

Cumulative Genetic Score and C9orf72 Repeat Status Independently Contribute to Amyotrophic Lateral Sclerosis Risk in 2 Case-Control Studies

John Dou, MPH, Kelly Bakulski, PhD, Kai Guo, PhD, Junguk Hur, PhD, Lili Zhao, PhD, Sara Saez-Atienzar, PhD, Ali Stark, BS, Ruth Chia, PhD, Alberto García-Redondo, PhD, Ricardo Rojas-Garcia, MD, PhD, Juan Francisco Vázquez Costa, MD, Ruben Fernandez Santiago, PhD, Sara Bandres-Ciga, PharmD, MS, PhD, Pilar Gómez-Garre, PhD, Maria Teresa Perifañán, PhD, Pablo Mir, MD, PhD, Jordi Pérez-Tur, PhD, Fernando Cardona, PhD, Manuel Menendez-Gonzalez, MD, Javier Riancho, MD, PhD, Daniel Borrego-Hernández, PhD, Lucia Galán-Dávila, MD, PhD, Jon Infante Ceberio, MD, PhD, Pau Pastor, MD, PhD, Carmen Paradas, MD, PhD, Oriol Dols-Icardo, PhD, Bryan J. Traynor, MD, PhD, Eva L. Feldman, MD, PhD, and Stephen A. Goutman, MD, MS, for the Spanish Neurological Consortium

Correspondence

Dr. Goutman
sgoutman@med.umich.edu

Neurol Genet 2023;9:e200079. doi:10.1212/NXG.000000000200079

Abstract

Background and Objectives

Most patients with amyotrophic lateral sclerosis (ALS) lack a monogenic mutation. This study evaluates ALS cumulative genetic risk in an independent Michigan and Spanish replication cohort using polygenic scores.

Methods

Participant samples from University of Michigan were genotyped and assayed for the chromosome 9 open reading frame 72 hexanucleotide expansion. Final cohort size was 219 ALS and 223 healthy controls after genotyping and participant filtering. Polygenic scores excluding the C9 region were generated using an independent ALS genome-wide association study (20,806 cases, 59,804 controls). Adjusted logistic regression and receiver operating characteristic curves evaluated the association and classification between polygenic scores and ALS status, respectively. Population attributable fractions and pathway analyses were conducted. An independent Spanish study sample (548 cases, 2,756 controls) was used for replication.

Results

Polygenic scores constructed from 275 single-nucleotide variation (SNV) had the best model fit in the Michigan cohort. An SD increase in ALS polygenic score associated with 1.28 (95% CI 1.04–1.57) times higher odds of ALS with area under the curve of 0.663 vs a model without the ALS polygenic score (p value = 1×10^{-6}). The population attributable fraction of the highest

From the Department of Epidemiology (J.D., K.B.), School of Public Health, Department of Neurology (K.G., E.L.F., S.A.G.), NeuroNetwork for Emerging Therapies (K.G., E.L.F., S.A.G.), University of Michigan, Ann Arbor; Department of Biomedical Sciences (J.H.), University of North Dakota, Grand Forks; Department of Biostatistics (L.Z.), School of Public Health, University of Michigan, Ann Arbor; Neuromuscular Diseases Research Section (S.S.-A., A.S., R.C., B.J.T.), Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD; ALS Unit (A.G.-R., D.B.-H.), Instituto de Investigación Sanitaria “i + 12” del Hospital Universitario 12 de Octubre de Madrid, SERMAS, CIBERER (A.G.-R., R.R.-G., J.F.V.C., D.B.-H.), Center for Networked Biomedical Research into Rare Diseases, Madrid; Neuromuscular Disorders Unit (R.R.-G.), Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona; Neuromuscular Unit (J.F.V.C.), Hospital Universitario y Politécnico La Fe, IIS La Fe; Department of Medicine (J.F.V.C.), Universitat de València; Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED) (R.F.S., P.G.-G., M.T.P., P.M., J.P.-T., F.C., O.D.-I.), Madrid; Lab of Parkinson’s disease and Other Neurodegenerative Movement Disorders (R.F.S.), IDIBAPS-Institut d’Investigacions Biomèdiques, Barcelona; Unitat de Parkinson i Trastorns del Moviment, Servei de Neurologia (R.F.S.), Hospital Clínic de Barcelona and Institut de Neurociències de la Universitat de Barcelona (Maria de Maetzu Center), Catalonia, Spain; Center for Alzheimer’s and Related Dementias (S.B.-C.), National Institute on Aging, Bethesda, MD; Unidad de Trastornos del Movimiento (P.G.-G., M.T.P., P.M.), Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC; Departamento de Medicina (P.M.), Universidad de Sevilla; Neurology and Molecular Genetics Mixed Investigation Unit (J.P.-T., F.C.), Instituto de Investigación Sanitaria La Fe, Molecular Genetics Unit (J.P.-T., F.C.), Institut de Biomedicina de València-CSIC; Department of Medicine (M.M.-G.), Universidad de Oviedo; Department of Neurology (M.M.-G.), Hospital Universitario Central de Asturias; Instituto de Investigación Sanitaria del Principado de Asturias (M.M.-G.), Oviedo, Spain; Service of Neurology (J.R.), Hospital Sierrallana, IDIVAL University of Cantabria, Torrelavega; Instituto de Investigación Marqués de Valdecilla (J.R., J.I.C.), Santander; Department of Neurology (L.G.-D.), ALS Unit, Hospital Clínico Universitario “San Carlos,” Madrid; Unit of Neurodegenerative Diseases (P.P.), Department of Neurology, University Hospital Germans Trias i Pujol; Neurosciences (P.P.), The Germans Trias i Pujol Research Institute (IGTP) Badalona; Department of Neurology (C.P.), Hospital Universitario Virgen del Rocío, Sevilla; and Memory Unit (O.D.-I.), Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Spain.

Go to [Neurology.org/NG](https://www.neurology.org/NG) for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the authors.

Coinvestigators are listed in the Appendix 2 at the end of the article.

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a “work of the United States Government” for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

Glossary

ALS = amyotrophic lateral sclerosis; AUC = area under the curve; C9orf72 = chromosome 9 open reading frame 72; GO = Gene Ontology; GWAS = genome-wide association study; SNV = single nucleotide variation.

20th percentile of ALS polygenic scores, relative to the lowest 80th percentile, was 4.1% of ALS cases. Genes annotated to this polygenic score enriched for important ALS pathomechanisms. Meta-analysis with the Spanish study, using a harmonized 132 single nucleotide variation polygenic score, yielded similar logistic regression findings (odds ratio: 1.13, 95% CI 1.04–1.23).

Discussion

ALS polygenic scores can account for cumulative genetic risk in populations and reflect disease-relevant pathways. If further validated, this polygenic score will inform future ALS risk models.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by rapidly progressive muscle weakness and death within 2–4 years from symptom onset^{1,2} with 50% of patients manifesting cognitive or behavioral dysfunction.^{1,2} Although ALS is traditionally divided into familial and sporadic forms, with familial ALS indicating those with an ALS family history, ALS genetic risk factors are present in both familial and sporadic patients.³ Under a monogenic model, a single risk gene is associated with a greater likelihood of developing ALS⁴ or contributes to a distinct phenotypic outcome, such as earlier age of disease onset.^{4,5} Since 1994, over 40 genes have been associated with ALS.⁶ The noncoding chromosome 9 open reading frame 72 (C9orf72) hexanucleotide expansion is the most common genetic form of ALS and is observed in 40% of familial and 10% of sporadic ALS in mixed European populations.^{7,8} Superoxide dismutase 1, TAR DNA binding protein 43, and fused in sarcoma are the next most common genes with variation frequencies of around 1% or less in sporadic cases.⁹ It is important that most patients with ALS do not carry a single causative ALS risk gene mutation. This highlights the notion of heritability, which captures the genetic and shared familial factors that contribute to disease risk.¹⁰ Heritability is as high as 38%–85% when considering twin data,¹¹ 36.9%–52.3% for parent-offspring pairs,¹⁰ 43% for all first-degree relatives,¹² and 7.2%–9.5% for common single nucleotide variation (SNV) (formerly single nucleotide polymorphism [SNP]).^{13–16} It is increasingly clear that many common SNVs may contribute a small amount of disease risk.¹⁷ Because most patients with ALS do not have a monogenic cause, it is crucial to understand the genetic contribution to ALS beyond single highly penetrant mutations to stratify population risk.

