Received: 12 December 2019 | Revised: 5 May 2020 | Accepted: 18 May 2020

DOI: 10.1002/humu.24055

DATA ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Revisiting the complex architecture of ALS in Turkey: Expanding genotypes, shared phenotypes, molecular networks, and a public variant database

Ceren Tunca[1,2] | Tuncay Şeker[3] | Fulya Akçimen[2] | Cemre Coşkun[2] | Elif Bayraktar[1] | Robin Palvadeau[1] | Seyit Zor[3] | Cemile Koçoğlu[2] | Ece Kartal[2] | Nesli Ece Şen[2] | Hamid Hamzeiy[2] | Aslıhan Özoğuz Erimiş[2] | Utku Norman[4] | Oğuzhan Karakahya[4] | Gülden Olgun[4] | Tahsin Akgün[5] | Hacer Durmuş[6] | Erdi Şahin[6] | Arman Çakar[6] | Esra Başar Gürsoy[7] | Gülsen Babacan Yıldız[7] | Barış İşak[8] | Kayıhan Uluç[8] | Haşmet Hanağası[6] | Başar Bilgiç[6] | Nilda Turgut[9] | Fikret Aysal[10] | Mustafa Ertaş[6] | Cavit Boz[11] | Dilcan Kotan[12] | Halil İdrisoğlu[6] | Aysun Soysal[13] | Nurten Uzun Adatepe[14] | Mehmet Ali Akalın[14] | Filiz Koç[15] | Ersin Tan[16] | Piraye Oflazer[6] | Feza Deymeer[6] | Öznur Taştan[17] | A. Ercüment Çiçek[4,18] | Erşen Kavak[3] | Yeşim Parman[6] | A. Nazlı Başak[1,2]

[1]Suna and İnan Kıraç Foundation, Neurodegeneration Research Laboratory (NDAL), Research Center for Translational Medicine (KUTTAM), Koç University School of Medicine, Istanbul, Turkey

[2]Suna and İnan Kıraç Foundation, Neurodegeneration Research Laboratory (NDAL), Department of Molecular Biology and Genetics, Boğaziçi University, Istanbul, Turkey

[3]Genomize Inc., Boğaziçi University Technology Development Region, Istanbul, Turkey

[4]Department of Computer Engineering, Bilkent University, Ankara, Turkey

[5]Department of Anesthesiology and Reanimation, American Hospital, Istanbul, Turkey

[6]Department of Neurology, Istanbul Medical School, Istanbul University, Istanbul, Turkey

[7]Department of Neurology, Faculty of Medicine, Bezmialem Vakıf University, Istanbul, Turkey

[8]Department of Neurology, Marmara University School of Medicine, Istanbul, Turkey

[9]Department of Neurology, Namık Kemal University School of Medicine, Tekirdağ, Turkey

[10]Department of Neurology, Medipol University School of Medicine, Istanbul, Turkey

[11]Department of Neurology, Karadeniz Technical University School of Medicine, Trabzon, Turkey

[12]Department of Neurology, Faculty of Medicine, Sakarya University, Sakarya, Turkey

[13]Department of Neurology, Bakırköy Research and Training Hospital for Neurologic and Psychiatric Diseases, Istanbul, Turkey

[14]Department of Neurology, Cerrahpaşa Medical School, Istanbul University-Cerrahpaşa, Istanbul, Turkey

[15]Department of Neurology, Çukurova University Medical School, Adana, Turkey

[16]Department of Neurology, Hacettepe University Medical School, Ankara, Turkey

[17]Department of Computer Science and Engineering, Sabancı University, Istanbul, Turkey

[18]Department of Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania

Tuncay Şeker, Fulya Akçimen, A. Ercüment Çiçek, Erşen Kavak, and Yeşim Parman contributed equally to this work.

Correspondence
A. Nazlı Başak, Suna and İnan Kıraç Foundation, Neurodegeneration Research Laboratory (NDAL), KUTTAM, Koç University School of Medicine, 34450 Istanbul, Turkey.
Email: nbasak@ku.edu.tr

Present address
Fulya Akçimen, Department of Human Genetics, McGill University, Montréal QC, Canada

Cemre Coşkun, Faculty of Biology, Ludwig-Maximilians-University of Munich, Munich, Germany

Cemile Koçoğlu, Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium

Ece Kartal, Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

Nesli Ece Şen, Experimental Neurology Department, University Hospital Frankfurt, Frankfurt am Main, Germany

Hamid Hamzeiy, Computational Systems Biochemistry, Max-Planck Institute of Biochemistry, Martinsried, Germany

Piraye Oflazer, Department of Neurology, Koç University School of Medicine, Istanbul, Turkey

Feza Deymeer, Department of Neurology, Memorial Şişli Hospital, Istanbul, Turkey.

Funding information
TÜBITAK, Grant/Award Number: 109S075; Bogaziçi University Research Funds, Grant/Award Number: 15B01P1; Suna and İnan Kıraç Foundation, Grant/Award Number: 2005–2020

Abstract

The last decade has proven that amyotrophic lateral sclerosis (ALS) is clinically and genetically heterogeneous, and that the genetic component in sporadic cases might be stronger than expected. This study investigates 1,200 patients to revisit ALS in the ethnically heterogeneous yet inbred Turkish population. Familial ALS (fALS) accounts for 20% of our cases. The rates of consanguinity are 30% in fALS and 23% in sporadic ALS (sALS). Major ALS genes explained the disease cause in only 35% of fALS, as compared with ~70% in Europe and North America. Whole exome sequencing resulted in a discovery rate of 42% (53/127). Whole genome analyses in 623 sALS cases and 142 population controls, sequenced within Project MinE, revealed well-established fALS gene variants, solidifying the concept of incomplete penetrance in ALS. Genome-wide association studies (GWAS) with whole genome sequencing data did not indicate a new risk locus. Coupling GWAS with a coexpression network of disease-associated candidates, points to a significant enrichment for cell cycle- and division-related genes. Within this network, literature text-mining highlights *DECR1*, *ATL1*, *HDAC2*, *GEMIN4*, and *HNRNPA3* as important genes. Finally, information on ALS-related gene variants in the Turkish cohort sequenced within Project MinE was compiled in the GeNDAL variant browser (www.gendal.org).

KEYWORDS
ALS, ALS variant database, genetics, clinical exome sequencing, coexpression network analysis, genome-wide association study, motor neuron disease, next generation sequencing, Turkish peninsula

1 | INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease of upper and lower motor neurons, leading to muscle wasting. The average age of onset (AO) is 55–60 years, however juvenile cases exist (Ghasemi & Brown, 2018). Two-thirds of patients present with spinal-onset, the rest shows bulbar-onset with dysarthria and/or swallowing problems. Cognitive and behavioral changes are seen in almost 50% of ALS cases (van Es et al., 2017). The disease results in death within 2–5 years due to respiratory failure, which can only be slightly extended with exceptional medical care (Brown & Al-Chalabi, 2017; Kiernan et al., 2011).

