’s disease. *PLoS Genet.* 12, e1004383 (2014).
113. 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).
114. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573, 61–68 (2019).
115. Sudlow, C. et al. Genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* 10, 4558 (2019).
116. Weary, 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).
117. Lamparter, D., Marbach, D., Ruedeii, 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).
118. Ligthart, S. et al. Genome analyses of 30 loci associated with bipolar disorder. *Nat. Genet.* 51, 793–803 (2019).
119. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427 (2014).
120. 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).
121. Greene, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444 (2019).
122. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75 (2019).
123. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383 (2014).
124. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649 (2018).
125. 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, 1122–1121 (2018).
126. 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).
127. 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).
128. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci in breast cancer. *Oncotarget* 6, 1332–1345 (2015).
129. 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).
130. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* 182, 1214–1221 (2020).
131. Ligh'tart, 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).
132. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 45, 1274–1283 (2013).
133. Deng, P., Wang, L., 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).
134. Cragg, J. G. & Donald, S. G. Testing identifiability and specification in instrumental variable models. *Econ. Theory* 9, 222–240 (1993).
Acknowledgements

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

Author contributions

Sample ascertainment: W.v.R., R.A.A.v.d.S., M.M., A.M.D., H.-J. Westeneng, G.H.P.T., N.T., J.C.-K., B.N.S., M. Gromicho, S. Chandran, S. Pal, K.E.M., F.J.S., J.H., R.W.O., M.S., T.M., N.B., A.J.v.d.k., A. Ratti, C. Gelleria, G. Lauria, G.P.C., C.C., D.S., S.D.A., G. Sorariti, G. Siciliano, M.E., A.P., A. Chiò, A. Calvo, C. Moglia, M. Brunetti, A. Canosa, M. Grassano, E.B., E.P., G. Logroscino, B.N., A.O., A.N., Y.L., M. Zafari, M. Gokline, R.H. Baloh, S.B., P. Corcione, P. Couratier, S. Millecamps, V.M., E.S., J.S.M.P., A. Assialouli, R.R.-G., P.A.D., J.P.R., A.C.L., J.H.W., D. Brenner, A. Freischmidt., G. Bensimon, A. Brice, A.D., C.A.M.P., S.S.-D., N.W.W., S.T., R. Rademakers, A. Braun, J.K., D.C.W., C.M.O., A.G.U., A.H., M.R., S. Cichon, M.M. Nöthen, P.A., B.J.T., A.B.S., M. Mitne Neto, R.J.C., R.A.O., M.W.-P., C.L.-H., Y.M.v.d., J.G., A. Roediger, N.G., A.J., T.B., E. Theele, B. Iule., B.S., O.W.W., R.S., C.A.H., C. Graff, L.B., V.F., V. Demeshonok, A. Atalina, B.R., B.K., J.Z., M.R.-G., D.G., Z.S., V. Drory, M.P., J.P.B., 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., S. Pronto, S. Petri, M.W., G.A.R., V.S., J.D.G., R.H. Brown, J.E.L., C.E.S., P.M.A., D. Fan, F.C.G., A.F.M., R.L.M., O.J.H., A.A.-C., F.C.G., L.H.v.d.B., J.H.V, SLACOM Consortium, PARALS Consortium, SLAGEN Consortium and SLAP Consortium. SNP array genotyping: W.v.R., R.A.A.v.d.S., A.M.D., A.S., J.F., G. Bensimon, A. Brice, A.D., C.A.M.P., S.S.-D., N.W.W., L. Tittmann, 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. Nöthen, 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.J.C., G. Breen, D. Fan, 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. Restuadi, R.P.B., M. Doberti, M.H., A.A.K., A.I., G.N.T., G.S., B.B., A. Fan, F.C.G., 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. Restuadi, R.P.B., M. Doberti, M.H., A.A.K., A.I., A.S., J.T., B.N.S., B.B., A. Fan, F.C.G., R.L.M., N.R.W. and J.H.V GWAS: W.v.R., R.A.A.v.d.S., J.J.F.A.v.c.v., P.J.H., R.A.J.Z., M.R.-G., M.H., 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 GWAS quality control: W.v.R., R.A.A.v.d.S., J.J.F.A.v.c.v., P.J.H., R.A.J.Z., M.R.-G., M.H., 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., R.A.A.v.d.S., M.K.B., N.d.K., H.-J. Westra, O.B.B., P.A.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. Westra, O.B.B., P.D., E.J.N.G., M.A.v.E., R.P., A.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. Baird, H.-J. Westra, G.D.S., T.R.G., E. Tsai, H.R. and J.H.V Writing the manuscript: W.v.R., M.K.B., D. Baird, 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.c.v., G.S., E.H., D. Baird, R. Restuadi, E.D., H.-J. Westra, G.H.P.T., K.R.v.E., J.J.F.A., M.A.v.E., R.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.