We hypothesize that polygenic scores will improve ALS risk prediction. The utility of a polygenic score for ALS, independent of C9orf72 status, has not been tested for ALS risk prediction. The goals of this study were to develop a genome-wide ALS polygenic score using an independent ALS cohort of participants not previously included in any genome-wide association study (GWAS) and test the score contribution to ALS risk models independently of C9orf72 status.

Methods

Michigan Study Participants and Sample Collection

All patients seen at the University of Michigan Pranger ALS Clinic are invited to participate, although the present case/control analysis is limited to those with ALS, thereby excluding participants with other forms of motor neuron disease. Healthy controls, without a personal or family history of a neurodegenerative disease in a first-degree or second-degree family member, are identified using a recruitment database available through the Michigan Institute for Clinical & Health Research and through population outreach via random address mailings. Participant demographics including sex (male, female), race/ethnicity (White or Caucasian, Black or African American, or Asian and Hispanic or Latino), and age (years) were obtained at the time of study enrollment. ALS diagnoses were confirmed by an ALS neurologist (S.A.G., E.L.F.) based on Gold Coast Criteria, who also recorded onset age (years), diagnosis age (years), onset segment (bulbar, cervical, lumbar, respiratory, thoracic), and presence of an ALS family history (yes or no) in the medical record. A family history of ALS in a first-degree or second-degree relative is considered positive. All participants provide venous blood, collected in an EDTA tube and frozen at –80°C for later DNA extraction.

Standard Protocol Approvals and Patient Consents

Study procedures of this Institutional Review Board (HUM28826)–approved longitudinal case/control study are published.^{18–20} All participants provided informed consent.

DNA Analysis

DNA was extracted using the QIAamp DNA Kit (Qiagen, Venlo, the Netherlands). Genome-wide genotypes at 1,748,250 positions were measured for 512 samples using the Infinium Multi-Ethnic Global-8 v1.0 array kit (Illumina, San Diego, CA) by the University of Michigan Advanced Genomics Core. All available clinical samples, including intentional duplicates ($n = 6$) and non-ALS diseased samples (6 primary lateral sclerosis, 12 other motor neuron disease), were included at this step to improve imputation

quality. DNA samples were also analyzed for the presence of the *C9orf72* repeat expansion per published protocols.⁷

PLINK (version 1.9) program was used to perform genetic microarray data quality control checks.²¹ Participants and SNVs were filtered using recommended thresholds.^{22,23} Participants were excluded for missing data at >1% of SNVs, discrepancies between genetic sex and predicted sex, and heterozygosity greater than 3 SDs from the mean. For intentional technical duplicate samples and unintentional related samples, the sample in each pair with the highest missingness was excluded. Participant inclusion based on genetic data quality control was visualized using a flow diagram, and 488 unique motor neuron disease and control participants met genetic quality filtering (eFigure 1A, links.lww.com/NXG/A610).

SNVs were excluded for missing genomic location data or missingness frequency in over 1% of samples. SNVs from autosomal chromosomes and the pseudoautosomal region of the sex chromosomes were handled separately from the non-autosomal regions of the sex chromosomes. Autosomal and pseudoautosomal region SNVs were further excluded for minor allele frequency <5% or for violating Hardy-Weinberg equilibrium (p value <10⁻⁶). SNV exclusion was described using a flow diagram, and 610,350 measured autosomal SNVs remained (eFigure 2, links.lww.com/NXG/A610).

Because population stratification by genetic ancestry can lead to confounding in genetic analyses,²⁴ principal components were computed to identify genetic ancestry groups in the cohort merged with the 1000 genomes version 5²⁵ reference panel. Individuals of all genetic ancestries were included in the main analysis, which adjusted for the first 5 multi-ancestry principal components. A sensitivity analysis limited participants to European ancestry by only including those clustered with known 1000 genomes European ancestry samples (principal component 1 < 0.02, principal component 2 < 0.08). Principal components were recomputed within the European ancestry sample for adjustment covariates.

To harmonize with the ALS GWAS,¹⁷ measured and cleaned genetic data were imputed with 1,000 genomes version 5²⁵ using the Minimac4 program.²⁶ After imputation, SNVs were filtered out if they had an imputation quality R^2 < 0.5 or a minor allele frequency <1% in the study sample and described using a flow diagram (eFigure 3, links.lww.com/NXG/A610).

Polygenic Score Development

Imputed and cleaned SNV data facilitated polygenic score creation for cohort ALS risk. ALS risk SNV weights were derived from a GWAS of 20,806 ALS cases and 59,804 controls.¹⁷ Eligible SNVs were those present in the ALS GWAS and this study's cleaned and imputed data. PRSice 2.0 generated polygenic scores,²⁷ using default pruning and clumping (250 kb window, R^2 threshold 0.1) parameters to account for linkage disequilibrium. Polygenic scores were defined as the sum of the weighted number of variant alleles per individual. SNVs were

included in the polygenic scores at a series of p value thresholds from the parent GWAS ranging from low p values (only most significant SNVs) to a 1.0 p value threshold (using all SNVs). The polygenic score with the highest R^2 in relation to ALS case-control status was selected for our primary analyses. Per Polygenic Score Reporting Standards,²⁸ for each SNV in the polygenic score, the identifier, chromosome, position, weight, and p value of association with ALS were provided (eTable 1, links.lww.com/NXG/A610). The cumulative ALS genetic risk by SNVs located beyond the *C9orf72* genomic region was determined by excluding chromosome 9 SNVs between 27,400,000 and 27,700,000 base pair positions in the primary polygenic score. A sensitivity analysis allowing SNVs in this *C9orf72* genomic region was also performed. A locus zoom plot²⁹ (Figure 1) visualized SNVs in the *C9orf72* region, and correlations of SNVs in this region with *C9orf72* expansion status were tested using the Fisher exact test.

Statistical Analyses

Statistical analyses were performed in R statistical software (version 4.1). Samples were excluded from analysis if they were duplicates or if they were from non-ALS or control participants ($n = 17$ non-ALS cases, $n = 5$ at-risk controls). Next, participants were excluded ($n = 24$) for missing data key covariates (sex, family history, age, *C9orf72* expansion status). A total of 442 participants met study inclusion criteria (eFigure 1B, links.lww.com/NXG/A610). The distributions of continuous covariates were described using mean and SD, and the distributions of categorical covariates were described using number and sample percent. Covariate distributions for included and excluded samples were provided. The Wilcoxon rank-sum test for continuous covariates and the χ^2 or Fisher exact test for categorical covariates tested for differences in the distributions of covariates between ALS and control participants.

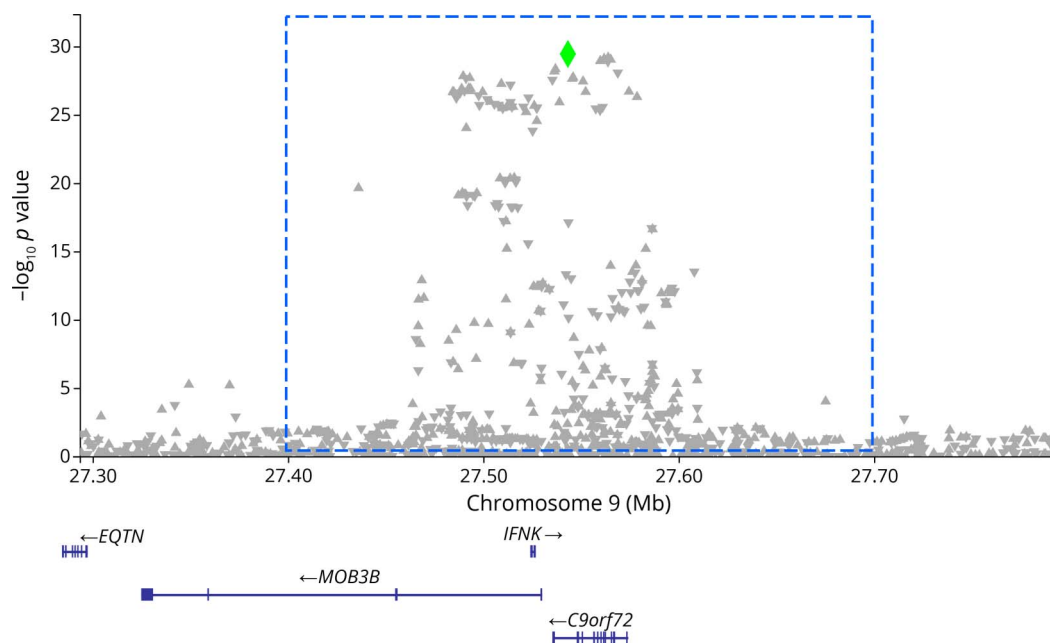
All regression models were adjusted for sex, age, family history of ALS, and 5 genetic principal components. The first analysis used multivariable logistic regression assessed the association between ALS and control status with ALS polygenic score. The second model tested for an association with *C9orf72* expansion status. The third model included both genetic components (ALS polygenic score and *C9orf72* expansion status) as predictors. Because family history and *C9orf72* expansion status had zero cell counts in controls, Firth penalized likelihood regression was used to avoid unstable effect estimates.