ALS genetics is complex; the familial form (fALS) is rare (5–10%), sporadic cases or isolated (singlet) patients (sALS) constitute 90% of cases. Familial and sporadic ALS are clinically indistinguishable and well-established fALS genes are implicated in sporadic disease, pointing to "apparently" sporadic cases with incomplete penetrance (Brown & Al-Chalabi, 2017). In populations in which consanguinity is common, juvenile atypical disease accompanies classical, resulting in a more heterogeneous genetic background that makes the differential diagnosis in the clinic challenging.

The Human Genome map, advances in next generation sequencing (NGS), genome-wide association studies (GWAS), and applications of whole exome sequencing (WES) changed paradigms in identifying ALS-associated alleles even with low power, both in family-based studies and in large admixed populations (Brenner et al., 2018; Cirulli et al., 2015; Smith et al., 2017). Moreover, GWAS performed using single nucleotide polymorphisms (SNPs) from whole genome sequencing (WGS) data, enabled association of rare variants with rare diseases, such as ALS (Nicolas et al., 2018; van Rheenen et al., 2016). Project MinE Sequencing Consortium, a large multinational ALS collaboration, is established based on this purpose, to define new disease-causing genes and risk loci associated with true sporadic disease to target novel therapeutics (van Rheenen et al., 2018).

2 | DATA SPECIFICATIONS

Data type	Tables and figures
Data acquisition method	Sanger sequencing, NGS
Data format	Filtered and analyzed
Experimental factors	None
Experimental features	Pathogenic genomic variation analysis in a large Turkish ALS cohort using conventional and next generation sequencing methods. Variant interpretation through in silico tools and clinician-researcher collaboration. Genotype-phenotype correlations in ALS and ALS-like disease. Data collection and sharing in public databases for ALS.
Data source location	Suna and İnan Kıraç Foundation Neurodegeneration Research Laboratory, Koç University Hospital, Davutpaşa Street No. 4, 34010, Istanbul, Turkey
Data accessibility	Data in this paper is published within the paper and deposited to ClinVar public database (https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB7287039).

3 | IMPACT OF DATA

The last decade has seen an unprecedented and exponential progress in data output, based on advances in genetics/genomics and on large international collaborations. Consequently, our knowledge of the genetic factors behind ALS has improved in an unparalleled fashion and the scientific scenario of ALS has dramatically changed. Today, the disease is accepted to be part of a continuum with other neurological diseases and a crossroads between genetic, neurometabolic and environmental factors.

This manuscript, five years apart from our previous publication on ALS in Turkey (Özoğuz et al., 2015), is not only an update with a triple increase in patient numbers, but it supersedes our earlier work by reflecting a whole new picture, being much more comprehensive in its scope, upgraded in cutting edge techniques, applying genome-wide and bioinformatic approaches to extract candidate disease genes and pathways, followed by a population-specific database.

Our understanding of the etiopathology of neurological diseases stem from the identification of disease genes and pathways. With its well-selected and large cohort, this study not only represents a distinct resource for ALS in Turkey, it also reveals the genetic variation in a highly inbred and admixed population, is thus expected to contribute to human disease at large.

4 | EXPERIMENTAL DESIGN, MATERIALS AND METHODS

This study includes 1,200 Turkish patients recruited from hospitals across Turkey between 2002 and 2019; 246 cases with a family history of ALS, plus 80 affected family members and 954 isolated ALS cases. Sample collection was approved by Boğaziçi University Ethics Committee. Genetic counseling was given to patients at the local institutions during blood collection and signed informed consent was obtained from all subjects. DNA samples from healthy relatives were obtained for research purposes only with their written approval. Genomic DNA was isolated from whole blood using the MagNa Pure Compact System (Roche, Switzerland).

4.1 | Screening for common ALS genes

Conventional screening for the hexanucleotide repeat expansion in the *C9orf72* gene was performed in all patients with or without family history of ALS, whereas *SOD1*, *TARDBP*, and *FUS* were screened only in familial cases. The *C9orf72* repeat expansion was tested using repeat-primed polymerase chain reaction (PCR) and flanking PCR was performed to identify the zygosity and the size of the repeats within the normal range. All five exons of *SOD1* were analyzed in fALS. Additionally, exon 4 of the *SOD1* gene was screened in all cases with consanguineous parents, independent of family history. Genomic variant analyses in *TARDBP* and *FUS* genes were restricted to their hotspots, exon 5 for *TARDBP* and exons 14 and 15 for *FUS*. Genotyping experiments were performed with GoTaq® Flexi DNA Polymerase (Promega), MyTaq™ DNA Polymerase (Bioline), FastStart Universal Master Mix (Roche, Switzerland) and One *Taq*® 2× Master Mix (New England Biolabs). The sequences of primers are available upon request. Sanger sequencing was outsourced (Macrogen Inc., Korea) and CLC Main software (Qiagen, Germany) was used for analysis.

4.1.1 | Bisulfite sequencing in the promoter region of *C9orf72*

The 5mC levels of the *C9orf72* promoter regions harboring 26 CpG sites were detected using direct bisulfite sequencing assay (BST-PCR). EZ DNA Methylation-Gold Kit (Zymo Research) was used for bisulfite conversion of genomic DNA according to the manufacturer's protocol. The converted genomic DNA was amplified using nested PCR with primers targeting the converted sequence in the promoter region. ZymoTaq Premix was used for these consecutive amplifications. Methylation levels were detected by direct evaluation of Sanger sequencing results. Commercially available human methylated (100%) and nonmethylated (0%) standards were used as controls, 50% control was prepared by mixing equal amounts of commercial standards (Zymo Research). The number of methylated

CpG sites was calculated for each individual and two-tailed Fisher's exact test was used to assess the association between promoter hypermethylation for the expansion carriers. The maximum number of methylated CpG sites among controls (2/26) was considered as the threshold for hypermethylation.

4.2 | WES

WES was applied in 250 individuals; 127 probands, 32 affected, and 91 healthy family members (Macrogen Inc., Korea). Selection criteria of the patients subjected to WES were (a) close consanguinity in the parents of the affected individual, (b) atypical clinical features, and (c) early/juvenile disease-onset. In addition, WES was also applied to cases with a positive family history of disease in the upper generations, if screening in four common dominant genes did not reveal any disease-associated variants. Suspected inheritance pattern was autosomal dominant for 25, autosomal recessive for 79 families, and in 23 cases the inheritance pattern could not be certified. Clinical information of the cases subjected to WES, their suspected inheritance patterns and initial clinical diagnoses are listed in Appendix (Table A1).