Competing interests

J.H.V. has sponsored research agreements with Biogen. L.H.v.d.B. receives personal fees from Cytokinetics outside of the submitted work. A.A.-C. has served on scientific advisory boards for Mitsubishi Tanabe Pharma, Orphan Pharma, Biogen, Lilly, GlaxoSmithKline, Amryx, Amylyx and Wave Therapeutics. A. Chiò, serves on scientific advisory boards for Mitsubishi Tanabe, Roche, Biogen, Denali and Cytokinetics. J.E.L. is a member of the scientific advisory board for Cerevel Therapeutics, a consultant for ACI Clinical LLC sponsored by Biogen, Inc. or Ionis Pharmaceuticals, Inc. J.E.L. is also a consultant for Perkins Coie LLP and may provide expert testimony. The remaining authors declare no competing interests related to this work.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00973-1. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00973-1. Correspondence and requests for materials should be addressed to Wouter van Rheenen or Jan H. Veldink.

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

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 × 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).
Extended Data Fig. 3 | PRS stratified by rare variant carrier status. Distribution of PRS 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 \textit{SOD1}, \textit{TARDBP}, \textit{FUS}, \textit{NEK1}, \textit{TBK1}, and \textit{CFAP410}, rare variants were included according to the model that yielded the strongest association in the rare variant burden association analyses. For \textit{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. \textit{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 ± 2 * 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.
Extended Data Fig. 6 | Colocalization signals. Loci were selected for colocalization analysis when the top associated SNP was associated with any neurodegenerative disease at $5 \times 10^{-5}$. For ALS, the European ancestries meta-analysis was used. Bayesian posterior probabilities for a shared variant driving risk of both traits ($PPH4$) are reported below locus names. Colors reflect LD between the variant and top associated SNP.
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 ($PP_{hu}$) is depicted by a connection between points.
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 x 10^{-6}) were suitable for tissue and cell-type enrichment analyses. The color represents enrichment coefficient and size indicates two-sided -log_{10}(P-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 x 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.
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. Downstream 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.
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.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- 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)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- 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

No software was used to collect the data.

Data analysis

The following software packages have been used for data analyses: R v3.6.3 with additional packages tidyr 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, MVVR 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, CppTools v1.0.9, matplolib v3.1.3, pylibfor v0.4, pyhpo v2.5.0 (www.anacoda.org), GenomicStudio v2.0 (https://emma.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html), GCTA v1.93.2beta (cnsengenomics/software/gcta), EIGENSOFT v6.1.4 (www.github.com/OreichLab/EIG), SNPTEST v2.4.4-beta3 (https://www.well.ox.ac.uk/~gav/snpTest/), PLINK v1.9 (www.cnsengenomics.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 (github.com/weizhoulUMICH/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://anovar.openbioinformatics.org/), Polyphen-2 (https://genetics.bwh.harvard.edu/pph2/), MutationTaster2 (http://www.mutationtaster.org/), Mutation Assessor release 3 (http://mutationassessor.org/3/), 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), LSCD v1.0.1 (www.github.com/bulkik/dsc), ExpansionHunter v4 (www.github.com/Illumina/ExpansionHunter), ExpansionHunter Denovo, (www.github.com/Illumina/ExpansionHunterDenovo), SMR (www.cnsengenomics/software/smr), MAGMA v1.6 (www.ctg.mrc.uk/software/magma), FUMA (https://fuma.ctg.mrc.uk/), FUMA Cell-type (https://fuma.ctg.mrc.uk/celltype), summary-BayesR (https://cnsengenomics/Software/gtcb/#SummaryBayesRAlphaBet), S-PrediXcan (https://github.com/hakymb/MetaXcan), TWAS (https://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.
Data

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: GCST0027163 and GCST0027164 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:

Wellcome Trust case-control consortium: www.wtccc.org.uk

dbGaP datasets:

- phs000101.v3.p1: NH Genome-Wide Association Studies of Amyotrophic Lateral Sclerosis
- phs000126.v1.p1: CDIR: 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
- phs000335: A Genome-Wide Association Study of Lung Cancer Risk
- phs000345: Genome-wide association study for Bladder Cancer Risk
- phs000789: Collaborative Study of Genes, Nutrients and Metabolites (CGSNM)
- 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
- phs000902: 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 (IAGWAS2)
- phs000237.v1: Northwestern NiGene 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
- phs000386.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 Fracture 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 Telangiectasis Equinovarus [Cluefoot] - 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 and Endophenotypes using Twin Studies
- phs000303.v1: Genetic Epidemiology of Refractive Error in the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) Study
- phs000125.v1: CDIR: Collaborative Study on the Genetics of Alcoholism Case Control Study
- phs000103.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.v1: Genome-Wide Association Study of Amyotrophic Lateral Sclerosis
- phs000268.v1.p1: Sporadic ALS Australia Systems Genomics Consortium (SALSA-SSC)

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.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences
Life sciences study design

All studies must disclose 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 throughdbGaP to achieve a ~1:10 case:control ratio per stratum at maximum. This ratio was roughly determined based on power calculations [https://227.bwh.harvard.edu/ggc/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 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 5: 14,402 ALS patients, 32,094 controls, IlluminaOmnihEx/IIlumina2M, European ancestries
stratum 6: 3,996 ALS patients, 20,632 controls, IlluminaGSA, European ancestries
stratum 8: 1,234 ALS patients, 2850 controls, IlluminaHumanOmni24kHuangJU, 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 1).

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 Ethikkomission des Kantons St. Gallen, Switzerland.

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