Additional statistical analyses, including classification testing, attributable fraction calculation, sensitivity analyses, and gene pathway analyses, are presented in eMethods (links.lww.com/NXG/A610).

Replication: Spanish Neurological Consortium

Participants were recruited across several sites in Spain as previously published³⁰ or as part of the ALS Genetic Spanish Consortium as previously published³¹ (eMethods, links.lww.com/NXG/A610). All participants provided informed consent, and the study received local ethics board approval. The

Figure 1 *C9orf72* Region of Chromosome 9 Visualized as a Locus Zoom Plot



Single nucleotide variations (SNVs) are plotted by genomic position. The y-axis corresponds to $-\log_{10}(p \text{ values})$ from the ALS genome-wide association study (Nicolas et al.¹⁷). The authors considered the *C9orf72* region to span from 27.4 Mb to 27.7 Mb on chromosome 9 as illustrated with the blue dashed box. In an independent sample, our primary polygenic score excluded the *C9orf72* region, and a sensitivity polygenic score included these SNVs. The SNV highlighted by the green diamond (rs3849943, located chr9:27543382) was associated with *C9orf72* repeat status (fisher $p \text{ value} = 0.00001$). Below the plot, positions of *C9orf72* as well as other genes in the region are shown. ALS = amyotrophic lateral sclerosis.

coordination and use of samples for this publication were approved by the Institutional Review Board of the National Institute on Aging. DNA extraction, genome-wide genotyping, *C9orf72* repeat expansion assay, and processing which followed published protocols are presented in eMethods. Statistical methods for assessing replication are also presented in eMethods.

Data Availability

Data may be shared by qualified investigators by reasonable request to the corresponding author. A data request proposal is reviewed and approved by a review panel, and a signed data sharing agreement will then be approved.

Results

Study Participants

The primary analysis included 442 participants (223 controls and 219 ALS cases) (Table 1). A family history of ALS was present in 7.8% of ALS cases and 0% of controls. The *C9orf72* repeat was present in 5.9% of ALS cases and 0% of controls. No age differences occurred between ALS and control participants, although the male participant proportion was higher in the ALS (59.0%) vs control (48%, $p \text{ value} = 0.027$) group. The 24 participants excluded for missing genetic, demographic, or ALS assessment data (eFigure 1B, links.lww.com/NXG/A610) had similar characteristics to the analysis cohort (eTable 2).

Genetic Data Characteristics and Polygenic Score Optimization

SNVs were measured at 1,748,250 positions. SNVs missing genomic location data, with missingness frequency of >1% of samples, with minor allele frequency <5%, or of Hardy-Weinberg equilibrium ($p \text{ value} < 10^{-6}$) were removed, leaving 601,350 measured autosomal SNVs (eFigure 2, links.lww.com/NXG/A610). Imputation resulted in 47,109,465 SNVs. Imputed SNVs with an imputation quality $R^2 \text{ value} < 0.5$ and SNVs with a minor allele frequency <1% were filtered. The final data set had 8,179,459 imputed SNVs (eFigure 3).

The *C9orf72* region on chromosome 9 spanned from 27.4 Mb to 27.7 Mb. After pruning, 5 SNVs were present in this region (Figure 1). Of these, 1 SNV rs3849943, located at position 27,543,382, associated with *C9orf72* expansion status (Fisher $p \text{ value} = 0.00001$). Because our goal was to estimate the cumulative genetic risk for ALS beyond the *C9orf72* expansion, out of caution, the primary polygenic score excluded this entire region. Polygenic score construction included SNVs and weights based on their association with ALS in an independent GWAS.¹⁷ Polygenic score performance was highest when constructed using a $p \text{ value threshold}$ of approximately 10^{-4} , using 275 SNVs (eFigure 4, links.lww.com/NXG/A610). At this threshold, the incremental R^2 for the polygenic score was approximately 1.2%.

For sensitivity analyses, a polygenic score using all available SNVs postpruning ($n = 254,307$ SNVs) showed an incremental R^2 of approximately 0.4%. Another sensitivity analysis included the 5

Table 1 Included Study Sample Characteristics by ALS Case and Control Status for Shared Ancestry Cohort

Characteristic	Control (N = 223) ^a	ALS (N = 219) ^a	p Value ^b
ALS polygenic score with <i>C9orf72</i> region removed	−0.08 (−0.73 to 0.64)	0.03 (−0.55 to 0.74)	0.11
ALS polygenic score with <i>C9orf72</i> region included	−0.13 (−0.74 to 0.65)	0.08 (−0.56 to 0.75)	0.080
<i>C9orf72</i> expansion status			<0.001
Negative	223 (100)	206 (94)	
Positive	0 (0)	13 (5.9)	
Family history of ALS	0 (0)	17 (7.8)	<0.001
ALS onset segment			—
Bulbar	0 (0)	59 (27)	
Cervical	0 (0)	80 (37)	
Lumbar	0 (0)	73 (33)	
Respiratory	0 (0)	1 (0.5)	
Thoracic	0 (0)	4 (1.8)	
Generalized	0 (0)	2 (0.9)	
Not applicable	209 (100)	0 (0)	
Age (y)	65 (58–71)	67 (59–73)	0.4
Sex			0.027
Female	115 (52)	90 (41)	
Male	108 (48)	129 (59)	
Self-reported race/ethnicity			0.037
Asian	2 (0.9)	1 (0.5)	
Black or African American	11 (5.0)	2 (0.9)	
Hispanic or Latino	5 (2.3)	3 (1.4)	
White or Caucasian	203 (92)	213 (97)	
Missing	2	0	
Multiancestry genetic PC1	0.0080 (0.0077–0.0082)	0.0080 (0.0078–0.0082)	0.9
Multiancestry genetic PC2	−0.020 (−0.020 to −0.020)	−0.020 (−0.020 to −0.019)	0.6
Multiancestry genetic PC3	−0.0079 (−0.0086 to −0.0071)	−0.0080 (−0.0085 to −0.0071)	0.7
Multiancestry genetic PC4	−0.0095 (−0.0101 to −0.0089)	−0.0098 (−0.0102 to −0.0090)	0.066
Multiancestry genetic PC5	−0.004 (−0.007 to 0.000)	−0.004 (−0.007 to −0.001)	0.8

Abbreviations: ALS = amyotrophic lateral sclerosis; *C9orf72* = chromosome 9 open reading frame 72; PC = principal component.

^a Median (25th percentile, 75th percentile); n (%).

^b Wilcoxon rank-sum test; Pearson χ^2 test; Fisher exact test.

SNVs in the *C9orf72* region that were previously removed, and the observed polygenic score performance was also highest using a *p* value threshold of approximately 10^{-4} (*n* = 280 SNVs) (eFigure 5, links.lww.com/NXG/A610).

Associations Between Genetic Predictors and ALS Cases Status

In unadjusted analyses, ALS cases had higher mean ALS polygenic scores (average standardized score of 0.03) vs

controls (average standardized score −0.08) (*p* value = 0.11) (eFigure 6, links.lww.com/NXG/A610). We examined the roles of genetic variables and family history in analyses adjusted for age, sex, and 5 genetic principal components. In the full study sample (*n* = 442 participants), a 1 SD increase in ALS polygenic score was associated with 1.28 times higher odds of ALS (95% CI 1.04–1.57) (Table 2), after also adjusting for *C9orf72* repeat expansion status and family history of ALS. These findings were consistent when limiting the sample to

Table 2 Regression Results in the Full Sample Used in Sensitivity Analyses (n = 223 Controls, n = 219 ALS Cases)

