Running Title: Polygenic Contributions to ALS

Machine learning suggests polygenic contribution to cognitive dysfunction in amyotrophic lateral sclerosis (ALS)

Katerina Placek[1], Michael Benatar[2], Joanne Wuu[2], Evadnie Rampersaud[3], Laura Hennessy[1], Vivianna M. Van Deerlin[4], Murray Grossman[1], David J. Irwin[1], Lauren Elman[1], Leo McCluskey[1], Colin Quinn[1], Volkan Granit[2], Jeffrey M. Statland[5], Ted. M Burns[6], John Ravits[7], Andrea Swenson[8], Jon Katz[9], Erik Pioro[10], Carlayne Jackson[11], James Caress[12], Yuen So[13], Samuel Maiser[14], David Walk[14], Edward B. Lee[4], John Q. Trojanowski[4], Philip Cook[15], James Gee[15], Rosa Rademakers[16], The CReATe Consortium[17], Wenan Chen[3], Gang Wu[3], J. Paul Taylor[3,18], & Corey T McMillan[1*]

[1] Department of Neurology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[2] Department of Neurology, University of Miami, Leonard M. Miller School of Medicine, Miami, FL
[3] Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN
[4] Department of Pathology & Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[5] Department of Neurology, University of Kansas Medical Center, Kansas City, KS
[6] Department of Neurology, University of Virginia Health System, Charlottesville, VA
[7] Department of Neurosciences, University of California San Diego, San Diego, CA
[8] Department of Neurology, University of Iowa
[9] Department of Neurology, Washington University in St. Louis, Missouri.
[10] Department of Neurology, Cleveland Clinic, Cleveland, OH
[11] Department of Neurology, University of Texas Health Science Center, San Antonio, San Antonio, TX
[12] Department of Neurology, Wake Forest University School of Medicine, Winston-Salem, NC
[13] Department of Neurology, Stanford University Medical Center, San Jose, CA
[14] Department of Neurology, University of Minnesota Medical Center, Minneapolis, MN
[15] Penn Image Computing Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[16] Department of Neuroscience, Mayo Clinic, Jacksonville, FL
[17] Rare Diseases Clinical Research Network, National Institutes of Health, Bethesda, MD
[18] The Howard Hughes Medical Institute, Chevy Chase, MS

[±] Refer to Appendix for full CReATe Consortium author list.
[*] **Correspondence:**
Penn Frontotemporal Degeneration Center, Department of Neurology
3700 Hamilton Walk, Suite 606B
Philadelphia, PA,19104
mcmillac@pennmedicine.upenn.edu; (215) 614 0987

Word/Figure Counts: Abstract=148 words (150 max); Body=3591 words (5000 max, excluding Supplementary Methods); Body References=44 (65 total); Figures/Tables=6.

Abstract

Amyotrophic lateral sclerosis (ALS) is a multi-system disorder characterized by progressive muscular weakness and, in addition, cognitive/behavioral dysfunction in nearly 50% of patients. The mechanisms underlying risk for cognitive dysfunction, however, remain elusive. Using sparse canonical correlation analysis (sCCA), an unsupervised machine-learning technique, we observed that 26 single nucleotide polymorphisms collectively associate with baseline cognitive performance in 330 ALS patients from the multicenter Clinical Research in ALS and Related Disorders for Therapeutic Development consortium. We demonstrate that a polygenic risk score derived from sCCA relates to longitudinal cognitive decline in the same cohort, and also to *in vivo* cortical thinning (N=80) and *post mortem* burden of TDP-43 pathology in the middle frontal and motor cortices (N=55) in independent validation cohorts of patients with sporadic ALS. Our findings suggest that common genetic polymorphisms contribute to the manifestation of cognitive dysfunction and disease vulnerability in a polygenic manner in ALS.

Introduction

Nearly half of patients with amyotrophic lateral sclerosis (ALS) manifest progressive decline in cognition consistent with extra-motor frontal and temporal lobe neurodegeneration, including 14% also diagnosed with frontotemporal dementia (FTD).[1,2] Comorbid cognitive dysfunction is a marker of poorer prognosis in this fatal disease and confers risk for shorter survival, more rapid functional decline, and greater caregiver burden.[3-6] While linkage analysis and genome-wide association studies (GWAS) have identified rare causal mutations[7-9] and common risk loci[10-14] linking the incidence of ALS with FTD, whether and how identified variants relate to phenotypic heterogeneity – including in cognition - has not yet been explored.

The genetic landscape of ALS is largely characterized by 'apparent sporadic' disease occurring in 90% of patients with no known family disease history or pathogenic mutation,[15] but recent evidence suggests polygenic contribution to disease incidence from common genetic variants.[16,17] Population-based studies estimate that only 5-10% of non-familial and 40-50% of familial ALS cases can be attributed to known pathogenic mutations[18] (e.g. *C9ORF72,*[7,19] *NEK1,*[20] *SOD1*[21]), but GWAS have revealed many loci of common genetic variation that confer risk for ALS and FTD. These include the largest ALS GWAS to-date which identified the novel *KIF5A* gene,[11] and genome-wide conjunction and conditional false discovery rate (FDR) analyses demonstrating genetic linkage between ALS and FTD via common single nucleotide polymorphisms (SNPs) at known and novel loci.[14] In patients with apparent sporadic ALS, single SNPs associated with risk of ALS and FTD demonstrate quantitative trait modification of cognitive decline,

in vivo cortical degeneration in the dorsolateral and ventromedial prefrontal cortices and anterior and middle temporal lobes, and overall *post mortem* pathologic burden of hyperphosphorylated transactive DNA response protein – 43 kiloDalton (TDP-43) in the middle frontal, temporal, and motor cortices.[22,23] Others have recently demonstrated shared polygenic risk between ALS and other traits (e.g. smoking, education) and diseases (e.g. schizophrenia)[16,17,24], suggesting that a single variant is unlikely to fully account for observed disease phenotype modification. However, we are unaware of any investigations evaluating polygenic contribution to cognitive dysfunction in ALS.

Here, we employ an unsupervised machine-learning approach - sparse canonical correlation analysis (sCCA) - to identify and evaluate polygenic contribution to cognitive dysfunction in ALS. Traditional approaches to polygenic scores identify variants associated with disease risk through GWAS in a univariate manner, and then compute the sum of alleles at each identified variant weighted by their effect sizes. In this study, we use data-driven sCCA to identify polygenic associations with a continuous phenotype of cognitive performance in ALS. This method is ideally suited to evaluating multivariate genotype-phenotype relationships, and employs sparsity to select maximally contributing variants and assigns corresponding weights based on model contribution with minimal *a priori* assumptions. We use sCCA to derive a polygenic risk score for cognitive dysfunction in the largest to-date longitudinal cohort of cognitively-characterized patients with ALS from the multicenter Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) consortium. We then evaluate independent neuroimaging and autopsy cohorts of patients with apparent

sporadic ALS[25] to test whether polygenic risk for cognitive dysfunction also relates to *in vivo* cortical neurodegeneration and *ex vivo* burden of underlying TDP-43 pathology. We focus our investigation on SNPs achieving genome-wide significance in the largest ALS GWAS to-date,[11] and SNPs identified as shared risk loci for both ALS and FTD.[14] We hypothesized that a sparse multivariate approach would reveal a subset of genetic loci associated with cognitive dysfunction profiles in ALS in a polygenic manner, and that follow-up analyses in independent neuroimaging and autopsy cohorts would converge to characterize quantitative traits associated with polygenic risk from identified loci.

Results

Heterogeneity of baseline cognitive and motor phenotype in ALS patients.

Smaller-scale studies have shown that ALS patients have impairments in executive, verbal fluency, and language domains, but with relative sparing of memory and visuospatial function.[4] The Edinburgh Cognitive and Behavioural ALS Screen (ECAS) was developed to measure cognitive function minimally confounded by motor disability and includes an ALS-Specific summary score that captures impairments in Language, Executive, and Verbal Fluency domains that are frequently observed in ALS patients, and an ALS-Non-specific summary score that captures less frequently observed impairments in Memory and Visuospatial function, in addition to overall performance (ECAS Total).[26] To measure the extent of heterogeneity in cognitive dysfunction, we evaluated performance of 330 patients with ALS or a related disorder (e.g., ALS-FTD, primary lateral sclerosis (PLS), progressive muscular atrophy (PMA)) participating in the

CReATe consortium's Phenotype-Genotype Biomarkers (PGB) study (*Table 1*). We used linear mixed-effects (LME) to model variability between individuals in baseline performance and rate of decline on the ECAS (Total Score, ALS-Specific Score, ALS-Non-specific Score, and Scores for each individual cognitive domain), physical disability on the ALS Functional Rating Scale – Revised (ALSFRS-R), and clinician ratings of upper motor neuron (UMN) and lower motor neuron (LMN) symptom severity; each model was adjusted for potential confounds in age, sex, education, disease duration, and mutation status (present or absent pathogenic mutation in genes causally associated with ALS or FTD). We confirm that cognitive and motor performance at baseline is heterogeneous across individuals (*Figure 1A*), and correlation analyses suggest that this is independent of disability in physical function or UMN/LMN symptom severity (all $R>0.2$; *Figure 1B*). Together, this establishes the heterogeneity of baseline and longitudinal cognitive and motor phenotypes within the PGB cohort, and validates assessment of cognitive function in a motor-disabled population.

Multivariate analyses indicate polygenic contributions to baseline cognitive performance.

To identify potential polygenic contributions to cognitive impairment in ALS we employed sCCA.[27] This unsupervised machine-learning approach enables identification of multivariate relationships between a dataset of one modality (e.g., allele dosage of SNPs) and another modality (e.g. measures of motor and cognitive function). Traditional CCA identifies a linear combination of all variables that maximize the correlation between datasets, resulting in an association of features from one dataset

(e.g., SNPs) and features from another dataset (e.g., clinical scores).[27] The "sparse" component of sCCA additionally incorporates an L1 penalty that shrinks the absolute value of the magnitude of coefficients to yield sparse models (i.e. models with fewer features) such that some coefficients are 0, and the features associated with them are effectively eliminated from the model. As a result, features that contribute little variance to the model are dropped and instead of a linear combination of all model features, we are able to identify a data-driven subset of SNPs that relate to a subset of clinical factors. Unstandardized regression coefficients resulting from sCCA serve as canonical weights indicating the direction and strength of the relationships between features.

We evaluated an allele-dosage dataset comprised of 33 SNPs identified as shared risk loci for both ALS and FTD,[14] and 13 SNPs identified as risk loci for ALS from the largest case-control GWAS to date,[11] with the latter chosen to include loci associated with ALS but not specifically with FTD. We excluded one SNP (rs12973192) since it was common to both sources, three SNPs (rs3849943, rs118082508, and rs116900480) identified as in high linkage disequilibrium (R^2>.8) with other included SNPs, and one SNP (rs10511816) due to high level of missingness across samples; this resulted in a total of 43 analyzed SNPs. We then related the allele-dosage dataset (see *Supplementary Methods*) to a dataset comprised of adjusted baseline clinical performance on functional, motor and cognitive measures extracted from LME models (see above).