Bioinformatic analyses of the samples were initially performed using in-house Burrows-Wheeler Aligner (BWA) (H. Li & Durbin, 2009) and Genome Analysis Toolkit (GATK; McKenna et al., 2010) pipeline. More recently, the online SEQ Platform, a cloud-based genomics software, was used and all samples have been retrospectively analyzed with this platform (Genomize Inc., Turkey). The SEQ Platform enables calculation of real-time minor allele frequency (MAF) for variants using NDAL- and SEQ-specific cohorts.

For the in-house pipeline, paired-end sequencing reads obtained from sequencing platforms were aligned to the human reference genome GRCh37 plus the decoy using BWA-MEM algorithm. Quality control and variant calling from binary sequence alignment/map format files were performed with HaplotypeCaller tool of GATK. The ANNOVAR software was used for structural and functional annotation of variants (Wang, Li, & Hakonarson, 2010). MAFs were recruited from 1000 Genomes Project (1000 G) and National Heart, Lung and Blood Institute Exome Sequencing Project (NHLBI-ESP6500; Auton et al., 2015; EVS, 2014). Functional consequences of variants were predicted via several sources (e.g., SIFT, PolyPhen2, and GERP++) and DANN scores were assigned to each variant.

For variant prioritization, the association of candidate genes with known human phenotypes was obtained from the OMIM database. Annotated variants were filtered using the VarSifter software (version 1.7) or SEQ Platform according to the inheritance mode and MAF (>0.01; Teer, Green, Mullikin, & Biesecker, 2012). Functional predictions were used for evaluation, but not for filtration. Center (NDAL)-specific MAF lower than 0.01 was used as a parameter during variant prioritization in the WES data of 600 Turkish patients and healthy family members. American College of Medical Genetics (ACMG) guideline verdict was determined for each candidate for further evaluation. Segregation analysis for candidate variants was performed by Sanger sequencing in the index case and in all available family members.

4.3 | WGS

Whole genome sequencing of 632 Turkish sALS cases and 151 neurologically healthy controls was performed within the scope of Project MinE. Samples were selected on the basis of definitive, late-onset ALS diagnosis, without a family history. The mean AO for the patients included in Project MinE was 51 years, in agreement with 52-year-old mean AO of the total sALS cohort; control subjects had a mean age of 55 years. *C9orf72* hexanucleotide repeat expansion was excluded in all patients before WGS. Project MinE guidelines were followed for sample selection and preparation (van Rheenen et al., 2018). PCR-based library free paired-end sequencing was performed on the Illumina HiSeq 2000 platform with an average of 40× coverage per sample (Illumina FastTrack Services, San Diego). Alignment to the hg19 reference genome and variant calling were performed using the Isaac pipeline and provided by Illumina as aligned reads in BAM files and individual-based gVCF files containing the single nucleotide variations (SNVs), short indels and structural variations (Raczy et al., 2013).

Protein coding variants in all ALS-causing and -associated genes reported (Ghasemi & Brown, 2018) were screened in annotated variant files. Candidate variants identified in sALS patients were further analyzed for pathogenicity using prediction tools, VarSome software (Kopanos et al., 2019) and our in-house exome database.

4.3.1 | WGS sample processing and quality control

All variants across the individuals were merged with AGG tool (Illumina). Individual/variant-level quality control was performed using PLINK (version: 1.9; Purcell et al., 2007) and VCF-tools (Danecek et al., 2011) (version:0.1.16). Samples with a deviated inbreeding coefficient (>3 *SD*) from the mean of the distribution, as well as related/duplicate samples and those with missingness rate higher than 10% were not included for further analyses (623 cases and 142 healthy controls remaining). A pruned set of high-quality SNPs were prepared using missingness rate (<10%), MAF (>5%) and Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$ for controls and $p < 1 \times 10^{-12}$ for cases) thresholds. SNPs within the *MHC* or *LCT* loci or, the inversions on chromosome 8/17 were excluded. Principal components (PCs) for each individual were calculated using PLINK.

4.3.2 | Genome-wide association study

Variants with MAF > 5% in the whole cohort were tested for association using a binary logistic regression in PLINK. First 10 PCs and gender were used as covariates.

TABLE 1 Clinical characteristics of the Turkish ALS cohort under study

	Total ALS	fALS	sALS
Number			
Probands	1,200	246 (20%)	954 (80%)
Affected family members	80	80	–
Male:female ratio	1.5	1.2	1.6
Consanguinity	301 (25%)	75 (30%)	226 (24%)
Dementia	30 (2.5%)	13 (5%)	17 (2%)
Age of onset			
Juvenile (<25 years)	101 (8%)	33 (13%)	68 (7%)
Middle (25–45 years)	292 (24%)	61 (25%)	231 (24%)
Late (>45 years)	718 (60%)	123 (50%)	595 (62%)
Not available	89	29	60
Mean age of onset (total ±SD)	50 ± 15.4	47 ± 16.9	51 ± 15.1
Site of onset			
Limb	773 (64%)	163 (66%)	610 (64%)
Bulbar	212 (18%)	36 (15%)	176 (18%)
Limb + bulbar	84 (7%)	12 (5%)	72 (8%)
Not available	131	35	96

Abbreviations: ALS, amyotrophic lateral sclerosis; fALS, familial ALS; sALS, sporadic ALS; *SD*, standard deviation.

4.3.3 | Gene-based burden testing

All variants were annotated using Ensembl Variant Effect Predictor version 92 (McLaren et al., 2016) and classified into two functional groups for gene-based association testing: (a) disruptive variants (stop-gained, stop-loss, start-loss, splice sites, and frameshift indels with high confident according to loftee prediction), and (b) missense variants predicted to be damaging using REVEL and MetaLR algorithms, on the basis of a combining approach (Dong et al., 2015; Ioannidis et al., 2016). ClinVar-benign variants were excluded. ClinVar-pathogenic variants were kept regardless of other elimination criteria (Landrum et al., 2014). In the remaining list, variants with MAF ≥ 1% in any public population databases (ExAC, gnomAD, 1000 G, and ESP6500; Auton et al., 2015; EVS, 2014; Exome Aggregate Consortium, 2016; Karczewski et al., 2019) and variants with MAF ≥ 5% in our cohort were excluded.

Gene-based burden testing was performed using the R-package of SKAT-O by aggregating disruptive and possibly damaging variants (missense variants) on genic regions and pathways (Ionita-Laza, Lee, Makarov, Buxbaum, & Lin, 2013). Pathway and associated lists were downloaded from the Broad Institute GSEA site (http://software.broadinstitute.org/gsea) using all canonical pathway dump (version 6.2). Tests were adjusted for gender and first 10 PCs.

4.3.4 | Gene coexpression network analysis

The ST-Steiner algorithm (Norman & Cicek, 2019) was used, which searches for a connected component (a tree) on a gene coexpression network or a cascade of gene-coexpression networks. The algorithm aims at maximizing the prizes of the selected genes and minimizes the cost of edges that are used to connect these genes.