Variable	C9orf72 region SNVs excluded from polygenic score		C9orf72 region included	
	Polygenic score ($p_{\text{threshold}} = 0.0001$, N = 275 SNVs) N = 442 participants	Polygenic score ($p_{\text{threshold}} = 0.0001$, N = 275 SNVs) N = 416 participants (without family history and/or C9orf72 expansion)	Polygenic score ($p_{\text{threshold}} = 1.0$, N = 254,280 SNVs) N = 442 participants	Polygenic score ($p_{\text{threshold}} = 0.001$, N = 280 SNVs) N = 442 participants
Polygenic score (1 SD increase)	1.28 (1.04–1.57)	1.28 (1.04–1.57)	1.13 (0.77–1.66)	1.28 (1.05–1.58)
C9orf72 repeat (positive)	22.8 (2.8–2,954)	—	20.7 (2.6–2,674)	21.4 (2.7–2,775)
Family history of ALS (yes)	33.2 (4.3–4,268)	—	32.6 (4.3–4,184)	32.7 (4.2–4,209)
Age (10-y increase)	1.1 (0.91–1.33)	1.1 (0.91–1.33)	1.09 (0.9–1.32)	1.09 (0.9–1.32)
Sex (male)	1.52 (1.02–2.27)	1.52 (1.02–2.28)	1.47 (0.99–2.19)	1.53 (1.03–2.29)

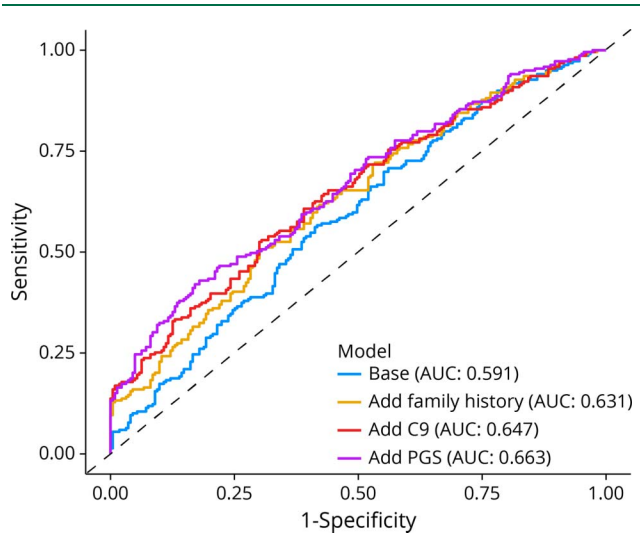
Abbreviations: ALS = amyotrophic lateral sclerosis; C9orf72 = chromosome 9 open reading frame 72; CI = confidence interval; SNV = single nucleotide variation. Regression results provided are odds ratios and 95% CIs within parentheses for association with ALS status. All Firth penalized logistic regression models were also adjusted for participant age, sex, and 5 genetic ancestry principal components. Polygenic scores for ALS are based on weights in an independent genome-wide association study.¹⁷

participants lacking a C9orf72 repeat or family history of ALS (N = 416 participants). A 1 SD increase in ALS polygenic score was again associated with 1.28 times higher odds of ALS (95% CI 1.04–1.57).

ALS Case Classification Performance

Beyond association testing, we were interested in the performance of genetic factors in ALS case classification (Figure 2).

Figure 2 ROC Curve



Base model has sex, age, and ancestry principal components (n = 442). Receiver operating characteristic curve (ROC) for classification of ALS and control participants. The base model includes sex, age, and 5 genetic principal components and has an area under the curve (AUC) of 0.591. Adding family history to the base model increases the AUC to 0.631 (likelihood ratio test p value = 0.06). Adding C9orf72 expansion in addition to family history increases the AUC to 0.647 (likelihood ratio test p value <0.001). Adding polygenic score (PGS, region around C9orf72 removed) in addition to family history and C9orf72 expansion improves the AUC to 0.663 (likelihood ratio test p value <0.001). ALS = amyotrophic lateral sclerosis.

The base classification model adjusted for sex, age, and 5 genetic principal components had an area under the curve (AUC) of 0.591. Adding family history of ALS alone to the base model increased AUC to 0.631 and improved classification over the base model (likelihood ratio test p = 0.06). Including C9orf72 repeat status as a covariate on top of the base model and family history increased the AUC to 0.647 and improved classification (likelihood ratio test p value <0.001). Adding the ALS polygenic score after family history and C9orf72 repeat status further raised AUC to 0.663 and improved classification (likelihood ratio test p value <0.001). To assess prediction accuracy, data sets were split into training and testing for 5-fold cross-validation. These AUC results were 0.539 for the base model, 0.588 adding family history, 0.603 adding C9orf72 repeat status, and finally 0.620 adding ALS polygenic score (eFigure 7, links.lww.com/NXG/A610). While the AUCs were attenuated, as a result of the sampling procedure, similar sequential prediction accuracy remained, highlighting the prediction capability.

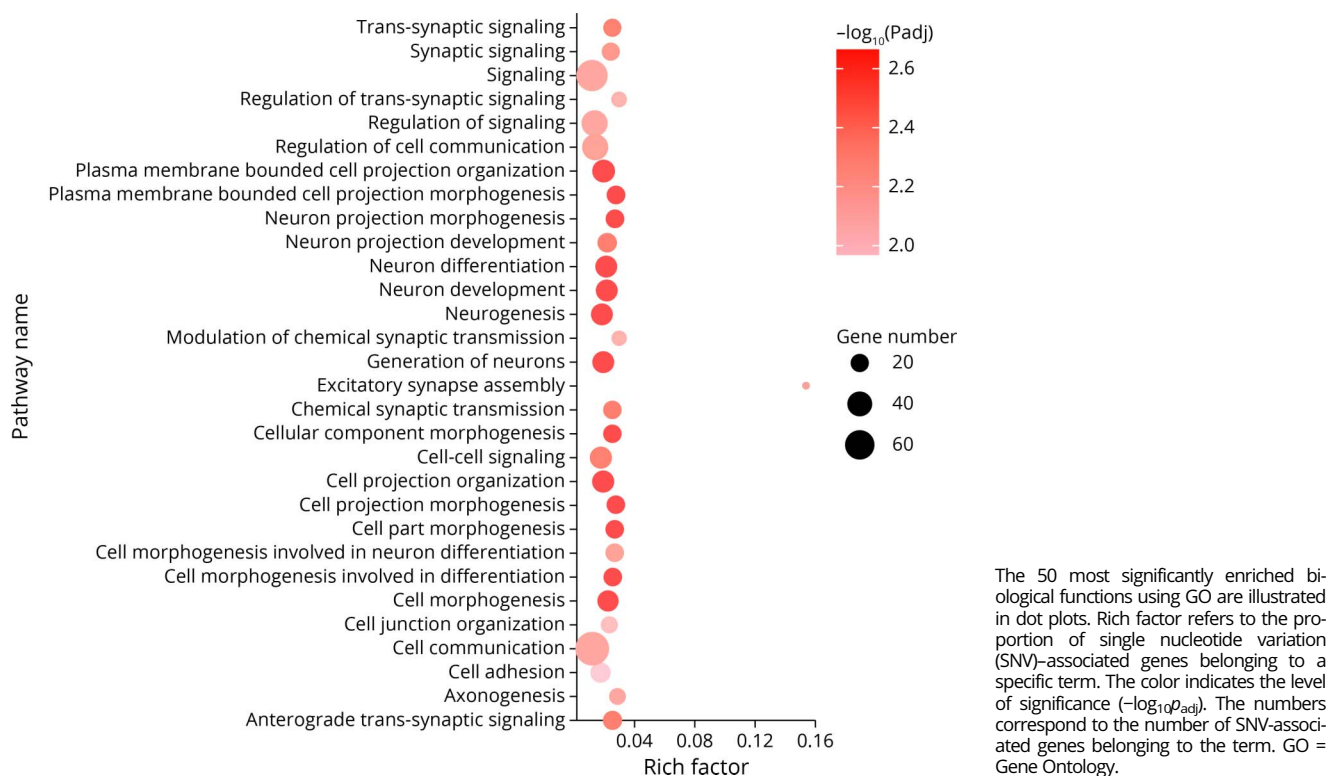
Attributable Fraction

To benchmark the fraction of ALS cases attributable to genetic factors, we compared those in the highest 20th percentile of ALS polygenic score with the rest of the sample. Here, 4.1% (95% CI –9.1 to 17.3) of ALS cases would be prevented if the highest 20th percentile of ALS polygenic score was at the level of the rest of the population. For the C9orf72 expansion, 6.3% (95% CI –2.7 to 15.3) of ALS cases would be avoided if they lacked the expansion.

Sensitivity Analyses

Sensitivity analysis (eResults, links.lww.com/NXG/A610, Table 2), including analysis around the C9orf72 region, and an analysis restricted to European ancestry participants (eTable 3, eTable 4, eFigure 8), overall showed findings consistent with the main analysis.