After optimizing model sparsity parameters (i.e., the L1 penalty; see *Supplementary Figure 1*), sCCA of 330 patients in the PGB Cohort of CReATe revealed that a subset of

26 genetic variants related to a single clinical feature – the ALS-Specific summary score on the ECAS - achieving the maximal canonical correlation between the two datasets (R=.34, p=.0018) (*Figure 2;* see *Supplementary Table 2* for canonical weights). Among the top five most highly weighted SNPs were rs1768208 (*MOBP*), rs538622 (*ATXN3*), and rs538622 (*ERGIC1*), and rs4239633 (*UNC13A*) and rs7224296 (*NSF*). Interestingly, 4 out of the 5 top SNPs and 20 of the 26 (77%) observed genetic variants have been previously identified as shared risk loci for ALS and FTD.[14] The association of genetic variants with the ECAS ALS-Specific score suggests these variants may contribute to impairment in domains of cognition frequently observed as impaired in patients with ALS (i.e. Language, Verbal Fluency, and Executive Function) that are also the most impaired domains of cognition observed in FTD. To evaluate model robustness, we additionally repeated sCCA 10,000 times each employing a random bootstrapped subsample of 70% of participants and observed similar results (*Supplementary Figure 3*).

Polygenic score captures baseline cognition as well as longitudinal rate of cognitive decline, but not motor decline.

Next, we investigated potential polygenic contributions to rate of decline in motor and cognitive performance in the CReATe PGB cohort. Investigation of baseline performance may only capture differences in cognitive impairment at a single (somewhat arbitrary) point in time, but not differences in the trajectory of cognitive decline over time.

We first calculated a weighted polygenic score (wPGS) by computing a weighted sum of allele dosage for each individual multiplied by the canonical weights (see *Supplementary Table 2*) for each genetic variant resulting from the prior sCCA. Spearman rank-order correlation between the wPGS and adjusted baseline estimates of ECAS ALS-Specific performance resulted in correlation value similar to the canonical correlation observed from sCCA (*rs(329)*=-.34, p=2.3x10^{-10}) (*Figure 3A*), suggesting construct validity.

We then conducted Spearman's rank order correlations between the wPGS and adjusted rate of decline on each measure of cognitive and motor performance using FWE correction. To obtain adjusted rates of decline, we extracted individual slope estimates from prior LME (see above) for individuals with 2 or more observations on the ECAS, ALSFRS-R, and UMN and LMN assessments. We observed significant negative relationships between the wPGS and rate of decline on the ECAS ALS-Specific score (*rs*(277)=-.25, p=2.7x10^{-4}) (*Figure 3B*) and each other ECAS score but not on the ALSFRS-R or UMN and LMN scores (all p >.9) (*Supplementary Figure 2*). These findings further suggest polygenic contribution to rates of cognitive – but not motor – decline in patients with ALS, in addition to baseline cognitive impairment.

Polygenic score relates to cortical thinning.

Cognitive dysfunction in ALS, including performance on the ECAS, has previously been attributed to sequential disease progression rostrally and caudally from the motor cortex[28-30] and advancing disease stage.[4] To evaluate convergent anatomical evidence

for polygenic contributions to cognitive performance in patients with ALS, we related the wPGS to voxel-wise *in vivo* measures of whole-brain cortical thickness (in mm^3). Cross-sectional measurements of cortical thickness were derived from T1-weighted magnetic resonance imaging (MRI) in an independent cohort of 80 patients with sporadic forms of ALS (i.e. no known pathogenic mutation; see *Supplementary Methods*) recruited for research from the University of Pennsylvania (UPenn) (*Table 2*). Nonparametric modeling relative to 10,000 random permutations adjusted for potential confounds in age, sex, disease duration, and scanning acquisition revealed a negative relationship between the wPGS and diffuse cortical thickness; all reported clusters survived FWE-corrected threshold-free cluster-enhancement (TFCE) ($p<0.05$) (*Figure 4A*; *Supplementary Table 4*). While relatively diffusely distributed, the neuroanatomic distribution of cortical thickness associations include the anterior prefrontal cortex, inferior frontal gyrus, and orbital frontal cortex which are known regions to support the domains of cognitive dysfunction characterized by the ECAS ALS-Specific Score. Cortical thickness associated with the wPGS also encompassed motor regions including primary and supplementary motor cortex consistent with ALS disease distribution as well as more posterior anatomic regions. These findings provide a potential neuroanatomical basis for observed polygenic relationships between the wPGS and baseline cognitive performance and rate of decline, and are consistent with prior associations of cortical neurodegeneration with cognitive dysfunction in patients with ALS.[29]

Polygenic score relates to frontal burden of TDP-43 neuropathology.

To complement these *in vivo* neuroanatomical data, we also explored a potential polygenic association with the neuropathological distribution of TDP-43 at autopsy. Extracellular inclusions composed of hyperphosphorylated TDP-43 are the predominant underlying neuropathology in 98% of patients with ALS and ~50% of patients with FTD.[31,32] We assessed the presence or absence of TDP-43 pathological inclusions in *a priori* specified regions of interest in the motor cortex and middle frontal cortex in 55 patients with autopsy-confirmed ALS due to underlying TDP-43 pathology (*Table 2*), and conducted nonparametric analyses to test whether *post mortem* burden of TDP-43 pathology was related to the wPGS. Wilcoxon rank-sum tests with continuity correction revealed that wPGS was significantly higher in those with greater TDP-43 pathology for both the motor cortex (W=212, p=0.041) and middle frontal cortex (W=195, p=0.04) (*Figure 4B*). Additional analyses using ordinal measures of TDP-43 pathology (i.e., none, rare, mild, moderate, or severe neuropathologic burden) remained significant for the middle frontal cortex (p<.01) and approached significance for the motor cortex (p=.07). These findings add converging neuropathological evidence suggesting that wPGS relates to the presence of pathology in the frontal cortex.

Discussion

In this study, we evaluated polygenic contribution to cognitive dysfunction in patients with ALS by employing an unbiased machine-learning technique. We identified a polygenic relationship between cognitive dysfunction and SNPs associated with incidence of ALS and FTD, which we further investigated through quantitative-trait evaluations of two independent sporadic ALS cohorts with *in vivo* neuroimaging and

post mortem neuropathology data. Our results indicate a polygenic contribution to the presence and rate of decline of cognitive dysfunction in domains specifically impaired in ALS. Converging evidence from independent cohorts further demonstrate the generalizability of the collective polygenic contribution to biologically plausible associations including reduced *in vivo* cortical thickness and TDP-43 pathologic burden at autopsy in the middle frontal and motor cortices. To our knowledge, our study provides novel evidence for a polygenic contribution to cognitive dysfunction in ALS, quantitative anatomic characterization of identified polygenic burden in ALS spectrum disease, and further phenotypic evidence for genetic linkage between ALS and FTD.

Our findings add to an increasing body of evidence for genetic contribution to phenotypic variability in ALS, and specifically demonstrate that polygenic variation accounts for variability in cognitive dysfunction and cortical disease burden in ALS. While cognitive impairment has been more frequently linked to genetic mutations causally associated with ALS, such as *C9ORF72* repeat expansions,[33] univariate studies have implicated SNPs including those in *UNC13A* as risk factors for ALS and/or FTD.[22,24] However, mounting evidence suggests that there are polygenic, rather than single allele, modifiers of disease risk and phenotype in ALS and related neurodegenerative diseases.[16,17,24] Our observation of 26 genetic variants collectively associating with the ECAS ALS-Specific score – a combined measure of executive, language, and verbal fluency domains most commonly affected in ALS - is consistent with the idea of polygenic contribution to phenotypic variability in ALS. Notably, our observed polygenic association in the CReATe PGB Cohort appears specific to

cognitive variability in ALS: we demonstrate relative independence of cognitive performance and motor disease severity (i.e. UMN or LMN symptoms, functional performance on the ALSFRS-R) and observe no evidence for polygenic association with motor disease severity. This suggests that, in this study, polygenic risk for cognitive dysfunction does not appear to by confounded by motor disease severity.

Of the 26 genetic variants collectively associated with the ECAS ALS-Specific score via sCCA, the majority (77%) are shared risk loci for ALS and FTD.[14] The genetic variants with the strongest model contributions (i.e. with the highest canonical weights) include SNPs in or nearest to *MOBP, NSF, ATXN3, ERGIC1,* and *UNC13A*. Two loci in *MOBP* (rs9820623, rs1768208) and two loci in *UNC13A* (rs12608932, rs4239633) were amongst the most heavily weighted. Our group recently demonstrated an additive dose-effect relationship between the minor allele of rs12608932 in *UNC13A* and *in vivo* cortical thinning in the dorsal prefrontal cortex, greater burden of TDP-43 pathology in the middle frontal and motor cortices at autopsy, and worse performance on a measure of working memory associated with executive function.[22] Our group has also shown that SNPs mapped to *MOBP*, including rs1768208, relate to regional neurodegeneration in patients with sporadic forms of FTD and shorter survival in FTD patients with underlying tau or TDP-43 pathology.[32,34] rs538622 near *ERGIC1*, originally identified through conditional FDR as a shared risk locus for ALS and FTD, has also been previously demonstrated to contribute to quantitative-trait modification in ALS by relating to reduced expression of the protein *BNIP1* in ALS patient motor neurons.[14] Other top-weighted variants near *NSF* and *ATXN3* indicate potential biological plausibility.

rs10143310 is found near *ATXN3* which encodes a de-ubiquitinating enzyme, and polyglutamine expansions in *ATXN3* cause spinocerebellar axatia – type 3.[35] rs7224296 near *NSF* tags the *MAPT* H1 haplotype[36] and is associated with increased risk for FTD syndromes including progressive supranuclear palsy, corticobasal degeneration,[37] as well as Alzheimer's and Parikinson's diseases.[38]

While the mechanism of polygenic contribution to cognitive dysfunction in ALS requires further investigation, we speculate based on our findings that identified SNPs may contribute to neuroanatomic disease burden. A weighted polygenic risk score derived from the observed multivariate genotype-phenotype correlation in the CReATe PGB cohort showed robust relationships in independent validation cohorts to both *in vivo* cortical thinning and *post mortem* burden of TDP-43 pathology. Anatomically, these findings were largely consistent with prior *in vivo* structural imaging studies of neurodegeneration associated with cognitive dysfunction and with *ex vivo* investigations of TDP-43 neuropathologic burden in ALS.[28,29,39] Thus, in addition to indicating polygenic contribution to cognitive dysfunction in ALS, our findings suggest a possible mechanisms of observed findings via disease pathophysiology.