To construct the gene coexpression network, we utilized the full BrainSpan microarray data set of the Allen Brain Atlas (Sunkin et al., 2013). This data set contains gene expression measurements of 524 brain samples from various brain regions obtained from 42 individuals that represent various time points in neurodevelopment, starting from embryonic period up to adulthood. We used correlation threshold of 0.7 (Pearson correlation, r^2). That is, an edge was added to a graph if two genes' correlation exceeded this threshold. This is a common threshold choice in the literature (Çiçek, 2017; Liu et al., 2014; Liu, Lei, & Roeder, 2015). After pruning for the genes that do not exist in the data, the final network contains 547,056 edges and 8,499 nodes.

As our edge cost $1-r^2$ was used. For the node (gene) prizes, we used the negative $\log_{10}$ transformed *p*-values derived from gene-based SKAT-O analysis. The ST-Steiner algorithm also inputs a list of terminal nodes, which have to be included in the tree (due to a very large artificial prize) and joined by other nodes. These are the genes with high level of risk confidence for ALS: *SOD1*, *TARDBP*, *SQSTM1*, *HNRNPA1*, *FUS*, *VCP*, *OPTN*, *PFN1*, *ATXN2*, *NEFH*, *SETX*, *ALS2*, *DCTN1*, *ANG*, *ELP3*, *FIG4*, *TAF15*, *SPG11*, *NEK1*, *PON1*, *PON3*, *TBK1*, *DAO*, *CHRNA3*, *CHRNB4*, *CREST* (*SS18L1*), *CHRNA4*, *NTE* (*PNPLA6*) (Ghasemi & Brown, 2018).

There are four hyperparameters to set in the ST-Steiner Algorithm. The first parameter is ω, which is the number of trees in an estimated forest. We set this parameter to 0 to obtain a single connected component with the assumption of a single functional cluster of genes as done in Norman and Cicek (2019). λ and α were set to zero, since the algorithm was run on a single network and

these parameters are used when a cascade of networks is employed. The third parameter β was used to put the node prizes and the edge costs on the same scale and adjust the size of the predicted subnetwork. After a line search to obtain a network, which includes approximately three predictions for every ground truth (terminal) gene, it was set to 0.17. Edge thickness denotes the correlation threshold (thicker = higher correlation).

Functional annotation clustering of the candidate genes predicted in the coexpression network was performed using DAVID Functional Annotation Tool (version 6.8; Huang et al., 2007). All 8,499 nodes (genes) used to create the coexpression network were given to the algorithm as background. Literature-mining for association between the predicted genes in the network and ALS was evaluated by screening GeneRif and DisGeNet databases (all species considered; Jimeno-Yepes, Sticco, Mork, & Aronson, 2013; Piñero et al., 2017). We determined the number of occurrences of genes in association with ALS based on specific search matching restricted words: "ALS," "fALS", "sALS," and "amyotrophic lateral sclerosis." The output counts were used as a score to denote the strength of the evidence for each gene with ALS in GeneRIF and DisGeNet databases (Table A2).

5 | DATA

5.1 | 45% of fALS and 10% of sALS are explained by known ALS genes in the Turkish cohort, indicating genetic heterogeneity in fALS and incomplete penetrance among sALS patients

A total of 1,200 Turkish patients diagnosed with ALS or ALS-like motor neuron disease were analyzed within the scope of this study, adopting a combination of conventional and NGS approaches. The clinical summary of the study cohort is compiled in Table 1. Sixty percent of the Turkish ALS cohort under study had an age of disease onset beyond 45 years and the mean ages of onset were 47 for fALS and 51 for sALS patients. In 64% of our cases, spinal symptoms were detected as the initial clinical features, 18% reported to suffer from bulbar symptoms and 7% showed mixed site of onset. The male to female ratio in the present cohort was 1.5.

Four common ALS genes (*C9orf72*, *SOD1*, *TARDBP*, and *FUS*) contribute to 35% of fALS and 6.1% of sALS in Turkey and analysis of pathogenic exonic variants obtained from WES and WGS data increases these numbers to 45% in fALS and 10% in sALS (Figure 1). This 10% of sALS cases explained by genomic variants in well-established and highly penetrant ALS genes, like *C9orf72*, *SOD1*, *TARDBP*, *FUS*, *OPTN* and *VCP*, are "apparently" sporadic, who are either (a) the only affected child of consanguineous couples, or (b) cases with low penetrance of the variant in the upper generations, or (c) carriers of de novo variants.

The GGGGCC hexanucleotide repeat expansion in the *C9orf72* gene was detected in 42 families (plus 8 affected family members) and in 38 sporadic cases in 1,200 ALS patients (Table A3). Mean age of disease onset among the expansion carriers was 54.5, representing classical ALS. A higher frequency of bulbar-onset ALS was observed among C9orf72 cases (23%), compared with 18% in all cases. Intrafamilial phenotypic variability was present among family members manifesting either ALS, ALS accompanied by frontotemporal dementia (ALS-FTD) or solely FTD symptoms. Dementia was reported in nine expansion carriers and two affected family members (13%).

In ALS cases with or without the expansion and in controls, two, five, and eight repeats were found to be the predominant allelic variants in the Turkish population for the nonexpanded allele of the *C9orf72* gene. In three cases the intermediate repeat sized $GGGGCC_{(30\text{-}35)}$ was detected, which did not segregate with the disease in the two families tested. Bisulfite sequencing assay of the 26 CpG sites, located in the promoter region of *C9orf72*, revealed a significant increase in promoter hypermethylation for the expansion carriers ($n = 52$ expansion carriers and 31 age- and sex-matched controls, Student's t test $p < .0023$); no significant correlation was

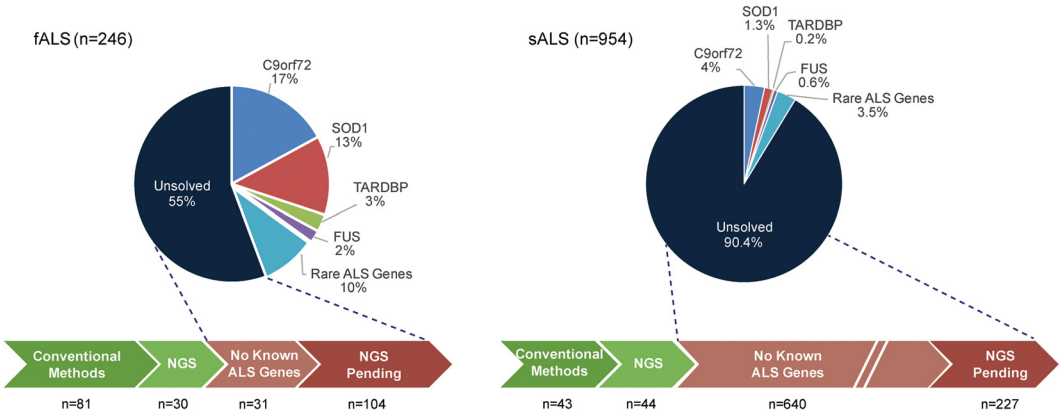

FIGURE 1 Frequency of ALS gene variants in the Turkish cohort. The four major ALS genes account for 35% of fALS, NGS increases this number to 45% (left pie). The same four ALS genes solve 6.1% of sALS, NGS increases it to almost 9% (right pie). The dark blue areas in the pies, represent unsolved cases and also samples not yet analyzed by NGS. ALS, amyotrophic lateral sclerosis; fALS, familial ALS; NGS, next generation sequencing; sALS, sporadic ALS

observed between number of CpGs methylated and the AO of patients (Hamzeiy et al., 2018).