Figure 3 Highly Enriched GO Biological Processes



Gene Pathway Analysis

In the 275 SNV-associated genes, included the polygenic score, *richR* identified 65 highly enriched Gene Ontology (GO) biological process terms, including several related to the neuronal system, such as “neuron differentiation”, “generation of neurons”, “neuron projection morphogenesis”, “neurogenesis”, and “neuron development” (Figure 3, eTable 5, links.lww.com/NXG/A610). A total of 9 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were significantly enriched at a nominal p value <0.05 , which included “Glycosphingolipid biosynthesis-ganglio series”, “Fatty acid degradation”, and “Pancreatic secretion” (Figure 4, eTable 6).

Replication Results

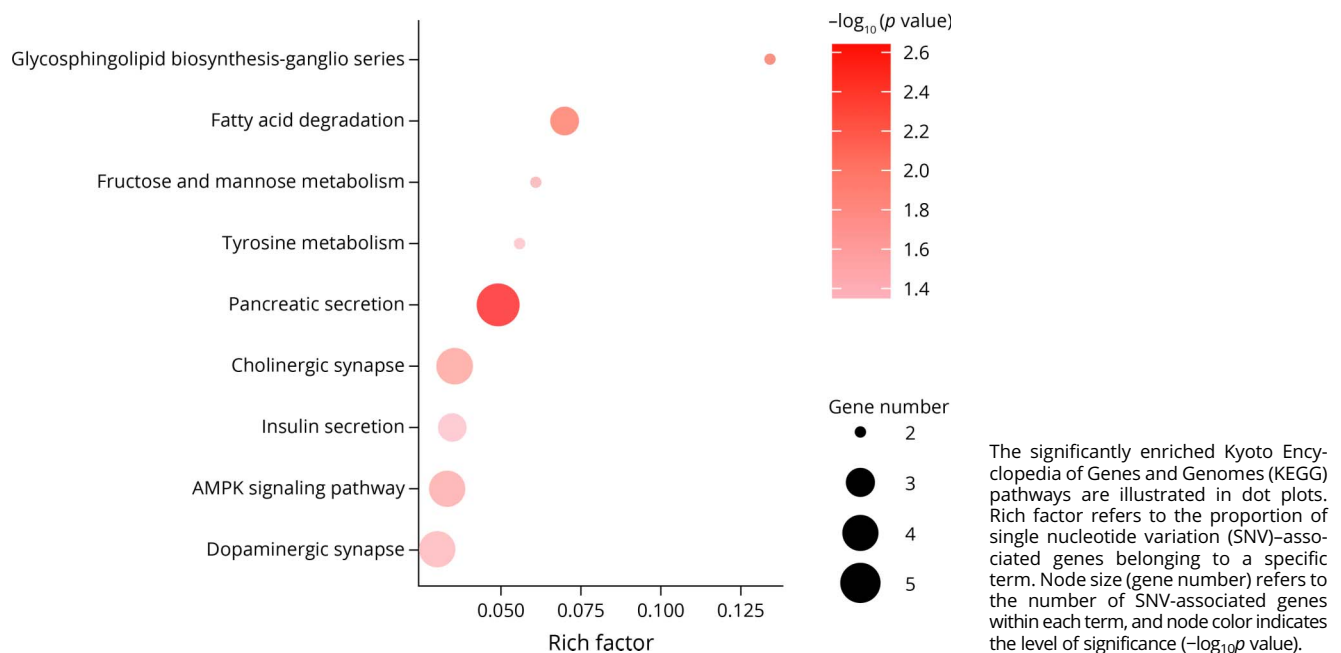
The Spanish cohort had 548 ALS cases and 2,756 controls, after removing 232 participants for missing age or *C9orf72* information. Family history, *C9orf72* expansion, and sex were associated with ALS case status (eTable 7, links.lww.com/NXG/A610). Owing to differences in genotyping arrays and allele frequencies between the Michigan and Spanish cohorts, available SNVs varied between the 2 cohorts. To harmonize analyses, SNVs were restricted to those available in both cohorts; the best performance in the Michigan cohort among overlapping SNVs resulted from a polygenic score consisting of 132 SNVs (p value threshold = 5×10^{-5}). In the Spanish cohort, a 1 SD increase in the harmonized ALS polygenic score was associated with 1.11 higher odds (95% CI 1.01–1.22) of ALS case status (p value = 0.028), adjusted for sex, age, *C9orf72* expansion, family history,

and 5 genetic principal components. In the Michigan cohort, a 1 SD increase in the harmonized ALS polygenic score was associated with 1.22 higher odds (95% CI 1.00–1.50) of ALS case status ($p = 0.04$) when including all ancestries, mirroring results above with the 275 SNV polygenic score. When limiting to European genetic ancestry in the Michigan cohort, the harmonized 132 SNV polygenic score had a stronger association, where 1 SD increase in ALS polygenic score was associated with 1.27 higher odds (95% CI 1.03–1.57) of ALS case status (p value = 0.02). Meta-analysis of the Spanish cohort and Michigan cohort (all ancestry) resulted in an estimate of 1 SD increase in ALS polygenic score being associated with 1.13 higher odds (95% CI 1.04–1.23) of ALS case status (p value = 0.004) (eFigure 9, links.lww.com/NXG/A610).

Discussion

ALS risk factors are incompletely understood. Models that predict the steps involved in developing ALS³² are necessary to generate ALS risk profiles. Representing this genetic risk³³ is critical because most individuals with ALS lack a monogenic ALS risk gene. Because genetic risk may be distributed throughout the genome, identifying polygenic risk facilitates an understanding of the multiple ALS pathologic pathways. Here we developed a weighted polygenic score using a large ALS-control GWAS.¹⁷ This score differed significantly in ALS cases vs controls from an independent Michigan cohort. Furthermore, this polygenic score

Figure 4 Highly Enriched KEGG Pathways



represents important genes and biological functions in the pathophysiology of ALS.

In this study, the ALS polygenic score with the best model fit and lowest p value was represented by 275 SNVs when excluding the region around *C9orf72* and 280 SNVs when including the region. We tested other SNV combinations as determined by default PRSice-2 p -value thresholds and a model including all SNVs. In each case, the model with fewer SNVs outperformed the larger models, suggesting that the genetic contributions to ALS are limited to a smaller subset of genes as opposed to a wide-ranging set of genes across more genomic regions. Next, we showed that an SD increase in the ALS polygenic score raised ALS odds by 1.28 times in both models without and with the *C9orf72* region. Of interest, risk increased when the *C9orf72* region was included, even after adjusting for the *C9orf72* expansion, suggesting a possible role for alleles around the *C9orf72* region on disease status, even in the absence of the repeat. Unsurprisingly, in these models, ALS risk was disproportionate for individuals with a family history or the *C9orf72* expansion. Removing individuals with an ALS family history or a *C9orf72* expansion did not change the impact of the polygenic score on ALS risk, meaning the polygenic score itself plays an essential role on the overall ALS risk profile. In addition, findings persisted when restricting to a European genetic ancestry population.

Polygenic scores summarize the combined effects that common and low-frequency alleles have on disease risk, thereby summarizing the genetic architecture of that disease.³⁴ Multiple fields use polygenic scores to explain risks such as cardiovascular

disease, cancers, neurodegenerative diseases,^{34,35} and other phenotypic traits.¹³ While polygenic scores are gaining traction for ALS,³⁶ few studies propose an ALS-specific polygenic score that can stratify populations at risk for ALS.

In contrast to our methods, an Australian group leveraged a list of 853 genetic variants with a changed amino acid sequence from a comprehensive literature search.⁴ After screening the population, 43 genetic variants from 18 genes were retained in the model, affecting 35.4% of their ALS population. However, the authors did not further develop polygenic scores.⁴ Another group³⁶ identified individuals in the Arivale Scientific Wellness cohort at elevated genetic risk for ALS using polygenic risk scores developed through literature and sought linkages to proteomics, metabolomics, and other clinical laboratory information. This group found that increased Ω -3 and decreased Ω -6 fatty acid levels and higher IL-13 levels correlated with ALS genetic risk. Based on KEGG analysis of the polygenic score developed herein, we found enrichment of the fatty acid degradation pathway,³⁷ which is consistent with ALS pathophysiology and suggests genes included in the polygenic score have biologic plausibility.