Beyond the potential biological mechanism of identifying polygenic contributions to ALS disease heterogeneity, we additionally suggest that sCCA can provide a tool for defining polygenic factors of disease risk. While sCCA has been widely applied to imaging-genetic studies,[40] we are unaware of prior applications using sCCA to define and polygenic score based on rich clinical phenotypic and biomarker data. Traditional

approaches to the generation of polygenic scores include using data from established, typically case-control, GWAS but practical considerations involve the selection of how many variants to include in a model and how to define the weights of an appropriate statistical model.[41] Critically, rather than an arbitrary selection of variants and their weights, the sparsity parameter of sCCA facilitates an unsupervised, data-driven manner to select the number of variants to include in the model and the canonical correlation provides data-driven weights to define the statistical model. The positive or negative direction of model-derived weights is potentially biologically informative, and could reflect 'risk' (i.e. positive weight) or 'protective' benefit (i.e. a negative weight). Further investigation is necessary to clarify the relationships between model-selected SNPs and model-derived canonical weight from both biological (e.g., some SNPs and/or genes may contribute more strongly to risk factors) and mathematical (e.g., weights may be constrained by minor allele frequency) perspectives. Nonetheless, sCCA may provide an optimal method for future studies of polygenic variation and direct research efforts towards model-selected variants.

Several limitations should be considered in the present study. Here, we focus our analysis on a set of SNPs selected *a priori* from previous large-scale GWAS based on genome-wide association with ALS[11] or shared risk between ALS and FTD.[14] Other genetic variants not included in the present study may also relate to cognitive dysfunction in ALS spectrum disease, and future, more densely-genotyped studies are necessary to elucidate discovery of novel genetic contributions to cognition that have not been identified through prior case-control studies. We derived a wPGS from the

sCCA to further investigate polygenic associations with longitudinal clinical and cognitive performance in the same cohort, and to investigate polygenic associations with *in vivo* and *post mortem* disease neuroanatomy in independent sporadic ALS cohorts. While we define our polygenic score from sCCA using a cross-sectional snapshot of baseline cognitive and motor performance estimates, future work using longitudinal data to define polygenic associations may further elucidate genetic risk for cognitive dysfunction in ALS. However, our finding that the wPGS also relates to longitudinal cognitive performance in the PGB cohort, and also relevant disease anatomy in independent cohorts, suggests its relevance to longitudinal cognitive phenotypes in ALS. Previous critique of polygenic scores suggest calculation based on GWAS-defined odds ratios for univariate risk loci[42] or undue influence by population variance[43] limit their use in clinical and prognostic settings. To avoid these potential confounds, our computation of a weighted polygenic risk score is based on model-selected parameters derived from an analysis including all genetic variants in an effort to account for multivariate genetic relationships. We also include the first two principal components in our model from the whole-genome analyses of the PGB CReATe cohort in an effort to account for differences in population heterogeneity. Our analyses focus on the investigation of genetic contribution to cognitive dysfunction in ALS, yet it is well established that behavioral impairment is also part of the ALS spectrum disease.[44] Further research is necessary to investigate polygenic risk for behavioral dysfunction in ALS, and whether loci included in our calculated wPGS confer risk for both cognitive and behavioral dysfunction. While this study demonstrates converging, multimodal evidence for polygenic risk in independent neuroimaging and autopsy cohorts,

replication in additional, large cohorts that allow for robust cross-validation is warranted; however alternative datasets that contain detailed genetic and cognitive phenotyping for ALS are lacking and PGB cohort represents the largest of its class. Future research investigating additional, large-scale patient cohorts with genetic and phenotypic data are necessary.

With these limitations in mind, our research demonstrates converging clinical, neuroimaging, and pathologic evidence for polygenic contribution to cognitive dysfunction and frontal cortical disease distribution in ALS. These findings should stimulate further investigation into polygenic risk for cognitive disease vulnerability in ALS, and suggest their importance in prognostic consideration and treatment trials. More broadly, this work provides insight into genetic contribution to heterogeneous phenotypes in neurodegenerative and neurological disease and supports evidence for polygenic architecture in these conditions.

Supplementary Methods

Participants: CReATe Consortium

Participants consisted of 339 individuals clinically diagnosed by a board-certified neurologist with a sporadic or familial form of amyotrophic lateral sclerosis (ALS), amyotrophic lateral sclerosis with frontotemporal dementia (ALS-FTD), progressive muscular atrophy (PMA), or primary lateral sclerosis (PLS) who were enrolled and evaluated through the CReATe Consortium's Phenotype-Genotype-Biomarker (PGB) study. All participants provided written informed consent. The PGB study is registered on clinicaltrials.gov (NCT02327845) and the University of Miami IRB (the central IRB for the CreATe Consortium) approved the study. This study entails participant blood DNA samples available for genetic screening and longitudinal evaluation at regularly-scheduled visits (ALS, ALS-FTD, and PMA: 0 (baseline), 3, 6, 12, and 18 months; PLS: 0 (baseline), 6, 12, 18, and 24 months). Participants were evaluated at each visit using alternate versions of the ECAS,[26] the ALSFRS-R,[45] with UMN and LMN burden scores calculated from a detailed elemental neuromuscular examination. Scores for UMN and LMN symptoms were calculating by summing within and across each spinal region resulting in a score ranging from 0 (none) to 10 (worst) reflective of burden of upper and lower motor neuron pathology. We excluded nine individuals with missing or incomplete longitudinal evaluation data that precluded subsequent analysis, resulting in a total of 330 participants.

Genotyping: CReATe Consortium

Peripheral blood mononuclear cell DNA was extracted using the QIAamp DNA Blood Mini Kit Qiagen #51106 and quantified using the Quant-iT dsDNA Assay Kit (Life Technologies cat#Q33130). The DNA integrity was verified by agarose gel electrophoresis (E-Gel, Life Technologies, cat#G8008-01). Unique samples were barcoded and whole genome sequencing (WGS) was performed at the HudsonAlpha Institute for Biotechnology Genomic Services Laboratory (Huntsville, Alabama) (HA) using Illumina HiSeq X10 sequencers to generate approximately 360 million paired-ends reads, each 150 base pairs (bp) in length. Peripheral DNA was extracted from participant blood samples and screened for known pathogenic mutations associated with ALS and related diseases.

Screening included sequencing for *C9ORF72* repeat expansions and WGS which is curated and validated via Sanger sequencing for pathogenic mutations associated with ALS and/or FTD in *ANG, CHCHD10, CHMP2B, FUS, GRN, hnRNPA1, hnRNPA2B1, MAPT, MATR3, OPTN, PFN1, SETX, SOD1, SPG11, SQSTM1, TARDBP, TBK1, TUBA4A, UBQLN2, VCP* (see *Table 1* for participant mutation status). The PGB study also includes patients with HSP that were excluded in the current analysis, but we additionally screened individuals for pathogenic mutations in 67 additional genes associated with hereditary spastic paraplegia (HSP) and 7 genes associated with distal hereditary motor neuropathy and all cases were negative for pathogenic mutations in these genes.

In terms of whole genome sequencing data, paired-end 150 bp reads were aligned to the GRCh38 human reference using the Burrows-Wheeler Aligner (BWA-ALN v0.7.12)[46] and processed using the GATK best-practices workflow implemented in the Genome Analysis Toolkit (GATK v3.4.0).[47] Variants for individual samples were called with HaplotypeCaller which produced intermediate gVCFs that we combined using a joint genotyping step to produce a multi-sample VCF. Variant filtration was performed using Variant Quality Score Recalibration (VQSR), which assigns a score (VQSOD) to each variant and a PASS/FAIL label and evaluated this in the context of hard filtering thresholds (Minimum Genotype Quality ≥ 20, minimum mean DP ≥ 10). Variant annotation was performed using Variant Effect Predictor (VEP)[48] and in-house pipelines including non-coding variant allele frequencies from Genome Aggregation Database (gnomAD).[49] In-house scripts were used to identify false positives resulting from paralogous mapping or/and gaps in the current human genome assembly. VCFs were further decomposed prior to analyses using the Decompose function of Vt.[50]

To control for population variance, we additionally derived the first two principal components scores for each in the CReATe PBG cohort using principal components analysis as implemented in Eigenstrat software.[51]

From the WGS data we extracted 43 hypothesized variants from WGS that previously achieved genome-wide significance for association with ALS[11] or joint association with ALS and FTD via GWAS[52] Proxy loci were genotyped (linkage disequilibrium (LD) R^2 > 0.80) when genetic data were not available for previously-published loci (see

Supplementary Table 1 for a complete list); one locus (rs10511816[52]) as excluded from analysis due to high level of missingness across samples; no LD proxy was identified. We examined linkage disequilibrium (LD) between loci on the same chromosome and pruned loci based on an R^2 value greater than or equal to 0.80; this resulted in the exclusion of rs3849943 (R^2 = 1.0 with rs3849942), rs118082508 (R^2 = 1.0 with rs117027576), and rs116900480 (R^2 = 0.929 with rs113247976). We then used Plink software[53] to recode participant genotypes according to additive genetic models (e.g. 0 = no minor allele copies, 1 = one minor allele copy, 2 = two minor allele copies), since the dominant or recessive nature of the loci included in this study remains unknown.

Linear Mixed-Effects Modeling of the ECAS and clinical measures

We conducted linear mixed-effects modeling of longitudinal performance on the ECAS, ALSFRS-R, and UMN and LMN scores using the *nlme* package in R. Each model was fit using maximum likelihood. In addition to the ECAS Total Score, we analyzed Executive Function, Language, Verbal Fluency, Memory, and Visuospatial subscores and summary scores (ALS-Specific Score, ALS-Non-Specific Score) each as dependent variables to analyze patient performance in separate cognitive domains and in clinically-grouped cognitive domains. Fixed effects included age at baseline visit (in years), lag between age of symptom onset and age at baseline visit (in years), sex, pathogenic mutation status (carrier / noncarrier), college education (yes / no) and visit time-point (in months), and we included individual by visit time-point as random effects. This allowed us to obtain adjusted estimates of baseline performance (i.e. intercept) and rate of

decline (i.e. slope) per individual, having regressed out potential confounding variables as fixed effects.

We conducted Spearman's rank-order correlations between baseline performance and rate of decline using family-wise error (FWE)-correction for multiple comparisons (see *Figure 1B*), and used the *car* package in R to conduct Levene's test for equality of variance to evaluate whether variance in baseline performance and rate of decline differed between measures.