Eighteen distinct pathogenic genomic variants in the *SOD1* gene were identified in 57 patients (32 fALS index cases plus 13 affected family members and 12 sporadic cases; Figure 2; Table 2). The human reference transcript NM_000454.4 was used for the nucleotide numbering of *SOD1*; as an exception, the old nomenclature (excluding the initiation codon) was used for the amino acid changes. The SOD1-p.(Leu144Phe; c.435G>T) Balkan variant (Battistini, Benigni, Ricci, & Rossi, 2013) was observed in 10 probands and seven affected family members, being the predominant *SOD1* variant in the study cohort. The highly characterized SOD1-p.(Asp90Ala; c.272A>C) genomic variant with a dual inheritance pattern, very common among Scandinavian populations in recessive form (Andersen et al., 1995), explained the disease in nine consanguineous Turkish cases. The dual inheritance pattern known for the SOD1-p.(Asp90Ala), was also true for three additional rare changes in *SOD1*; p.(Asn86Ser) (c.205T>C), p.(Leu117Val; c.352G>C), and p.Glu133Lys (c.400G>A), which were detected in 10 different pedigrees with or without family history of ALS (Table 2). Apart from the highly penetrant and frequent pathogenic *SOD1* variations, others identified in our cohort are present in relatively small families with few affected children in the same generation (Table 2). Examples of *SOD1* variants with evidence of reduced penetrance are the p.(Glu22Leu) (c.68A>T), p.(Glu40Gly) (c.122A>G), p.(His71Tyr) (c.214C>T), p.(Val87Met) (c.262G.A) and p.(Thr137Ala) (c.412A>G) variations with asymptomatic carriers in the families.

Pathogenic *TARDBP* (NM_007375.3) and *FUS* (NM_004960.3) genomic variants explained the disease in 20 probands and four affected family members (Figures 1 and 2; Table 2). The heterozygous FUS-p.(Pro525Leu) (c.1574C>T) and FUS-p.(Tyr526Cys) (c.1577A>G) variations were detected in four isolated juvenile cases without a family history, in whom de novo occurrence of the variants was shown via variation-negative parents. Additionally, the intermediate CAG repeat expansions in the *ATXN2* gene, associated with an increased ALS risk, were reanalyzed in an extended cohort of 519 sALS cases as compared with 236 fALS and sALS patients in a previous study from our laboratory (Elden et al., 2010; Lahut et al., 2012). Analysis, using the control cohort of Lahut et al. (2012) ($n = 420$), confirmed increased ALS risk in carriers with (27–33) CAGs (21/519; Fisher's exact test $p = .0086$).

Analysis of exonic variants in the WES ($n = 127$) and WGS ($n = 623$) data revealed the presence of pathogenic gene variants in 74 cases, out of which, 19 were previously published by our group (Tables 2 and 3; Akçimen et al., 2019; Özoğuz et al., 2015; Tunca et al., 2018). Variant information and clinical features of the cases solved are compiled in Table 3 and in Appendix (Table A4). Homozygous *OPTN* variants were observed in a total of eight cases in the current Turkish cohort. In five families out of eight, the homozygous AA deletion in the *OPTN* gene leading to a premature stop codon (p.(Lys360Valfs*18), c.1078_1079delAA, NM_021980.4) was identified. *OPTN*-based disease in our cohort, presents as classical ALS with an earlier onset at 38 years on average (Table 3). *SPG11* and *ALS2* genomic variants are the second most frequent causes of autosomal recessive ALS in the cohort, with average ages of onset of 14 and 1, respectively.

Rare autosomal dominant ALS genes predominating in our cohort include *VCP*, *ANG*, and *TBK1*, identified in both familial and "apparently sporadic" cases without any reported disease history in the family (Table 3). The novel heterozygous *ERBB4* (p.Arg1096Cys, c.3286C>T, NM_001042599.1) and *KIF5A* (p.Asp1002Gly, c.3005A>G, NM_004984.2) pathogenic genomic variations were detected in two distinct families, in which the causative variants segregated with the disease in at least three affected family members. Among the ALS gene variants with unknown pathogenicity identified through WGS, two variations in the *PON1* and *PON3* genes, were

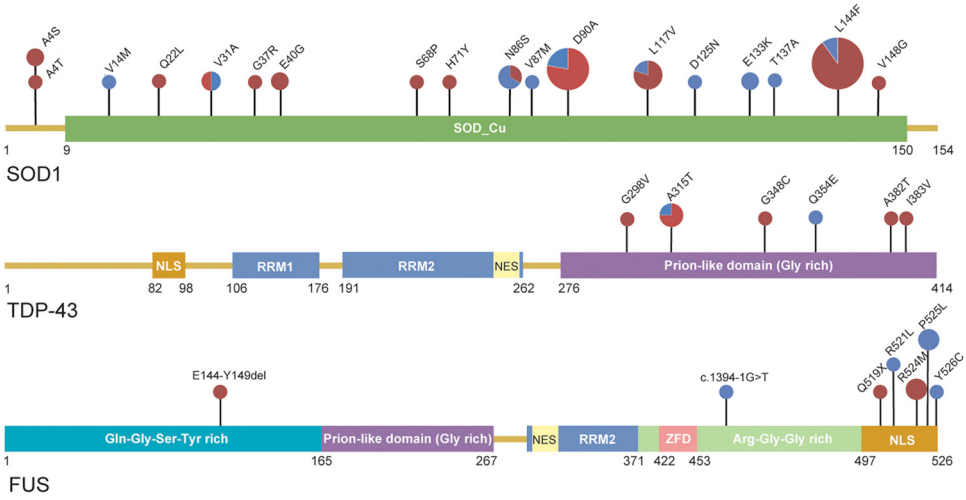

FIGURE 2 Amino acid changes identified in SOD1, TDP-43, and FUS proteins. Variant-specific pie charts represent the variant's proportion in fALS (red) or sALS (blue) cases, smallest circle corresponding to one case and the largest to 11 cases. fALS, familial ALS; NES, nuclear export signal; NLS, nuclear localization signal; RRM, RNA recognition motif; sALS, sporadic ALS; ZFD, Zinc finger domain