Another study used sparse canonical correlation analysis to identify a polygenic score of cognitive dysfunction in an ALS population.³⁸ Like our methods, the authors focused on SNVs achieving genome-wide significance in the Nicolas study¹⁷ and with risk loci in ALS and frontotemporal dementia. Of the 45 SNVs used in their models, 27 were associated with cognitive performance in their ALS population,

involving the genes *MOBP*, *NSF*, *ATXN3*, *ERGIC1*, and *UNC13A*. Our polygenic score also included SNVs in *MOBP*, *ATXN3*, and *UNC13A*, thereby supporting its validity. Additional uses of polygenic scores in ALS include examining polygenic traits for other diseases that overlap with ALS.^{13,39} Although this was not our approach, such studies have yielded linkages between ALS and traits of schizophrenia, cognitive performance, and educational attainment.^{13,39} Our findings are consistent with an Australian case-control study that observed a polygenic score for ALS was associated with case status.³⁹

Polygenic scores have shown utility in other neurodegenerative conditions, such as Alzheimer disease, to find those at high and low genetic risk.⁴⁰ For example, a polygenic score derived from the International Genomics of Alzheimer's Disease Project GWAS showed it could predict participants who would transition from mild cognitive impairment to late-onset Alzheimer disease.⁴¹ A similar approach using a polygenic score created from an Alzheimer's cohort GWAS data set associated with incident dementia in a large Swedish birth cohort.⁴²

Our disease classification model further supports the utility of our polygenic score. Our polygenic score improved model performance, even one that included the most prevalent ALS risk gene, the *C9orf72* expansion. In Alzheimer's, similar findings are noted, where a polygenic risk score was able to classify Alzheimer cases vs controls with an AUC of 0.83, even when excluding *APOE4* carriers.⁴³ This indicates that these genetic models are beneficial in case classification, even when considering strong genetic risks, which superimpose on polygenic risk. Another analysis similarly showed that polygenic scores in Alzheimer disease could classify patients accurately and that the prediction improved when incorporating additional variables such as sex and age.⁴⁴ In other disorders with large effect size mutations, a polygenic score has also provided additional classification information.⁴⁵

Because polygenic scores often overlap in persons with and without a disease of interest, focusing on patients with polygenic scores in distribution tails may offer better predictive power.⁴⁶ Thus, to add further perspective to this polygenic risk, we showed that 4.1% of ALS cases could be avoided for individuals with the highest 20% of polygenic score if an intervention was possible. While this population attributable risk approach considers the fraction of disease caused by exposure, this idea can also be applied to genetic data.^{47,48} For example, a study of polygenic scores in cutaneous squamous cell carcinoma showed that removing all risk alleles from a population would decrease the risk of this cancer by 62%.⁴⁹ The authors argue that identifying those at the highest genetic risk could inform programs for skin cancer screening, with the caveat that interactions of SNVs with environmental factors⁵⁰ are not included in the model. A parallel approach is also proposed for breast cancer to help identify populations that would benefit from targeted risk reduction strategies.⁵¹ A similar analysis has shown changes in the prevalence of type 2

diabetes, breast cancer, hypertension, and myocardial infarctions, if a proportion of polygenic risk is removed or enhanced in the population.⁵² Currently, there is no biomarker or tool that can definitively predict who will develop ALS later in life. Therefore, even if the polygenic score can only explain a small number of individuals at risk, it could be an important screening method for risk reduction.

Replication of these findings is important to determine the generalizability of the results. We used genotype and ALS phenotype data from an independent Spanish cohort as a replication cohort. Although the SNVs included in the polygenic score were adjusted due to the available overlap of SNVs in both data sets, there was consistency in the magnitude of the polygenic score effect, further providing support for our proposed polygenic score. Replication of polygenic scores is critical to ascertain that the methods and population background used to develop the score is generalizable.⁴⁶ Furthermore, replication cohorts can determine which risk variants are applicable across diverse populations.⁵³ Future work may incorporate a very recent updated ALS GWAS, although we selected the older GWAS here to maintain consistency with the existing literature.⁵⁴ Replicating polygenic scores in ALS remains important, although this requires large numbers of samples from participants not included in GWAS used to derive SNV weights.⁵⁵

We next queried how this set of SNVs affects disease pathobiology. Through gene enrichment and pathway analysis, we showed that this polygenic score selects multiple pathways relevant to ALS biology, including synaptic signaling, regulation of protein metabolic process, neuron projection, and axon guidance. Using KEGG pathways, we also identified important ALS biological functions, including glycosphingolipid biosynthesis and fatty acid degradation.^{20,56} A cohort of 78,500 individuals developed a polygenic score for biological pathways and cell types to determine involvement in ALS.⁵⁷ Significant pathways included those involved in neuronal development and differentiation with an emphasis on the cytoskeleton.⁵⁷ Of these pathways, the cytoskeleton pathway was significant for individuals both with and without the *C9orf72* repeat expansion, whereas the autophagosome pathway was only significant for *C9orf72* carriers. Overlapping enriched GO pathways in our polygenic score with those of that study⁵⁸ included neuron projection morphogenesis, cell morphogenesis involved in differentiation, neuron development, cellular component morphogenesis, cell development, and cell projection organization. The overall overlap shows that these 2 different methods for developing a polygenic score selects similar pathways. Other studies of gene expression in ALS have also identified dysregulated metabolic pathways and cytoskeletal pathways.⁵⁸

This study has limitations. Owing to cost and a research interest in common genetic variants, we performed genome-wide genotyping instead of whole genome sequencing. While whole genome sequencing would allow us to better account

for genetic background, the method we used is validated across many studies. In addition, the study population size is small compared with the number of individuals affected by ALS. However, the sample size here was limited to participants not included in prior GWASs and is thus a strength. This is important because developing polygenic scores from participants who are already in the reference GWAS may lead to biased results. In addition, because we did use a lower-cost genotyping strategy imputed to maximize overlap with the ALS GWAS used for weights, these methods could be beneficial for population screening where the cost of whole genome sequencing is not economically feasible. In addition, this study mainly consisted of participants with a European genetic ancestry. To support the generalizability of these findings, improving enrollment of and study of genotypes from participants with diverse backgrounds is required.

In conclusion, we find that a polygenic score for ALS can account for cumulative genetic risk in the population and reflect cellular processes that are relevant to ALS. If further validated, this polygenic score can be a valuable tool for ALS risk models and the design of ALS prevention studies.

Acknowledgment

The authors thank Crystal Pacut, Caroline Piecuch, and Stacey Sakowski, PhD, from the University of Michigan (Ann Arbor, MI) for study support, Masha Savelieff, PhD, and Emily Koubek, PhD, from the University of Michigan (Ann Arbor, MI) for editorial assistance, and the following members of the Spanish Neurological Consortium for study assistance: Mónica Povedano Panadés, Antonio Guerrero-Sola, Tania García-Sobrino, Marilina Puente Hernández, María Jesús SobridoGómez, Ivonne Jericó-Pascual, López, Adriano Jimenez-Escrig, Mario Ezquerro, and Ana Gorostidi Pagola. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH.

Study Funding

National ALS Registry/CDC/ATSDR (1R01TS000289); National ALS Registry/CDC/ATSDR CDCP-DHHS-US (CDC/ATSDR 200-2013-56856); NIEHS K23ES027221; NIEHS R01ES030049; NINDS R01NS127188, ALS Association (20-IIA-532), the Dr. Randall W. Whitcomb Fund for ALS Genetics, the Peter R. Clark Fund for ALS Research, the Scott L. Pranger ALS Clinic Fund, and the NeuroNetwork for Emerging Therapies at the University of Michigan. This work was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (Z01-AG000949-02). Project “ALS Genetic study in Madrid Autonomous Community” funded by “ESTRATEGIAS FRENTE A ENFERMEDADES NEURODEGENERATIVAS” from Spanish Ministry of Health.

Disclosure

J.F. Vázquez-Costa receives payment for lectures and presentations from Biogen. P. Mir receives payments for honoraria or lectures from Abbvie, Abbott, and Zambon. L. Galán-Dávila

receives consulting fees, payment, or honoraria from Akcea, Alnylan, Genzyme, Sobi, Pfizer, and equipment donation from Pfizer. J.I. Ceberio receives payment for lectures and presentations from Abbvie, Bial, and Zambon. B.J. Traynor holds a patent for “Diagnostic and therapeutic implications for the C9orf72 repeat expansion” and has collaborative research agreements with Ionis Pharmaceuticals, Roche, and Optimeos. E.L. Feldman receives consulting fees from Novartis and is an inventor on a patent held by University of Michigan titled, “Methods for Treating Amyotrophic Lateral Sclerosis.” S.A. Goutman is an inventor on a patent held by University of Michigan titled, “Methods for Treating Amyotrophic Lateral Sclerosis.” The other authors declare no competing interests. Go to Neurology.org/NG for full disclosures.