Sparse Canonical Correlation Analysis

We next conducted sparse canonical correlation analysis (sCCA) to select a parsimonious linear combination of variables that maximize the correlation between two multivariate datasets using the *PMA* package in R.[27] One dataset comprised scaled intercepts from each clinical variable per participant (see *Mixed-Effects Modeling of Clinical Data*), and the other comprised raw genotypes and, in an effort to account for population differences in genetics, raw estimates for the first two principle components per participant.

We assumed standard (e.g. unordered) organization of each dataset, and selected regularization parameters using a grid search of 100 combinations of L1 values between 0 (most sparse) and 1 (least sparse) in increments of 0.1. The combination of L1 values yielding the highest canonical correlation of the first variate were then used in the sCCA, as similarly reported.[54] We calculated a *p* value by reporting the probability of

observing a canonical correlation greater than or equal to the model-fitted value relative to 10,000 random permutations. To evaluate model robustness, we additionally repeated sCCA 10,000 times each employing a random bootstrapped subsample of 70% of participants and observed similar results.

Polygenic Score

We calculated a weighted polygenic score (wPGS) by computing a weighted sum of allele dosage for each individual using canonical weights for each genetic variant resulting from prior sCCA. We then conducted Spearman's rank-order correlations between the wPGS and 1) adjusted estimates of baseline performance on clinical measure(s) selected from sCCA and 2) adjusted rates of decline on all clinical measures using FWE correction for multiple comparisons.

Participants: Neuroimaging cohort from UPenn

We retrospectively evaluated 80 patients with ALS recruited for research between 2006 and 2019 from the Penn Comprehensive ALS Clinic and Penn Frontotemporal Degeneration Center at UPenn (*Table 2*). Inclusion criteria consisted of negative screening for known pathogenic mutations associated with ALS and FTD, complete genotyping at analyzed SNPs, white non-Latino racial and ethnic background (population diversity is known to influence allele frequencies across individuals), disease duration from symptom onset < 2.5 standard deviations from respective group means (to avoid confounds associated with clear outliers), and T1-weighted MRI. All patients (N= 80) were diagnosed with ALS by a board-certified neurologist (L.E., L.M.,

M.G., D.I) using revised El Escorial criteria[55] and assessed for ALS frontotemporal spectrum disorder using established criteria;[56] those patients enrolled in research prior to 2017 were retrospectively evaluated through chart review. All patients participated in an informed consent procedure approved by an Institutional Review Board convened at the University of Pennsylvania.

Participants: Autopsy cohort from UPenn

We evaluated brain tissue samples from 55 ALS autopsy cases identified from the UPenn Integrated Neurodegenerative Disease Database[57] who were diagnosed by a board-certified neuropathologist (J.Q.T., E.B.L.) with ALS due to TDP-43 pathology using immunohistochemistry[31] and published criteria;[58] this cohort included 11 patients from the ALS neuroimaging cohort. Inclusion criteria consisted of negative screening for known pathogenic mutations associated with ALS or FTD, complete genotyping at analyzed SNPs, white non-Latino racial and ethnic background (population diversity is known to influence allele frequencies across individuals), disease duration from symptom onset < 2.5 standard deviations from respective group means (to avoid confounds associated with clear outliers), and sufficient brain tissue samples from the motor cortex and middle frontal cortex for pathological analysis.

Genetic Screening and SNP Genotyping: UPenn cohorts

DNA was extracted from peripheral blood or frozen brain tissue following the manufacturer's protocols (Flexigene (Qiagen) or QuickGene DNA whole blood kit (Autogen) for blood, and QIAsymphony DNA Mini Kit (Qiagen) for brain tissue). All

patients were screened for *C9ORF72* hexanucleotide repeat expansions using a modified repeat-primed polymerase-chain reaction (PCR) as previously described[59] and we excluded any patient with >=30 hexanucleotide repeats. Of the remaining individuals, we evaluated family history using a three-generation pedigree history as previously reported.[60] For cases with a family history of the same disease we sequenced 45 genes previously associated with neurodegenerative disease, including genes known to be associated with ALS (e.g. *SOD1*[21], TBK1[9]). Sequencing was performed using a custom-targeted next-generation sequencing panel (MiND-Seq)[25] and analyzed using Mutation Surveyor software (Soft Genetics, State College, PA). We excluded any individuals identified as having a pathogenic mutation.

For our cross-validation analyses of PENN MRI and autopsy samples, we genotyped peripheral or brain cerebellum DNA of each case using the Illumina Infinium Global Screening Array through the Children's Hospital of Philadelphia (CHOP) Center for Applied Genomics Core according to manufacturer's specifications. Plink was then used to remove variants with <95% call rate, Hardy-Weinberg equilibrium (HWE) p-value < 0.0001 and individuals with >5% missing genotypes. From the remaining genotypes, if the tag SNP was not available, we used the LDlink tool [61] to select a proxy SNP on the same chromosome <100kb in distance and a D'prime >0.85, with the exception of one SNP in which the maximum D'prime available was 0.57 (see *Supplementary Table 3*).

Neuroimaging Processing and Analyses

High-resolution T1-weighted MPRAGE structural scans were acquired for neuroimaging participants using a 3T Siemens Tim Trio scanner with an 8-channel head coil, with T=1620ms, T=3.09ms, flip angle=15°, 192x256 matrix, and $1mm^3$ voxels. T1-weighted MRI images were then preprocessed using Advanced Normalization Tools (ANTs) software.[62] Each individual dataset was deformed into a standard local template space in a canonical stereotactic coordinate system. ANTs provide a highly accurate registration routine using symmetric and topology-preserving diffeomorphic deformations to minimize bias toward the reference space and to capture the deformation necessary to aggregate images in a common space. Then, we used N4 bias correction to minimize heterogeneity[63] and the ANTs Atropos tool to segment images into six tissue classes (cortex, white matter, CSF, subcortical grey structures, brainstem, and cerebellum) using template-based priors and to generate probability maps of each tissue. Voxel-wise cortical thickness was measured in millimeters (mm^3) from the pial surface and then transformed into Montreal Neurological Institute (MNI) space, smoothed using a 3 sigma full-width half-maximum Gaussian kernel, and downsampled to 2mm isotropic voxels.

We used *randomise* software from FSL to perform nonparametric, permutation-based statistical analyses of cortical thickness images from our neuroimaging cohort. Permutation-based statistical testing is robust to concerns regarding multiple comparisons since, rather than a traditional assessment of two sample distributions, this method assesses a true assignment of factors (e.g. wPGS) to cortical thickness relative to many (e.g., 10,000) random assignments.[64]

We evaluated whether wPGS relates to magnitude of reduced cortical thickness, covarying for age, sex, disease duration, and scanner acquisition in an effort to control for factors associated with reduced cortical thickness but not specifically associated with polygenic risk. We constrained analysis using an explicit mask restricted to high probability cortex (>.4). We report clusters that survive $p<.05$ threshold-free cluster enhancement (TFCE)[65] corrected for family-wise error (FWE) relative to 10,000 random permutations.

Neuropathology Processing and Analyses

The presence of phosphorylated TDP-43 intraneuronal inclusions (dots, wisps, skeins) for sampled regions from the middle frontal and motor cortices were assessed on a semi-quantitative ordinal scale: 0=none/rare, 1=mild, 2=moderate, 3=severe/numerous. All neuropathological ratings were performed by an expert neuropathologist (J.Q.T., E.B.L.) blinded to patient genotype. For each region, we divided the autopsy cohort into two groups based on ordinal score of TDP-43 pathology: None/Rare=pathology score of 0, Mild to Severe=pathology score of 1 or greater. We did this to maximize statistical power, as analyses considering ordinal scales were statistically under-powered. We conducted Wilcoxon rank-sum tests using base *R* to determine whether wPGS differed between cases with absent or present TDP-43 pathology in each region.

Acknowledgements

The CReATe consortium (U54NS092091) is part of the Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), NCATS. CReATe is funded through a collaboration between NCATS, and the NINDS. Additional research support was provided by the National Institutes of Health (NS106754, AG017586, NS092091). The genomics sequencing was funded by St. Jude Children's Research Hospital American Lebanese Syrian Associated Charities (ALSAC) and the ALS Association (grants 17-LGCA-331 and 16-TACL-242).

Author Contributions

Study concept/design: K.P., M.B., C.T.M.

Acquisition, analysis, or interpretation of data: K.P., M. B., J. W., E. R., L. H., V. V. D., D. J. I., L. E., L. M., C. Q., V. G., J. S., T. B., J. R., A. S., J. K., E. P., C. J. J. C., Y. S., S. M., D. W., E. B. L., J. Q. T., P. C., J. G., R. R., G. W., J. P. T., and C. T. M.

Drafting/revising manuscript: K.P., M. B., C.T.M., M.G., E.B.L., D.W., V.G., E. R., G. W., W.C., J.S., T.B.

Competing Interests statement

The following authors declare the following competing interests:

C. M. receives financial support from Biogen and has provided consulting for Axon Advisors. M.B. reports grants from National Institutes of Health, the ALS Association, the Muscular Dystrophy Association, the Centers for Disease Control and Prevention, the Department of Defense, and Target ALS during the conduct of the study; personal

fees from Mitsubishi Tanabe Pharma, AveXis, and Genentech, outside the submitted work. In addition, M.B. has a provisional patent entitled ‘Determining Onset of Amyotrophic Lateral Sclerosis,’ and serves as a site investigator on clinical trials funded by Biogen and Orphazyme. All other authors declare no competing interests.