TABLE 2 Clinical data of patients with *SOD1*, *TARDBP*, *FUS* genomic variants

Gene	ALS ID	Nucleotide change	Protein change[c]	Gender	Age of onset	Site of onset	Gene dosage	Family history	Phenotype
SOD1(NM_000454.4)	**1398**	c.13G>A	p.(Ala4Thr)	F	25	B	het	Yes	Juvenile ALS
	221	c.13G>T	p.(Ala4Ser)	M	20	L	het	Yes	Juvenile ALS
	308			F	44	L			ALS
	907			F	NA	NA			ALS
	1167[a]	c.43G>A	p.(Val14Met)	M	42	L	het	No	ALS
	960	c.68A>T	p.(Gln22Leu)	F	30	L	het	Yes	ALS
	1327	c.95T>C	p.(Val31Ala)	M	45	B	het	Yes	ALS
	1547[a]			F	64	L		No	ALS
	1453	c.112G>C	p.(Gly37Arg)	M	41	L	het	Yes	ALS
	802	c.122A>G	p.(Glu40Gly)	F	39	L	het	Yes	ALS
	816			F	32	L		Yes	ALS
	1450	c.205T>C	p.(Ser68Pro)[n]	M	54	L	het	Yes	ALS
	226	c.214C>T	p.(His71Tyr)[n]	M	19	L	het	Yes	Juvenile ALS
	707			F	57	L			ALS
	191	c.260A>G	p.(Asn86Ser)	M	28	L	hom	No	ALS
	623[a]			F	42	L	het	No	ALS
	1207			M	48	L		Yes	ALS
	102[a]	c.262G>A	p.(Val87Met)	F	29	L	het	No	ALS
	147	c.272A>C	p.(Asp90Ala)	M	49	L	hom	Yes	Lower limb dominant stereotyped Scandinavian phenotype
	310			M	55	L		Yes	
	429			F	45	L		Yes	
	741			F	32	L		Yes	
	810			F	29	L		Yes	
	1256[b]			M	44	L		Yes	
	1359			M	35	L+B		Yes	
	1545			F	64	L		No	
	1579			M	51	L		No	
	561	c.352C>G	p.(Leu117Val)	F	62	L	het	Yes	ALS
	1527			F	38	L			ALS
	1396			F	62	L		Yes	ALS
	1412			F	40	L			ALS
	1472[a]			F	36	L		No	ALS
	1888			M	50	L		Yes	ALS
	1882			F	58	L			ALS
	1439			F	24	L	hom	Yes	juvenile ALS
	355[a]	c.376G>A	p.(Asp125Asn)	M	50	L	het	No	ALS
	1716[b]	c.400G>A	p.(Glu133Lys)	F	37	L	hom	No	ALS
	1064[a]			M	34	L	het	No	ALS
	1655[b]	c.412A>G	p.(Thr137Ala)	F	49	L	het	No	ALS
	61	c.435G>T	p.(Leu144Phe)	F	52	L	het	Yes	ALS
	281[a]			M	57	L		No	ALS
	607			F	45	L		Yes	ALS
	713			F	53	L			ALS
	724			M	52	L			ALS
	727			F	NA	L			ALS
	1773			M	60	L			ALS
	635			F	54	L		Yes	ALS
	772			M	49	L		Yes	ALS
	1059			F	51	L			ALS
	1063			F	56	L			ALS
	1235			F	59	L			ALS

TABLE 2 (Continued)

Gene	ALS ID	Nucleotide change	Protein change[c]	Gender	Age of onset	Site of onset	Gene dosage	Family history	Phenotype
	935			M	37	L		Yes	ALS
	1036			F	60	L		Yes	ALS
	1633			M	64	L		Yes	ALS
	1691			M	60	L		Yes	ALS
	1715			F	34	L		Yes	ALS
	97[b]	c.446T>G	p.(Val148Gly)	F	46	L	het	Yes	ALS
TARDBP (NM_007375.3)	**1082**	c.893G>T	p.(Gly298Val)[n]	F	66	L	het	Yes	ALS
	356[b]	c.943G>A	p.(Ala315Thr)	M	58	L	het	Yes	ALS
	357			F	59	L		Yes	ALS (man in barrel)
	600			F	57	L		Yes	ALS
	1448			M	62	L		Yes	ALS
	408[a]			M	48	L		No	ALS
	910	c.1042G>T	p.(Gly348Cys)	M	37	L	het	Yes	ALS
	919			M	42	L			ALS
	911			M	NA	L			ALS
	660	c.1060C>G	p.(Gln354Glu)[n]	F	42	L	het	No	ALS
	976	c.1144G>A	p.(Ala382Thr)	M	39	L	het	Yes	ALS
	277[b]	c.1147A>G	p.(Ile383Val)	M	67	B	het	Yes	ALS
	311			F	42	NA			ALS
FUS (NM_004960.3)	**264**[b]	c.430_447del	p.(Glu144_-Tyr149del)	M	16	L	het	No	Juvenile ALS
	1034	(NC_000016.10) c.1394−1G>T[n]	p.(=)	F	47	L	het	Yes	ALS
	1208	c.1555C>T	p.(Gln519*)	F	42	L		Yes	ALS
	485[a]	c.1562G>T	p.(Arg521Leu)	M	39	L	het	No	ALS
	227[b]	c.1571G>T	p.(Arg524Met)	F	32	L	het	Yes	ALS
	581			M	53	L		Yes	ALS
	1647			F	29	L		Yes	ALS
	549	c.1574C>T	p.(Pro525Leu)	M	14	L	het (de novo)	No	Juvenile ALS with fast progression
	1610[b]			M	17	L	het (de novo)	No	
	377[a]			F	16	L	het (de novo)	No	
	1423[b]	c.1577A>G	p.(Tyr526Cys)	M	12	L	het (de novo)	No	

Note: Bold indicates index cases.

Abbreviations: B, bulbar; het, heterozygous; hom, homozygous; L, limb; n, novel; NA, not available; WES, whole exome sequencing; WGS, whole genome sequencing.

[a]Identified in the framework of Project MinE (WGS).

[b]Identified with WES.

[c]The old nomenclature of the *SOD1* gene is adopted here for amino acid changes.

detected in Turkish sALS cases as compared to none of the controls (*PON1*, rs755475189, 2 patients/0 controls; *PON3*, rs147006695, 5 patients/0 controls; Table A5).