Publication History

Previously published at medRxiv (MEDRXIV/2022/281377) on October 31, 2022. Received by *Neurology: Genetics* March 1, 2023. Accepted in final form April 6, 2023. Submitted and externally peer reviewed. The handling editor was Associate Editor Raymond P. Roos, MD, FAAN.

Appendix 1 Authors

Name	Location	Contribution
John Dou, MPH	Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Kelly Bakulski, PhD	Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Kai Guo, PhD	Department of Neurology; NeuroNetwork for Emerging Therapies, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Jungkuk Hur, PhD	Department of Biomedical Sciences, University of North Dakota, Grand Forks, ND	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Lili Zhao, PhD	Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Sara Saez-Atienzar, PhD	Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Ali Stark, BS	Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data

Appendix 1 (continued)

Name	Location	Contribution
Ruth Chia, PhD	Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Alberto García-Redondo, PhD	ALS Unit, Instituto de Investigación Sanitaria “i + 12” del Hospital Universitario 12 de Octubre de Madrid, SERMAS; CIBERER, Center for Networked Biomedical Research into Rare Diseases, Madrid, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Ricardo Rojas-García, MD, PhD	CIBERER, Center for Networked Biomedical Research into Rare Diseases, Madrid; Neuromuscular Disorders Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Spain	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Juan Francisco Vázquez-Costa, MD	CIBERER, Center for Networked Biomedical Research into Rare Diseases, Madrid; Neuromuscular Unit, Hospital Universitario y Politécnico la Fe, IIS La Fe; Department of Medicine, Universitat de València, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Ruben Fernandez Santiago, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Lab of Parkinson's disease and Other Neurodegenerative Movement Disorders, IDIBAPS-Institut d'Investigacions Biomèdiques, Barcelona; Unitat de Parkinson i Trastorns del Moviment. Servicio de Neurología, Hospital Clínic de Barcelona and Institut de Neurociències de la Universitat de Barcelona (Maria de Maetzu Center), Catalonia, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Sara Bandres-Ciga, PharmD, MS, PhD	Center for Alzheimer's and Related Dementias, National Institute on Aging, Bethesda, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Pilar Gómez-Garre, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/ Universidad de Sevilla, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data

Appendix 1 (continued)

Name	Location	Contribution
Maria Teresa Perrián, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Spain	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Pablo Mir, MD, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC; Departamento de Medicina, Universidad de Sevilla, Spain	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Jordi Pérez-Tur, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Neurology and Molecular Genetics Mixed Investigation Unit. Instituto de Investigación Sanitaria La Fe, Valencia; Molecular Genetics Unit. Institut de Biomedicina de València-CSIC, Valencia, Spain	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Fernando Cardona, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Neurology and Molecular Genetics Mixed Investigation Unit, Instituto de Investigación Sanitaria La Fe; Molecular Genetics Unit. Institut de Biomedicina de València-CSIC, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Manuel Menéndez-González, MD	Department of Medicine, Universidad de Oviedo; Department of Neurology, Hospital Universitario Central de Asturias; Instituto de Investigación Sanitaria del Principado de Asturias, Oviedo, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Javier Riancho, MD, PhD	Service of Neurology, Hospital Sierrallana. IDIVAL University of Cantabria, Torrelavega; Instituto de Investigación Marqués de Valdecilla, Santander, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data

Continued

Appendix 1 (continued)

Name	Location	Contribution
Daniel Borrego-Hernández, PhD	ALS Unit, Instituto de Investigación Sanitaria “i + 12” del Hospital Universitario 12 de Octubre de Madrid, SERMAS, Madrid, Spain; CIBERER, Center for Networked Biomedical Research into Rare Diseases, Madrid, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Lucía Galán-Dávila, MD, PhD	Department of Neurology, ALS Unit, Hospital Clínico Universitario “San Carlos,” Madrid, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Jon Infante Ceberio, MD, PhD	Instituto de Investigación Marqués de Valdecilla, Santander, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Pau Pastor, MD, PhD	Unit of Neurodegenerative Diseases, Department of Neurology, University Hospital Germans Trias i Pujol; Neurosciences, The Germans Trias i Pujol Research Institute (IGTP) Badalona, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Carmen Paradas, MD, PhD	Department of Neurology, Hospital Universitario Virgen del Rocío, Sevilla, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Oriol Dols-Icardo, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid; Memory Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Bryan J. Traynor, MD, PhD	Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Eva L. Feldman, MD, PhD	Department of Neurology; NeuroNetwork for Emerging Therapies, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Stephen A. Goutman, MD, MS	Department of Neurology; NeuroNetwork for Emerging Therapies, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data

Appendix 2 Coinvestigators

Name	Location	Role	Contribution
Jesús Esteban-Pérez, MD	Instituto de Investigación Sanitaria “i + 12” del Hospital Universitario 12 de Octubre de Madrid, SERMAS, Spain	Site Investigator	Collection and coordination of samples
Pilar Cordero-Vázquez, MScN	Instituto de Investigación Sanitaria “i + 12” del Hospital Universitario 12 de Octubre de Madrid, SERMAS, Spain	Site Investigator	Collection and coordination of samples
Sevilla Teresa, MD	Hospital Universitario y Politécnico la Fe, IIS La Fe, Valencia, Spain	Site Investigator	Collection and coordination of samples
Adolfo López de Munain, MD, PhD	Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain	Site Investigator	Collection and coordination of samples
Julio Pardo-Fernández, MD, PhD	Hospital Clínico, Santiago de Compostela, Spain	Site Investigator	Collection and coordination of samples
Ivonne Jericó-Pascual, MD	Complejo Hospitalario de Navarra, Pamplona, Spain	Site Investigator	Collection and coordination of samples
Ellen Gelpi Mantius, MD	Neurological Tissue Bank of the Biobank-Hospital Clinic-IDIBAPS, Barcelona, Spain	Site Investigator	Collection and coordination of samples
Janet Hoenicka, PhD	Institut de Recerca Sant Joan de Déu, Barcelona, Spain	Site Investigator	Collection and coordination of samples
Victoria Alvarez Martinez, PhD	Hospital Universitario Central de Asturias, Spain	Site Investigator	Collection and coordination of samples
Francisco Javier Rodríguez de Rivera Garrido, MD	Hospital Universitario La Paz, Madrid, Spain	Site Investigator	Collection and coordination of samples
Katrin Beyer, PhD	University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain	Site Investigator	Collection and coordination of samples
Jordi Clarimón Echevarria, PhD	Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Spain	Site Investigator	Collection and coordination of samples

References

- Goutman SA, Hardiman O, Al-Chalabi A, et al. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol.* 2022; 21(5):465-479. doi:10.1016/S1474-4422(21)00414-2
- Goutman SA, Hardiman O, Al-Chalabi A, et al. Recent advances in the diagnosis and prognosis of amyotrophic lateral sclerosis. *Lancet Neurol.* 2022;21(5):480-493. doi: 10.1016/S1474-4422(21)00465-8
- Feldman EL, Goutman SA, Petri S, et al. Amyotrophic lateral sclerosis. *Lancet.* 2022; 400(10360):1363-1380. doi:10.1016/S0140-6736(22)01272-7