References

1. Montuschi, A. *et al.* Cognitive correlates in amyotrophic lateral sclerosis: a population-based study in Italy. *Journal of Neurology, Neurosurgery & Psychiatry* **86,** 168–173 (2015).
2. Beeldman, E. *et al.* The cognitive profile of ALS: a systematic review and meta-analysis update. *Journal of Neurology, Neurosurgery & Psychiatry* **87,** 611–619 (2016).
3. Elamin, M. *et al.* Cognitive changes predict functional decline in ALS: A population-based longitudinal study. *Neurology* **80,** 1590–1597 (2013).
4. Crockford, C. *et al.* ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. *Neurology* **91,** e1370–e1380 (2018).
5. Hu, W. T. *et al.* Behavior Matters—Cognitive Predictors of Survival in Amyotrophic Lateral Sclerosis. *PLoS ONE* **8,** e57584 (2013).
6. Caga, J., Hsieh, S., Lillo, P., Dudley, K. & Mioshi, E. The Impact of Cognitive and Behavioral Symptoms on ALS Patients and Their Caregivers. *Front Neurol* **10,** 942 (2019).
7. 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).
8. Van Deerlin, V. M. *et al.* TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *The Lancet Neurology* **7,** 409–416 (2008).
9. Freischmidt, A. *et al.* Haploinsufficiency of *TBK1* causes familial ALS and fronto-temporal dementia. *Nat Neurosci* **18,** 631–636 (2015).
10. 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).
11. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* **97,** 1268–1282.e6 (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. Diekstra, F. P. *et al.* C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: A genome□wide meta□analysis. *Ann Neurol.* **76,** 120–133 (2014).
14. Karch, C. M. *et al.* Selective Genetic Overlap Between Amyotrophic Lateral Sclerosis and Diseases of the Frontotemporal Dementia Spectrum. *JAMA Neurol* **75,** 860–16 (2018).
15. Turner, M. R. *et al.* Genetic screening in sporadic ALS and FTD. *Journal of Neurology, Neurosurgery & Psychiatry* **88,** 1042–1044 (2017).
16. McLaughlin, R. L. *et al.* Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun* **8,** 14774 (2017).
17. Ciga, S. B. *et al.* Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. *Ann Neurol.* **85,** 470–481 (2019).
18. Umoh, M. E. *et al.* Comparative analysis of C9orf72 and sporadic disease in an

ALS clinic population. *Neurology* **87,** 1024–1030 (2016).
19. 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).
20. Kenna, K. P. *et al.* NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet* **48,** 1037–1042 (2016).
21. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362,** 59–62 (1993).
22. Placek, K. *et al.* UNC13A polymorphism contributes to frontotemporal disease in sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging* **73,** 190–199 (2019).
23. Vass, R. *et al.* Risk genotypes at TMEM106B are associated with cognitive impairment in amyotrophic lateral sclerosis. *Acta Neuropathologica* **121,** 373–380 (2011).
24. Hagenaars, S. P. *et al.* Genetic risk for neurodegenerative disorders, and its overlap with cognitive ability and physical function. *PLoS ONE* **13,** e0198187 (2018).
25. Toledo, J. B. *et al.* A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. *Alzheimer's & Dementia* **10,** 477–484.e1 (2014).
26. Abrahams, S., Newton, J., Niven, E., Foley, J. & Bak, T. H. Screening for cognition and behaviour changes in ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **15,** 9–14 (2014).
27. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10,** 515–534 (2009).
28. Lulé, D. *et al.* Cognitive phenotypes of sequential staging in amyotrophic lateral sclerosis. *CORTEX* **101,** 163–171 (2018).
29. Agosta, F. *et al.* Structural brain correlates of cognitive and behavioral impairment in MND. *Hum Brain Mapp* **37,** 1614–1626 (2016).
30. Müller, H.-P. & Kassubek, J. MRI-Based Mapping of Cerebral Propagation in Amyotrophic Lateral Sclerosis. *Front Neurosci* **12,** 655 (2018).
31. Neumann, M. *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science* **314,** 130–133 (2006).
32. Irwin, D. J. *et al.* Myelin oligodendrocyte basic protein and prognosis in behavioral-variant frontotemporal dementia. *Neurology* **83,** 502–509 (2014).
33. Byrne, S. *et al.* Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. *The Lancet Neurology* **11,** 232–240 (2012).
34. McMillan, C. T. *et al.* Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiology of Aging* **35,** 1473–1482 (2014).
35. Burnett, B., Li, F. & Pittman, R. N. The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. *Human Molecular Genetics* **12,** 3195–3205 (2003).
36. Yokoyama, J. S. *et al.* Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia. *Acta Neuropathol* **133,** 825–837 (2017).

37. Ferrari, R. *et al.* Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *Journal of Neurology, Neurosurgery & Psychiatry* **88,** 152–164 (2017).
38. Desikan, R. S. *et al.* Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. *Mol. Psychiatry* **20,** 1588–1595 (2015).
39. Prudlo, J. *et al.* TDP-43 pathology and cognition in ALS: A prospective clinicopathologic correlation study. *Neurology* **87,** 1019–1023 (2016).
40. Parkhomenko, E., Tritchler, D. & Beyene, J. Sparse canonical correlation analysis with application to genomic data integration. *Stat Appl Genet Mol Biol* **8,** Article 1–34 (2009).
41. Sugrue, L. P. & Desikan, R. S. What Are Polygenic Scores and Why Are They Important? *JAMA* **321,** 1820–1821 (2019).
42. Wald, N. J. & Old, R. The illusion of polygenic disease risk prediction. *Genetics in Medicine 2019* **319,** 1 (2019).
43. Mostafavi, H., Harpak, A., Conley, D., Pritchard, J. K. & Przeworski, M. Variable prediction accuracy of polygenic scores within an ancestry group. *bioRxiv* **51,** 629949 (2019).
44. Lillo, P., Mioshi, E., Zoing, M. C., Kiernan, M. C. & Hodges, J. R. How common are behavioural changes in amyotrophic lateral sclerosis? *Amyotrophic Lateral Sclerosis* **12,** 45–51 (2010).
45. Cedarbaum, J. M. *et al.* The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences* **169,** 13–21 (1999).
46. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
47. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
48. Hunt, S. E. *et al.* Ensembl variation resources. *Database (Oxford)* **2018,** 1193 (2018).
49. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* **49,** 531210 (2019).
50. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinformatics* **31,** 2202–2204 (2015).
51. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38,** 904–909 (2006).
52. Karch, C. M. *et al.* Selective Genetic Overlap Between Amyotrophic Lateral Sclerosis and Diseases of the Frontotemporal Dementia Spectrum. *JAMA Neurol* (2018). doi:10.1001/jamaneurol.2018.0372
53. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81,** 559–575 (2007).
54. Xia, C. H. *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* 1–14 (2018). doi:10.1038/s41467-018-05317-y

55. Brooks, B. R., Miller, R. G., Swash, M., Munsat, T. L.World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. in **1,** 293–299 (2000).
56. Strong, M. J. *et al.* Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 1–22 (2017). doi:10.1080/21678421.2016.1267768
57. Toledo, J. B. *et al.* A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. *Alzheimers Dement* **10,** 477–84.e1 (2014).
58. Mackenzie, I. R. A. *et al.* A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathol* **122,** 111–113 (2011).
59. Suh, E. *et al.* Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. *Acta Neuropathologica* **130,** 363–372 (2015).
60. Wood, E. M. *et al.* Development and Validation of Pedigree Classification Criteria for Frontotemporal Lobar Degeneration. *JAMA Neurol* **70,** 1411–1417 (2013).
61. Machiela, M. J. & Chanock, S. J. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics* **34,** 887–889 (2018).
62. Tustison, N. J. *et al.* Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage* **99,** 166–179 (2014).
63. Tustison, N. J. *et al.* N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging* **29,** 1310–1320 (2010).
64. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92,** 381–397 (2014).
65. Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44,** 83–98 (2009).

Figure Legends.

1. **Clinical and genetic heterogeneity in ALS.** A) Differences in baseline performance and rate of decline on each measure of cognitive performance and motor symptom severity for each individual in the CReATe PGB cohort included in this analysis; the heatmap indicates each individual's standard deviation from the mean, with warmer colors denoting positive values (i.e. baseline or rate of decline that is greater than the mean) and cooler colors denoting negative standard deviation values (i.e. baseline or rate of decline that is lesser than the mean). B) Correlation between baseline performance and rate of decline for all cognitive and motor measures across the CReATe PGB cohort; the heatmap indicates Spearman correlation coefficient, with warmer colors indicative of positive correlations and cooler colors indicative of negative correlations. C) Variation in SNP genotype for each individual in the CReATe PGB cohort; the color key indicates the number of minor alleles at each SNP for each individual: Blue = 0 minor alleles, Yellow = 1 minor allele, Red = 2 minor alleles.

2. **Sparse, polygenic relationship between clinical and genetic variation in ALS.** sCCA selected the Edinburgh Cognitive and Behavioral ALS Screen (ECAS) ALS-Specific score and 26 SNPs which resulted in the maximal correlation between clinical and genetic variates in the CReATe PGB cohort. The color key denotes which variates were selected for inclusion in the model: Green = selected, Gray = not selected. The heatmap denotes the canonical weight strength associated with each relationship between each selected SNP and the ECAS ALS-Specific Score, with warmer colors indicative of positive weights and cooler colors indicative of negative weights.

3. **Polygenic risk score correlates with cognitive performance on the ECAS.** A) Weighted polygenic risk score (wPGS) correlates with A) adjusted baseline and B) rate of decline on the Edinburgh Cognitive and Behavioral ALS Screen (ECAS) ALS-Specific score in the CReATe PGB cohort.

4. **Reduced cortical thickness and greater burden of pTDP-43 pathology relates to higher polygenic risk score in independent validation cohorts.** A) Sporadic ALS (N=80) from an independent validation cohort at UPenn with higher weighted polygenic risk score (wPGS) exhibited greater reduction of cortical thickness in diffuse neocortical regions. The heatmap indicates the associated T-statistic for each voxel, with yellow representing the highest value. B) Beeswarm boxplots of wPGS in sporadic ALS at autopsy (N=55) from an independent validation cohort at UPenn show that presence of pTDP-43 pathology in the motor cortex and middle frontal cortex are associated with higher wPGS.

Supplementary Figure Legends.

1. **Gridsearch for sparse canonical correlation analysis (sCCA) L1 parameters.** Each column indicates 1 of 100 unique combinations of L1 parameters (ranging 0.1 to 1) applied to clinical and genetic datasets, and each row lists a feature entered into the sCCA. The heatmap denotes the canonical weight strength between clinical features and SNPs within each column; warmer colors indicate positive weights and cooler colors indicate negative weights.

2. **Polygenic risk score correlates with rate of decline in cognitive performance on the ECAS.** Scatterplots demonstrating relationships between weighted polygenic risk score (wPGS) and rate of decline on each Edinburgh Cognitive and Behavioral ALS Screen (ECAS) score and upper motor neuron (UMN), lower motor neuron (LMN), and ALS Functional Rating Scale – Revised (ALSFRS-R) scores in the CReATe PGB cohort.

3. **Evaluation of sparse canonical correlation analysis (sCCA) model robustness in bootstrapped subsamples.** Each column indicates 1 of 10,000 random bootstrapped subsamples of 70% of participants in the CReATe PGB cohort, and each row lists a feature entered into the sCCA. The heatmap denotes the canonical weight strength between clinical features and SNPs within each column; warmer colors indicate positive weights and cooler colors indicate negative weights.

Tables

Table 1: Demographic Characteristics of the CReATe PGB cohort.