Apart from classical ALS genes, WES revealed variants in rare genes associated with diverse motor neuron phenotypes either with upper or lower motor neuron predominance (e.g., *ZFYVE26*, *DNAJB2*, *PLEKHG5*, *TRPV4*, *FBXO38*, and *VRK1*; Table 3 and Table A4). The *C19orf12* genomic variant, implicated in neurodegeneration with brain iron accumulation, mimicking ALS, was detected in three cases with an autosomal recessive inheritance (AO: 9–24), who were initially diagnosed with a probable juvenile ALS (Deschauer et al., 2012).

From the clinical exome sequencing perspective, WES-only success rate of NDAL is calculated as 42% for patients diagnosed with ALS and ALS-like disease in the Turkish cohort (Tables 2 and 3, Table A4). This rate increases to almost 50% in familial cases and in cases with consanguineous parents, regardless of family history.

5.2 | GeNDAL, a web-based variant browser for ALS-related genes

Fully anonymized information regarding ALS-related variants with known or unknown pathogenicity identified in WGS analysis are

presented in the Genome Browser of NDAL, GeNDAL (http://www.gendal.org). GeNDAL is a platform which allows the users to query variants by dbSNP ID, amino acid change, gene symbol, Human Genome Variant Server ID, transcript or sequence ontology defined by the Sequence Ontology Consortium (http://www.sequenceontology.org/). Detailed variant annotations and graphical representations of variant-related information from public databases (ClinVar, gnomAD, etc.) can be visualized (Figure 3). In addition, the phenotypes can be distinguished as ALS or unaffected control. The GeNDAL database currently constructed for ALS-related gene variants will be complemented in future for other phenotypes in NDAL's cohort.

5.3 | Whole genome sequencing analysis of 623 Turkish sALS cases and 142 neurologically healthy controls did not reveal significant risk loci

A joint cohort of ALS patients and control samples worldwide were analyzed by Project MinE Sequencing Consortium, which included 224 Turkish samples (van der Spek et al., 2019; van Rheenen et al., 2016). Analysis of WGS data of an expanded Turkish cohort consisting of 623 Turkish sALS patients and 142 neurologically healthy controls, revealed 47,971,649 novel variants which are not represented in gnomAD. Out of all variants detected, 23,410,513 had a MAF smaller than 0.1% in the cohort (Table A6). GWAS analysis and gene-based burden testing (SKAT-O) on this cohort did not reveal a significant risk/protective variant or gene (Figure 4). The top 25 SNPs detected in GWAS and the top ten genes from SKAT-O are listed in Tables A7 and A8 in the Appendix. Even though there is a lack of association with new variants or genes from the GWAS, we scanned the literature for the candidate genes and none of them was associated with ALS or similar phenotypes.

On the basis of the hypothesis that ALS genes are working as a functional cluster, we conducted a gene coexpression network analysis (a) to search for other candidate genes which might also confer ALS risk, and (b) to investigate the function of the predicted cluster and its relation to ALS. The "guilt-by-association principle"-based gene discovery approach has been applied on many complex neurologic/psychiatric disorders to discover more risk genes as in the autism spectrum disorder (De Rubeis et al., 2014; Sanders et al., 2015) and schizophrenia (Torkamani, Dean, Schork, & Thomas, 2010). For this purpose, the ST-Steiner algorithm (Norman & Cicek, 2019) was used to create a network around established ALS genes using the SKAT-O *p*-values as the prize for the network analysis. The resulting subnetwork contains 98 newly predicted genes around the 28 ALS-associated terminal genes (Figure 5). Ninety-five of 98 predicted genes had higher variation rates in patients and were significantly enriched for cell cycle (DAVID enrichment score: 15,03) and cell division genes (DAVID enrichment score: 6,8; Table A9; Huang et al., 2007). Coexpression network analysis, coupled with literature text-mining using GeneRIF and DisGeNet databases, pointed to *DECR1* (case count: 3 and control count: 0), *ATL1* (case count: 1 and control count: 0), *HDAC2* (case count: 1 and control count: 0), *GEMIN4* (case count: 1 and control count: 0) and *HNRNPA3* (case count: 1, control count: 0) genes which are marked in Figure 5. Finally, accumulation of all variants in WGS to canonical disease pathways obtained from Broad Institute also did not suggest a significant association (Figure A1); top ranking pathways are listed in Table A10. Even with a Turkish sample size four times larger than the two initial Project MinE studies, neither GWAS nor network analysis point to any known or new significant association with ALS, indicating the crucial requirement for larger sample sizes.

6 | DISCUSSION

6.1 | Clinical presentation of ALS in Turkey

For almost two decades, well-established patient registers operating in European countries gather organized patient data to understand the epidemiology of ALS (Hardiman et al., 2017). These registers work countrywide and are unbiased in terms of origin, socioeconomic status and the disease stage of the patient in contrast to a local clinic. The incidence reported for ALS worldwide is argued to be misleading in the absence of long-running patient registers which are more efficient in recognizing family history and hidden symptoms like FTD (Hardiman et al., 2017). Turkey's ALS patient registry operated by the Turkish ALS-MND Association (www.als.org.tr) since 2001 with headquarters in İstanbul and İzmir, is relatively recent and may not represent the whole country. Hence, information regarding the incidence and prevalence of ALS in Turkey and survival rates are still restricted and scattered. NDAL, being the only reference center for the molecular analysis of ALS in Turkey, has been recruiting patients from across the country for over 20 years and gathers available patient data to investigate the clinical and molecular basis of this complex neurodegenerative disease in an admixed population inhabiting the Turkish peninsula in the crossroads of many civilizations since several centuries.

This study, with 1,200 probands, offers an update on the phenotypic and genetic landscape of ALS in Turkey. The fALS percentage of 20% (246/1,200), exceeding North American and European populations (5–10%; Ghasemi & Brown, 2018), is explained by population-specific and social factors, such as extensive kindreds consisting of many generations and offspring. A unique aspect in Turkey, common to countries in the Near and Middle East, is the high proportion of close consanguineous marriages, approaching 50% in the eastern parts of the country. Consanguinity in the ALS cohort under study is calculated as 30% in fALS and 23% in sALS. This suggests even a higher percentage for Mendelian inheritance in yet unexplained cases that are classified originally as sALS due to singlet patients in the family. Thus, familial ALS, harboring a simplex genetic component, seems to be above 20% among Turkish cases.