4. McCann EP, Henden L, Fifita JA, et al. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. *J Med Genet.* 2020;58(2):87-95. doi: 10.1136/jmedgenet-2020-106866
5. Sheppard SR, Parker MD, Cooper-Knock J, et al. Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry.* 2021;92(5):510-518. doi:10.1136/jnnp-2020-325014
6. Gregory JM, Fagegaltier D, Phatnani H, Harms MB. Genetics of amyotrophic lateral sclerosis. *Curr Genet Med Rep.* 2020;8(4):121-131. doi:10.1101/cshperspect.a024125
7. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011;72(2):257-268. doi:10.1016/j.neuron.2011.09.010
8. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011;72(2):245-256. doi:10.1016/j.neuron.2011.09.011
9. Mezzini R, Flynn LL, Pitout IL, Fletcher S, Wilton SD, Akkari PA. ALS genetics, mechanisms, and therapeutics: where are we now? *Front Neurosci.* 2019;13:1310. doi: 10.3389/fnins.2019.01310
10. Ryan M, Heverin M, McLaughlin RL, Hardiman O. Lifetime risk and heritability of amyotrophic lateral sclerosis. *JAMA Neurol.* 2019;76(11):1367. doi:10.1001/jamaneurol.2019.2044
11. McLaughlin R, Vajda A, Hardiman O. Heritability of amyotrophic lateral sclerosis: insights from disparate numbers. *JAMA Neurol.* 2015;72(8):857. doi:10.1001/jamaneurol.2014.4049
12. Trabjerg BB, Garton FC, van Rheenen W, et al. ALS in Danish Registries: heritability and links to psychiatric and cardiovascular disorders. *Neurol Genet.* 2020;6(2):e398. doi:10.1212/NXG.0000000000000398
13. Bandres-Ciga S, Noyce AJ, Hemani G, et al. Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. *Ann Neurol.* 2019;85(4):470-481. doi:10.1002/ana.25431
14. Nakamura R, Misawa K, Tohno G, et al. A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. *Commun Biol.* 2020;3(1):526.
15. Li C, Ou R, Wei Q, Shang H. Shared genetic links between amyotrophic lateral sclerosis and obesity-related traits: a genome-wide association study. *Neurobiol Aging.* 2021;102:211.e1-211.e9. doi:10.1016/j.neurobiolaging.2021.01.023
16. van Rheenen W, Shatunov A, Dekker AM, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet.* 2016;48(9):1043-1048. doi: 10.1038/ng.3622
17. Nicolas A, Kenna KP, Renton AE, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron.* 2018;97(6):1268-1283.e6. doi:10.1016/j.neuron.2018.02.027
18. Su FC, Goutman SA, Chernyak S, et al. Association of environmental toxins with amyotrophic lateral sclerosis. *JAMA Neurol.* 2016;73(7):803-811. doi:10.1001/jamaneurol.2016.0594
19. Goutman SA, Boss J, Patterson A, Mukherjee B, Batterman S, Feldman EL. High plasma concentrations of organic pollutants negatively impact survival in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry.* 2019;90(8):907-912. doi:10.1136/jnnp-2018-319785
20. Goutman SA, Boss J, Guo K, et al. Untargeted metabolomics yields insight into ALS disease mechanisms. *J Neurol Neurosurg Psychiatry.* 2020;91(12):1329-1338. doi: 10.1136/jnnp-2020-323611
21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):7. doi:10.1186/s13742-015-0047-8
22. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res.* 2018;27(2):e1608. doi: 10.1002/mpr.1608
23. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet.* 2011;Chapter 1:Unit1 19. doi: 10.1002/0471142905.hg0119s68
24. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population stratification in genetic association studies. *Curr Protoc Hum Genet.* 2017;95:1.22.21-1.22.23. doi:10.1002/cphg.48
25. The 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.
26. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics.* 2015;31(5):782-784. doi:10.1093/bioinformatics/btu704
27. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience.* 2019;8(7):giz082. doi:10.1093/gigascience/giz082
28. Wand H, Lambert SA, Tamburro C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature.* 2021;591(7849):211-219. doi:10.1038/s41586-021-03243-6
29. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336-2337. doi:10.1093/bioinformatics/btq419
30. Bandres-Ciga S, Ahmed S, Sabir MS, et al. The genetic architecture of Parkinson disease in Spain: characterizing population-specific risk, differential haplotype structures, and providing etiologic insight. *Mov Disord.* 2019;34(12):1851-1863. doi:10.1002/mds.27864
31. García-Redondo A, Dols-Icardo O, Rojas-García R, et al. Analysis of the C9orf72 gene in patients with amyotrophic lateral sclerosis in Spain and different populations worldwide. *Hum Mutat.* 2013;34(1):79-82. doi:10.1002/humu.22211
32. Al-Chalabi A, Calvo A, Chio A, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol.* 2014;13(11):1108-1113. doi:10.1016/S1474-4422(14)70219-4
33. Yanes T, McInerney-Leo AM, Law MH, Cummings S. The emerging field of polygenic risk scores and perspective for use in clinical care. *Hum Mol Genet.* 2020;29(R2):R165-R176. doi:10.1093/hmg/ddaa136
34. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19(9):581-590. doi:10.1038/s41576-018-0018-x
35. Bakulski KM, Vadari HS, Faul JD, et al. Cumulative genetic risk and APOE ε4 are independently associated with dementia status in a multiethnic, population-based cohort. *Neurol Genet.* 2021;7(2):e576. doi:10.1212/NXG.0000000000000576
36. Wainberg M, Magis AT, Earls JC, et al. Multiomic blood correlates of genetic risk identify presymptomatic disease alterations. *Proc Natl Acad Sci U S A.* 2020;117(35):21813-21820. doi:10.1073/pnas.2001429117
37. Shi N, Kawano Y, Tateishi T, et al. Increased IL-13-producing T cells in ALS: positive correlations with disease severity and progression rate. *J Neuroimmunol.* 2007;182(1-2):232-235. doi:10.1016/j.jneuroim.2006.10.001
38. Placek K, Benatar M, Wu J, et al. Machine learning suggests polygenic risk for cognitive dysfunction in amyotrophic lateral sclerosis. *EMBO Mol Med.* 2021;13(1):e12595. doi:10.15252/emmm.202012595
39. Restuadi R, Garton FC, Benyamin B, et al. Polygenic risk score analysis for amyotrophic lateral sclerosis leveraging cognitive performance, educational attainment and schizophrenia. *Eur J Hum Genet.* 2021;30(5):532-539. doi:10.1038/s41431-021-00885-y
40. Leonenko G, Baker E, Stevenson-Hoare J, et al. Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat Commun.* 2021;12(1):4506.
41. Chaudhury S, Brookes KJ, Patel T, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. *Transl Psychiatry.* 2019;9(1):154.
42. Najar J, van der Lee SJ, Joas E, et al. Polygenic risk scores for Alzheimer's disease are related to dementia risk in APOE ε4 negatives. *Alzheimers Dement (Amst).* 2021;13(1):e12142. doi:10.1002/dad2.12142
43. Escott-Price V, Myers A, Huentelman M, Shoaib M, Hardy J. Polygenic risk score analysis of Alzheimer's disease in cases without APOE4 or APOE2 alleles. *J Prev Alzheimers Dis.* 2019;6(1):16-19. doi:10.14283/jpad.2018.46
44. Escott-Price V, Shoaib M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging.* 2017;49:214.e7-214.e11. doi:10.1016/j.neurobiolaging.2016.07.018
45. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun.* 2020;11:3635.
46. Baker E, Escott-Price V. Polygenic risk scores in Alzheimer's disease: current Applications and future directions. *Front Digital Health.* 2020;2:14. doi:10.3389/fdgh.2020.00014
47. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet.* 2014;15(11):765-776.
48. Arnold N, Koenig W. Polygenic risk score: clinically useful tool for prediction of cardiovascular disease and benefit from lipid-lowering therapy? *Cardiovasc Drugs Ther.* 2021;35(3):627-635. doi:10.1007/s10557-020-07105-7
49. Sordillo JE, Kraft P, Wu AC, Asgari MM. Quantifying the polygenic contribution to cutaneous squamous cell carcinoma risk. *J Invest Dermatol.* 2018;138(7):1507-1510. doi:10.1016/j.jid.2018.01.031
50. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat.* 2012;132(2):365-377. doi:10.1007/s10549-011-1818-2
51. Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and non-modifiable risk factors among white women in the United States. *JAMA Oncol.* 2016;2(10):1295-1302. doi:10.1001/jamaoncol.2016.1025
52. Lello L, Raben TG, Hsu SDH. Sibling validation of polygenic risk scores and complex trait prediction. *Scientific Rep.* 2020;10(1):13190.
53. Bogumil D, Conti DV, Sheng X, et al. Replication and genetic risk score analysis for pancreatic cancer in a diverse multiethnic population. *Cancer Epidemiol Biomarkers Prev.* 2020;29(12):2686-2692. doi:10.1158/1055-9965.EPI-20-0963
54. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet.* 2021;53(12):1636-1648. doi:10.1038/s41588-021-00973-1
55. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet.* 2013;14(7):507-515. doi:10.1038/nrg3457
56. Goutman SA, Guo K, Savelieff MG, et al. Metabolomics identifies shared lipid pathways in independent amyotrophic lateral sclerosis cohorts. *Brain.* 2022;145(12):4425-4439. doi:10.1093/brain/awac025
57. Saez-Atienzar S, Bandres-Ciga S, Langston RG, et al. Genetic analysis of amyotrophic lateral sclerosis identifies contributing pathways and cell types. *Sci Adv.* 2021;7(3):eabd9036. doi:10.1126/sciadv.abd9036
58. Maniatis S, Aijo T, Vickovic S, et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science.* 2019;364(6435):89-93. doi:10.1126/science.aav9776