		ALS	ALS-FTD	PLS	PMA
N		280	15	22	13
Sex	**Male (%)**	163 (58%)	12 (80%)	11 (50%)	8 (625)
Number of Visits	**Mean (SD)**	3.08 (1.35)	2.87 (1.13)	2.82 (1.26)	3.31 (1.38)
Age at Symptom Onset	**Mean (SD)**	56.34 (12.54)	62.27 (10.18)	49.68 (7.39)	48.08 (15.31)
Symptom Onset to Baseline (years)	**Mean (SD)**	3.59 (4.97)	3.53 (2.47)	8.45 (6.12)	7.77 (7.17)
Site of Symptom Onset					
Bulbar		46 (16.4)	4 (26.7)	5 (22.7)	-
Bulbar & Limb		7 (2.5)	-	3 (13.6)	-
Bulbar & Other	**N (%)**	7 (2.5)	1 (6.7)	-	-
Limb		175 (62.5)	4 (26.7)	13 (59.1)	11 (84.6)
Limb & Other		22 (7.6)	-	1 (4.5)	1 (7.7)
Other		7 (2.5)	5 (33.3)	-	1 (7.7)
College Education or greater	**N (%)**	197 (70.4)	9 (60.0)	20 (90.9)	10 (76.9)
Mutation Carrier		34 (12.1)	3 (20.0)	0 (0.0)	0 (0.0)
C9ORF72		22 (7.9)	3 (20.0)	-	-
C9ORF72 and UBQLN2		1 (0.4)	-	-	-
SOD1	**N (%)**	8 (2.9)	-	-	-
SQSTM1		1 (0.4)	-	-	-
TARDBP		1 (0.4)	-	-	-
TBK1		1 (0.4)	-	-	-
Baseline ALSFRS-R (0-48)	**Mean (SD)**	35.02 (7.09)	34.53 (6.42)	36.50 (5.95)	33.62 (7.83)
UMN Score (0-10)	**Mean (SD)**	2.71 (1.68)	2.65 (1.85)	4.4 (1.33)	0.87 (0.73)
LMN Score (0-10)	**Mean (SD)**	2.53 (1.49)	2.49 (1.74)	0.59 (0.96)	4.84 (1.93)
ECAS					
ALS Specific (0-100)		80.89 (10.85)	51.73 (13.22)	87.95 (7.47)	81.62 (11.61)
Language (0-28)		25.84 (2.66)	20.60 (4.12)	26.82 (1.97)	26.62 (1.26)
Verbal Fluency (0- 24)		16.62 (5.10)	7.29 (5.41)	19.50 (3.38)	16.77 (4.36)
Executive (0-48)	**Mean (SD)**	38.43 (5.97)	24.33 (10.59)	41.64 (3.29)	38.23 (7.50)
ALS Non-Specific (0-36)		28.01 (3.81)	17.93 (8.56)	29.73 (2.76)	27.62 (6.31)
Memory (0-24)		16.44 (3.54)	8.27 (7.00)	17.95 (2.84)	15.69 (6.20)
Visuospatial (0- 12)		11.57 (0.88)	10.36 (2.17)	11.77 (0.43)	11.92 (0.28)
ECAS Total (0-136)		108.90 (13.05)	69.67 (20.62)	117.68 (9.12)	109.23 (16.47)

PGB = Phenotype-Genotype-Biomarker; CReATe = Clinical Research in ALS and Related Disorders for Therapeutic Development ; ALS = Amyotrophic lateral sclerosis, ALS-FTD = ALS-Frontotemporal dementia; PLS = Primary lateral sclerosis, PMA = Progressive muscular atrophy; ALSFRS-R = Revised ALS Functional Rating Scale; UMN = upper motor neuron; LMN = lower motor neuron; ECAS = Edinburgh Cognitive and Behavioral ALS Screen; SD = standard deviation

Table 2. Demographics for neuroimaging (A) and autopsy (B) validation amyotrophic lateral sclerosis (ALS) cohorts from UPenn.

A. Neuroimaging Cohort

N (Male)	80 (49)
Age at MRI in Years, M (SD)	59.76 (11.45)
Education in Years, M (SD)	15.10 (3.03)
Disease Duration in Years, M (SD)	2.90 (2.33)
Site of Symptom Onset, N (%)	
Bulbar	18 (22.50)
Limb	49 (61.25)
Cognitive	8 (10.00)
Unknown	5 (6.25)
ALSFRS-R, M (SD)	34.51 (7.26)

B. Autopsy Cohort

N (Male)	55 (30)
Age at Death Years, M (SD)	65.31 (10.36)
Education in Years, M (SD)	15.54 (3.49)
Disease Duration at Death in Years, M (SD)	4.11 (2.92)
Site of Symptom Onset, N (%)	
Bulbar	8
Limb	30
Cognitive	1
Respiratory	1
Unknown	15
ALSFRS-R, M (SD)	18.09 (6.10)

Abbreviations: ALSFRS-R = ALS Functional Rating Scale – Revised, M = Mean, SD = standard deviation

Supplementary Tables

Supplementary Table 1: List of genetic variants analyzed in the PGB Study. Variants in red were excluded due to linkage disequilibrium ($R^2 > 0.80$).

Marker Name	Nearest Gene	Chr	1000 Genome GMAF	GRCh38 Position	Proxy Marker	Proxy HG19 Position
rs2068667	*NFASC*	1	0.208	chr1:204948552	rs11240317	chr1:204920322
rs11185393	*AMY1A*	1	0.368	chr1:104209379	rs67205957	chr1:104752258
rs515342	*ASB1*	2	0.214	chr2:238458655	rs508986	chr2:239337691
rs9820623	*MOBP*	3	0.406	chr3:39452367	rs6765697	chr3:39493239
rs13079368	*MOBP*	3	0.275	chr3:39471060	rs1464047	chr3:39526874
rs1768208	*MOBP*	3	0.323	chr3:39481512	rs616147	chr3:39534481
rs10463311	*TNIP1*	5	0.431	chr5:151031274	-	-
rs3828599	*GPX3*	5	0.417	chr5:151022235	rs4958872	chr5:150402334
rs538622	*ERGIC1*	5	0.32	chr5:172920676	rs2446192	chr5:172352369
rs17111695	*NAF1*	5	0.183	chr5:151052885	rs12518386	chr5:150438085
rs757651	*REEP2*	5	0.016	chr5:138455779	rs149312547	chr5:137792021
rs10488631	*TNPO3*	7	0.059	chr7:128954129	rs12539741	chr7:128596805
rs17070492	*LOC101927815*	8	0.208	chr8:2563763	-	-
rs7813314	*BC045738*	8	0.2	chr8:2558274	rs6996532	chr8:2417678
rs10869188	*C9ORF72*	9	0.49	chr9:72614090	rs7032232	chr9:75229116
rs870901	*AK097706*	9	0.133	chr9:107086201	rs60743641	chr9:109854824
rs3849942	*C9ORF72*	9	0.183	chr9:27543283	rs3849943	chr9:27543382
rs13302855	*C9ORF72*	9	0.086	chr9:27595997	rs34460171	chr9:27594491
rs732389	*AK294518*	10	0.205	chr10:78584745	rs7071538	chr10:80338173

rs7118388	*CAT*	11	0.454	chr11:34432600	rs1962369	chr11:34456941
rs12803540	*CAT*	11	0.138	chr11:34471200	rs17881488	chr11:34492443
rs117027576	*KIF5A*	12	0.00913	chr12:56922819	-	-
rs113247976	*KIF5A*	12	0.007	chr12:57581917	-	-
rs142321490	*KIF5A*	12	0.006	chr12:58282349	-	-
rs74654358	*TBK1*	12	0.012	chr12:64488187	-	-
rs1578303	*HTR2A*	13	0.204	chr13:47389011	rs144877054	chr13:47962781
rs10492593	*PCDH9*	13	0.121	chr13:66919985	rs73208976	chr13:67486924
rs17446243	*TTL/TEL*	13	0.116	chr13:40174794	rs78375967	chr13:40751567
rs10139154	*SCFD1*	14	0.428	chr14:30678292	-	-
rs10143310	*ATXN3*	14	0.339	chr14:92074037	-	-
rs12886280	*NUBPL*	14	0.412	chr14:31829453	rs35875023	chr14:32298974
rs6603044	*BTBD1*	15	0.332	chr15:83015059	rs12904695	chr15:83700365
rs9901522	*PMP22*	17	0.18	chr17:14770617	-	-
rs739439	*KIAA0524*	17	0.105	chr17:28396803	rs35714695	chr17:26719788
rs2240601	*MSI2*	17	0.192	chr17:57673751	rs16942143	chr17:55748611
rs2285642	*GGNBP2*	17	0.407	chr17:36556904	rs10707226	chr17:34916453
rs7224296	*NSF*	17	0.472	chr17:46722680	rs9912530	chr17:44836302
rs12973192	*UNC13A*	19	0.278	chr19:17642430	-	-
rs12608932	*UNC13A*	19	0.43	chr19:17641880	rs12973192	chr19:17753239
rs4239633	*UNC13A*	19	0.28	chr19:17631660	rs71162163	chr19:17744075
rs2425220	*DLGAP4*	20	0.241	chr20:36371037	rs2425223	chr20:35001520

rs75087725	*C21orf72*	21	0.003	chr21:44333234	-	-
rs3849943	*C9ORF72*	9	0.183	chr9:27543384	-	-
rs118082508	*KIF5A*	12	0.005	chr12:5692503	-	-
rs116900480	*KIF5A*	12	0.006	chr12:58262322	-	-

Abbreviations: GMAF = global minor allele frequency; Chr = chromosome; GRCh38 = Genome Reference Consortium Human Build 38; HG19 = Human Genome Project 19

Supplementary Table 2: Features and canonical weights selected from sparse canonical correlation analysis (sCCA).

Feature	Canonical Weight
ECAS ALS-Specific Score	-1
rs1768208 *(MOBP)*	-0.393
rs10143310 *(ATXN3)*	-0.338
rs538622 *(ERGIC1)*	-0.269
rs11185393 *(AMY1A)*	-0.196
rs9901522 *(PMP22)*	-0.186
rs12886280 *(NUBPL)*	-0.174
rs10488631 *(TNPO3)*	-0.11
rs739439 *(KIAA0524)*	-0.107
rs757651 *(REEP2)*	-0.047
rs2240601 *(MSI2)*	-0.046
rs113247976 *(KIF5A)*	-0.023
rs75087725 *(C21orf72)*	-0.019
rs10492593 *(PCDH9)*	0.002
rs7813314 *(BC045738)*	0.007
rs7118388 *(CAT)*	0.049
rs17111695 *(NAF1)*	0.05
rs10463311 *(TNIP1)*	0.075
rs10869188 *(C9ORF72)*	0.092
rs10139154 *(SCFD1)*	0.099
rs3828599 *(GPX3)*	0.167

rs12608932 *(UNC13A)*	0.17
rs9820623 *(MOBP)*	0.179
rs2068667 *(NFASC)*	0.185
rs6603044 *(BTBD1)*	0.202
rs4239633 *(UNC13A)*	0.21
rs7224296 *(NSF)*	0.538

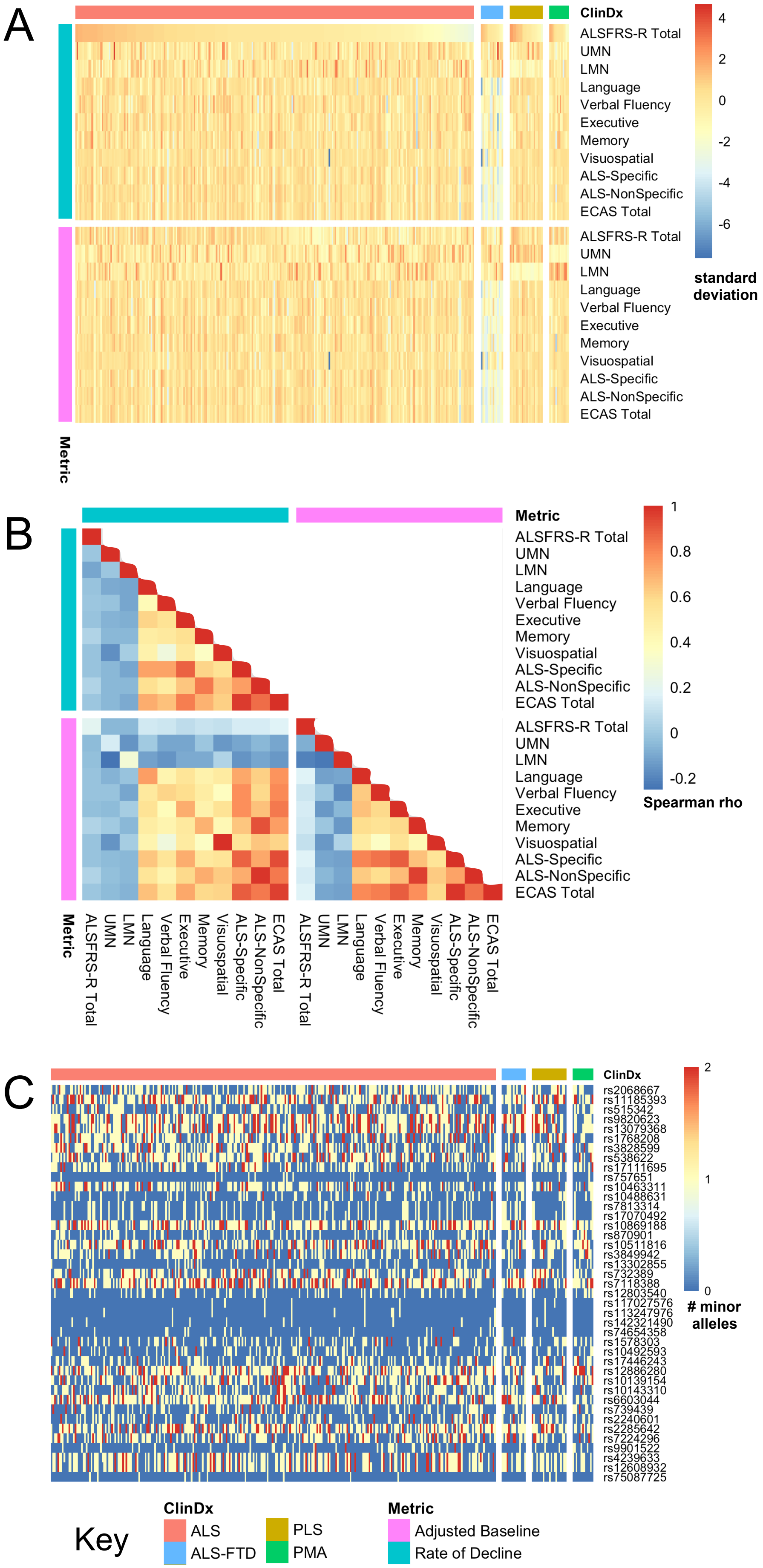
Supplementary Table 3: List of genetic variants and proxies analyzed in the Penn MRI and neuropathology samples.

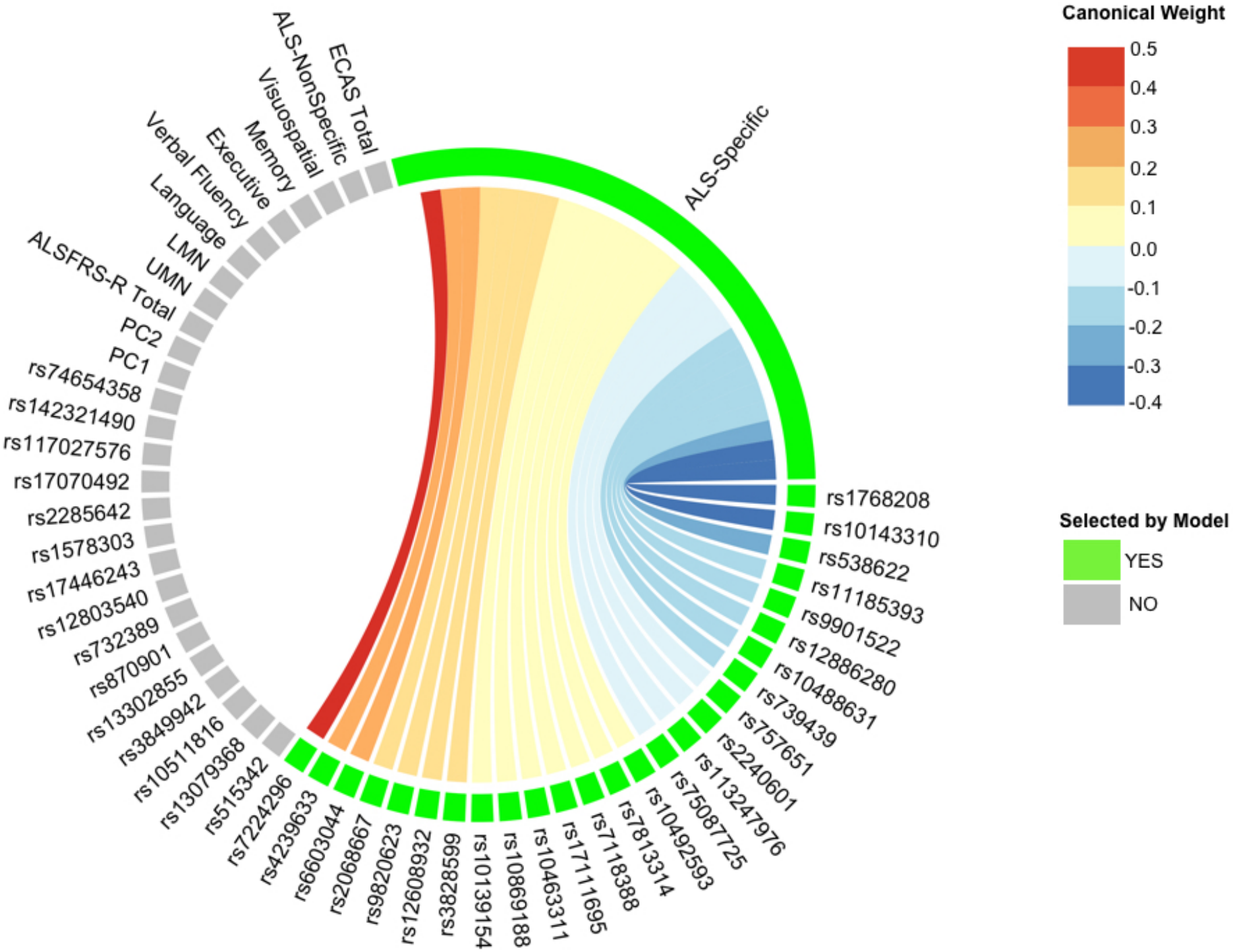
Marker Name	Nearest Gene	Chr	GRCh37 Position	Proxy Marker	Proxy GRCh37 Position	D’
rs10139154	*SCFD1*	14	31147498	rs230340	31125745	1.00
rs10143310	*ATXN3*	14	92540381	rs12588287	92536959	1.00
rs10463311	*TNIP1*	5	150410835	rs3924	150410135	1.00
rs10488631	*TNPO3*	7	128594183	rs10488631	N/A	N/A
rs10492593	*PCDH9*	13	67494117	rs1924311	67546580	1.00
rs10869188	*C9ORF72*	9	75229006	rs10217204	75231217	0.85
rs11185393	*AMY1A*	1	104752001	rs11185392	104751861	1.00
rs113247976	*KIF5A*	12	57975700	rs113247976	N/A	N/A
rs12608932	*UNC13A*	19	17752689	rs12608932	N/A	N/A
rs12886280	*NUBPL*	14	32298659	rs1953273	32368934	0.85
rs17111695	*NAF1*	5	150432446	rs60646696	150432657	1.00
rs1768208	*MOBP*	3	39523003	rs1768208	N/A	N/A
rs2068667	*NFASC*	1	204917680	rs12737855	204912843	1.00
rs2240601	*MSI2*	17	55751112	rs1005631	55733373	0.98
rs3828599	*GPX3*	5	150401796	rs2233287	150440097	0.91
rs4239633	*UNC13A*	19	17742469	rs34474857	17653298	0.57
rs538622	*ERGIC1*	5	172347679	rs514627	172355039	1.00
rs6603044	*BTBD1*	15	83683811	rs7172665	83717146	1.00
rs7118388	*CAT*	11	34454147	rs7118388	N/A	N/A
rs7224296	*NSF*	17	44800046	rs183211	44788310	0.99
rs739439	*KIAA0524*	17	26723822	rs117288663	26749941	0.85
rs75087725	*C21orf72*	21	45753117	rs8133730	45727242	1.00
rs757651	*REEP2*	5	137791468	rs255997	137889505	0.96
rs7813314	*BCO45738*	8	2415366	rs7005619	2421058	0.91
rs9820623	*MOBP*	3	39493858	N/A	N/A	N/A
rs9901522	*PMP22*	17	14673934	N/A	N/A	N/A

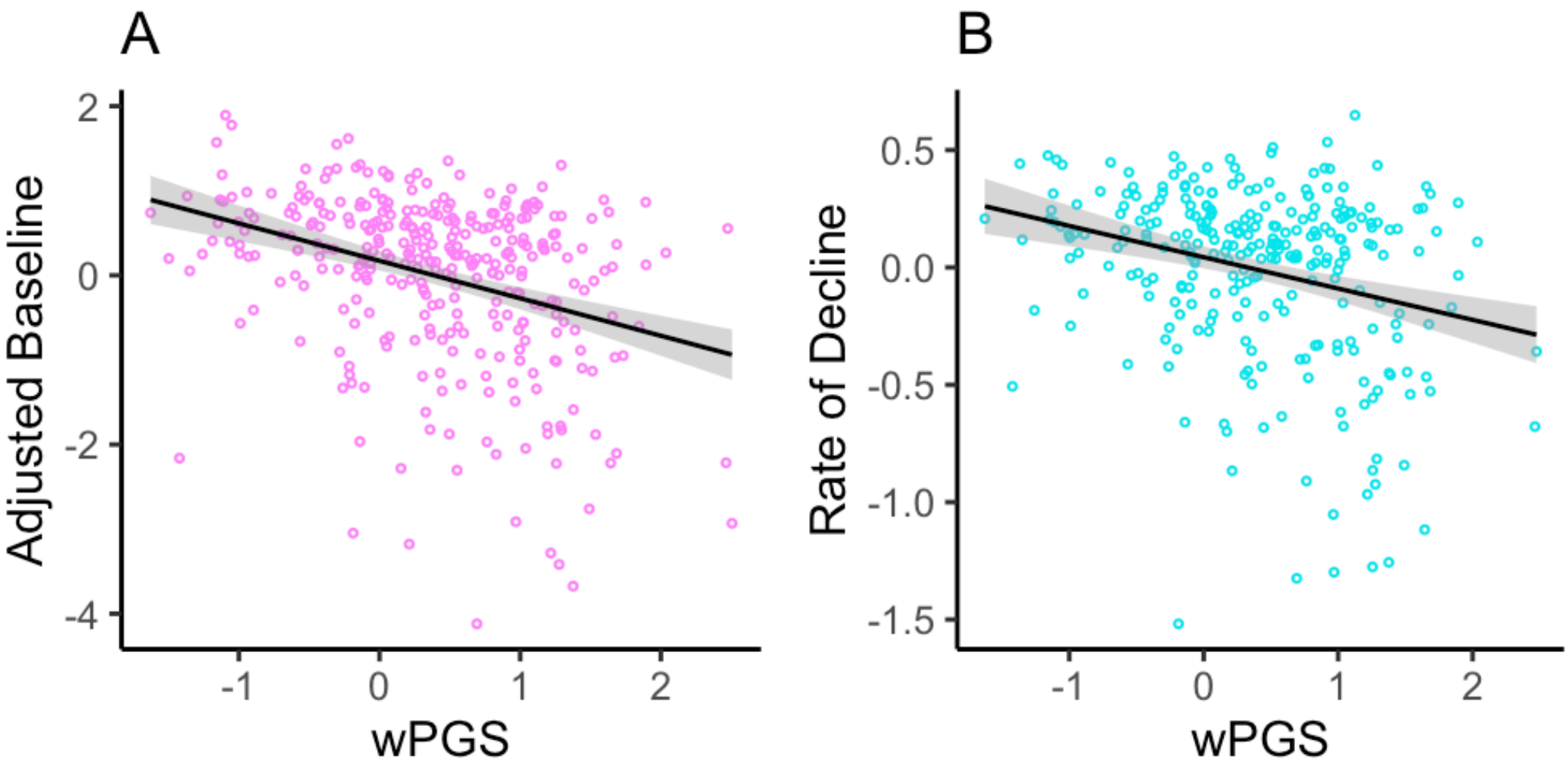
Supplementary Table 4: Peak voxel coordinates for regions of reduced cortical thickness associated with increased weighted polygenic score (wPGS) in patients with sporadic ALS from the independent validation cohort from UPenn.

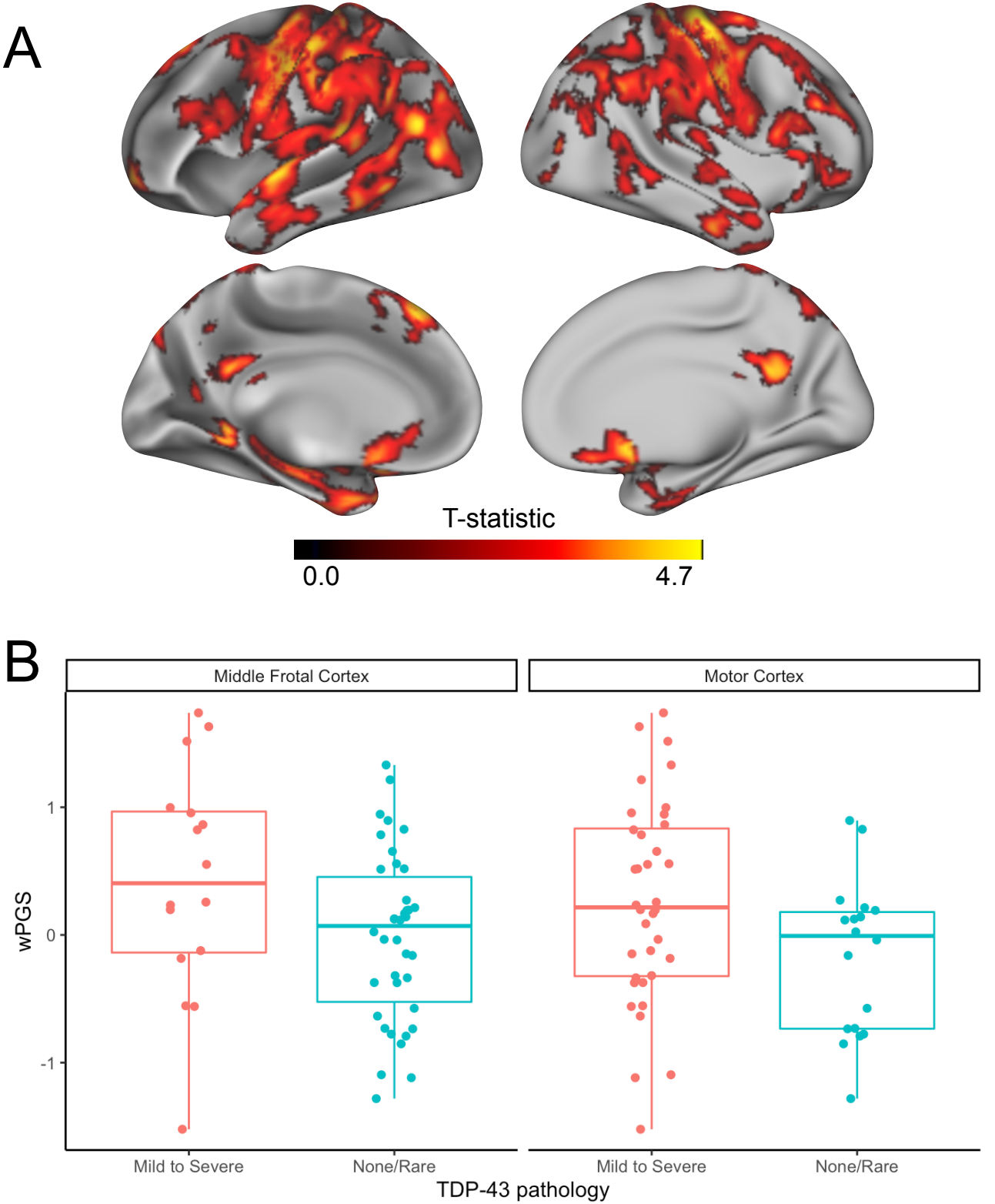
Neuroanatomic region (BA)	**L/R**	**MNI Coordinates**			***T* statistic**	***p* value (*FWE* corrected)**	**Voxels**
		x	**y**	**z**			
Primary Sensory Cortex (1)	L	-54	-16	52	4.7	0.008	11921
Premotor Cortex (6)	R	38	-16	68	4.39	0.007	9847
Subgenual Cingulate (25)	R	6	14	-8	4.21	0.018	805
Anterior Frontal Cortex (8)	L	-8	34	54	4.37	0.018	661
Posterior Cingulate Cortex (23)	R	2	-52	24	4.19	0.02	543
Anterior Prefrontal Cortex (10)	L	-36	60	-6	3.56	0.032	257
Visual Association Cortex (18)	L	-14	-46	-2	3.9	0.032	101
Inferior Frontal Gyrus (45)	R	30	24	10	2.94	0.042	57
Extrastriate Cortex (19)	R	46	-76	10	3.24	0.045	35
Orbital Frontal Cortex (47)	R	30	26	-10	2.64	0.05	26
Visual Association Cortex (18)	R	38	-84	-6	3.16	0.05	22

Abbreviations: BA = Brodmann area, L/R = Left/Right, MNI = Montreal Neurological Institute


A
ClinDx
ALSFRS-R Total
UMN
LMN
Language
Verbal Fluency
Executive
Memory
Visuospatial
ALS-Specific
ALS-NonSpecific
ECAS Total
Metric
4
2
0
-2
-4
-6
standard deviation
B
Metric
1
0.8
0.6
0.4
0.2
0
-0.2
Spearman rho
C
ClinDx
rs2068667
rs11185393
rs515342
rs9820623
rs13079368
rs1768208
rs3828599
rs538622
rs17111695
rs757651
rs10463311
rs10488631
rs7813314
rs17070492
rs10869188
rs870901
rs10511816
rs3849942
rs13302855
rs732389
rs7118388
rs12803540
rs117027576
rs113247976
rs142321490
rs74654358
rs1578303
rs10492593
rs17446243
rs12886280
rs10139154
rs10143310
rs6603044
rs739439
rs2240601
rs2285642
rs7224296
rs9901522
rs4239633
rs12608932
rs75087725
2
1
0
minor alleles
Key
ClinDx
ALS
ALS-FTD
PLS
PMA
Metric
Adjusted Baseline
Rate of Decline


ECAS Total
ALS-NonSpecific
Visuospatial
Memory
Executive
Verbal Fluency
Language
LMN
UMN
ALSFRS-R Total
PC2
PC1
rs74654358
rs142321490
rs117027576
rs17070492
rs2285642
rs1578303
rs17446243
rs12803540
rs732389
rs870901
rs13302855
rs3849942
rs10511816
rs13079368
rs515342
rs7224296
rs4239633
rs6603044
rs2068667
rs9820623
rs12608932
rs3828599
rs10139154
rs10869188
rs10463311
rs17111695
rs7118388
rs7813314
rs10492593
rs75087725
rs113247976
rs2240601
rs757651
rs739439
rs10488631
rs12886280
rs9901522
rs11185393
rs538622
rs10143310
rs1768208
ALS-Specific
Canonical Weight
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
Selected by Model
YES
NO


A
Adjusted Baseline
2
0
-2
-4
-1
0
1
2
wPGS
B
Rate of Decline
0.5
0.0
-0.5
-1.0
-1.5
-1
0
1
2
wPGS


A
T-statistic
0.0
4.7
B
Middle Frotal Cortex
Motor Cortex
wPGS
1
0
-1
Mild to Severe
None/Rare
Mild to Severe
None/Rare
TDP-43 pathology