6.2 | Impact of common genes on ALS in Turkey

The *C9orf72* hexanucleotide repeat expansion and pathogenic missense variants in *SOD1* together explain 30% of familial cases in our

TABLE 3 Variants identified via WES and WGS in rare genes

Gene	Transcript ID	Variation		Gene dosage	AO	Family history	Consanguinity	Phenotype
		DNA change	Protein change					
ERBB4	NM_001042599.1	c.3286C>T	p.(Arg1096Cys)	Het	48	Yes	No	ALS
KIF5A	NM_004984.2	c.3005A>G	p.(Asp1002Gly)[n]	Het	50	Yes	No	ALS
TBK1	NM_013254.4	c.922C>T	p.(Arg308*)[n][a]	Het	46	No	No	ALS
		c.1436_1437delTG	p.(Val479Glufs*4)	Het	20	No	No	Juvenile ALS
VCP	NM_007126.5	c.463C>T	p.(Arg155Cys)[a]	Het	62	No	No	ALS
		c.475C>T	p.(Arg159Cys)[a]	Het	52	No	No	ALS
		c.572G>C	p.(Arg191Pro)	Het	60	Yes	No	ALS-FTD
UBQLN2	NM_013444.3	c.1516C>T	p.(Pro506Ser)	Hemi	26	Yes	No	ALS
		c.1573C>T	p.(Pro525Ser)	Hemi	22	Yes	Yes	ALS
TFG	NM_001007565.2	c.854C>T	p.(Pro285Leu)	Het	47	Yes	No	ALS with sensory neuropathy
ANG	NM_001145.4	c.208A>G	p.(Ile70Val)[a]	Het	52	No	No	ALS
		c.208A>G	p.(Ile70Val)[a]	Het	40	No	No	ALS
		c.208A>G	p.(Ile70Val)	Het	28	No	No	ALS
CHCHD10	NM_213720.2	c.176C>T	p.(Ser59Leu)[a]	Het	51	No	No	ALS
FBXO38	NM_001271723.1	c.1577G>A	p.(Arg526Gln)[n]	Het	congenital	No	Yes	Juvenile MND
TRPV4	NM_147204.2	c.943C>T	p.(Arg315Trp)	Het	infancy	Yes	Yes	Juvenile MND
TRPM7	NM_017672.6	c.4445C>T	p.(Thr1482Ile)	Het	teenage	Yes	No	Juvenile ALS
SETX	NM_015046.7	c.5839G>A	p.(Ala1947Thr)[n]	Het	11	Yes	No	Juvenile ALS
ERLIN1	NM_006459.4	c.281T>C	p.(Val94Ala)[n]	Hom	15	Yes	Yes	Juvenile ALS
SPG11	NM_025137.4	c.1432C>T	p.(Gln478*)[n]	Hom	20	No	Yes	Juvenile ALS
		c.1966_1967delAA	p.(Lys656Valfs*11)	Hom	16	Yes	Yes	Juvenile ALS
		c.2250delT	p.(Phe750Leufs*3)[n]	Hom	16	Yes	Yes	Juvenile ALS
		c.7155T>G	p.(Tyr2385*)[n]	Hom	23	Yes	Yes	Juvenile ALS
OPTN	NM_021980.4	c.76delC	p.(His26Thrfs*19)[a]	Hom	35	No	Same village	ALS
		c.875dupC	p.(Glu293Glyfs*19)[n]	Hom	33	Yes	Yes	ALS
		c.1078_1079delAA	p.(Lys360Valfs*18)	Hom	32	Yes	Yes	ALS
		c.1078_1079delAA	p.(Lys360Valfs*18)	Hom	43	Yes	Yes	ALS
		c.1078_1079delAA	p.(Lys360Valfs*18)	Hom	31	Yes	Yes	ALS
		c.1078_1079delAA	p.(Lys360Valfs*18)[a]	Hom	42	No	No	ALS
		c.1078_1079delAA	p.(Lys360Valfs*18)[a]	Hom	42	Yes[b]	No	ALS
		c.1217delC	p.(Thr406Lysfs*5)[n][a]	Hom	42	No	No	ALS

(Continues)

TABLE 3 (Continued)

Gene	Transcript ID	Variation		Gene dosage	AO	Family history	Consanguinity	Phenotype
		DNA change	Protein change					
ALS2	NM_020919.4	c.1718C>A	p.(Ala573Glu)	Hom	2,5	No	Yes	Juvenile ALS
		c.2761C>T	p.(Arg921*)	Hom	1	No	Yes	juvenile ALS
		c.4381C>T	p.(Arg1461*)[n]	Hom	1	Yes	Yes	Juvenile ALS
		c.4808C>T	p.(Pro1603Leu)	Hom	1	No	Yes	Juvenile ALS
C19orf12	NM_001031726.3	c.32C>T	p.(Thr11Met)	Hom	24	No	Yes	Juvenile ALS
		c.194G>T	p.(Gly65Val)	Hom	10	Yes	Yes	Juvenile ALS
		c.194G>T	p.(Gly65Val)	Hom	9	No	Yes	Juvenile ALS
SYNE1	NM_182961.3	c.22930C>T	p.(Gln7644*)[n]	Hom	21	Yes	Yes	Juvenile ALS
		c.23524C>T	p.(Arg7842*)	Hom	17	Yes	Yes	Juvenile ALS
ZFYVE26	NM_015346.3	c.2074delC	p.(Leu692Serfs*52)[n]	Hom	17	No	Yes	Juvenile MND
		c.2615–2617delGCTinsTGAA	p.(Arg872Hisfs*17)[n]	Hom	22	No	Yes	Juvenile MND
DNAJB2	NM_006736.5	c.14A>G	p.(Tyr5Cys)	Hom	31	No	Yes	Juvenile ALS
		c.757G>A	p.(Glu253Lys)	Hom	22	No	Yes	Juvenile MND
PLEKHG5	NM_198681.3	c.1648C>T	p.(Gln550*)	Hom	20	Yes	Yes	Juvenile ALS
		c.2120C>A	p.(Pro707His)	Hom	14	No	Yes	Juvenile ALS
SIGMAR1	NM_147157.2	c.355G>A	p.(Glu119Lys)	Hom	2	No	Yes	Juvenile ALS
		c.358A>G	p.(Thr120Ala)[n]	Hom	17	No	Yes	Juvenile ALS
VRK1	NM_003384.3	c.961C>T	p.(Arg321Cys)	Hom	22	Yes	Yes	Juvenile ALS
		c.1135_1136delCA	p.(Gln379Aspfs*23)[n]	Hom	17	No	Yes	Juvenile MND
DJ1	NM_007262.5	c.133C>T	p.(Gln45*)	Hom	24	Yes	Yes	ALSPDC
IGHMBP2	NM_002180.2	c.638A>G	p.(His213Arg)	Hom	9	Yes	Yes	Juvenile MND
SLC52A3	NM_033409.4	c.802C>T	p.(Arg268Trp)	Hom	1,5	Inconclusive	Yes	Madras MND

Abbreviations: ALS, amyotrophic lateral sclerosis; ALSPDC, ALS-Parkinsonism-Dementia Complex; AO, age of onset; fALS, familial ALS; Het, heterozygous; Hemi, hemizygous; Hom, homozygous; MND, motor neuron disease; sALS, sporadic ALS.

[a]Cases solved in the framework of Project MinE (WGS).

[b]Patient status changed from sALS to fALS upon diagnosis of a younger sister with ALS, [n]novel variant.

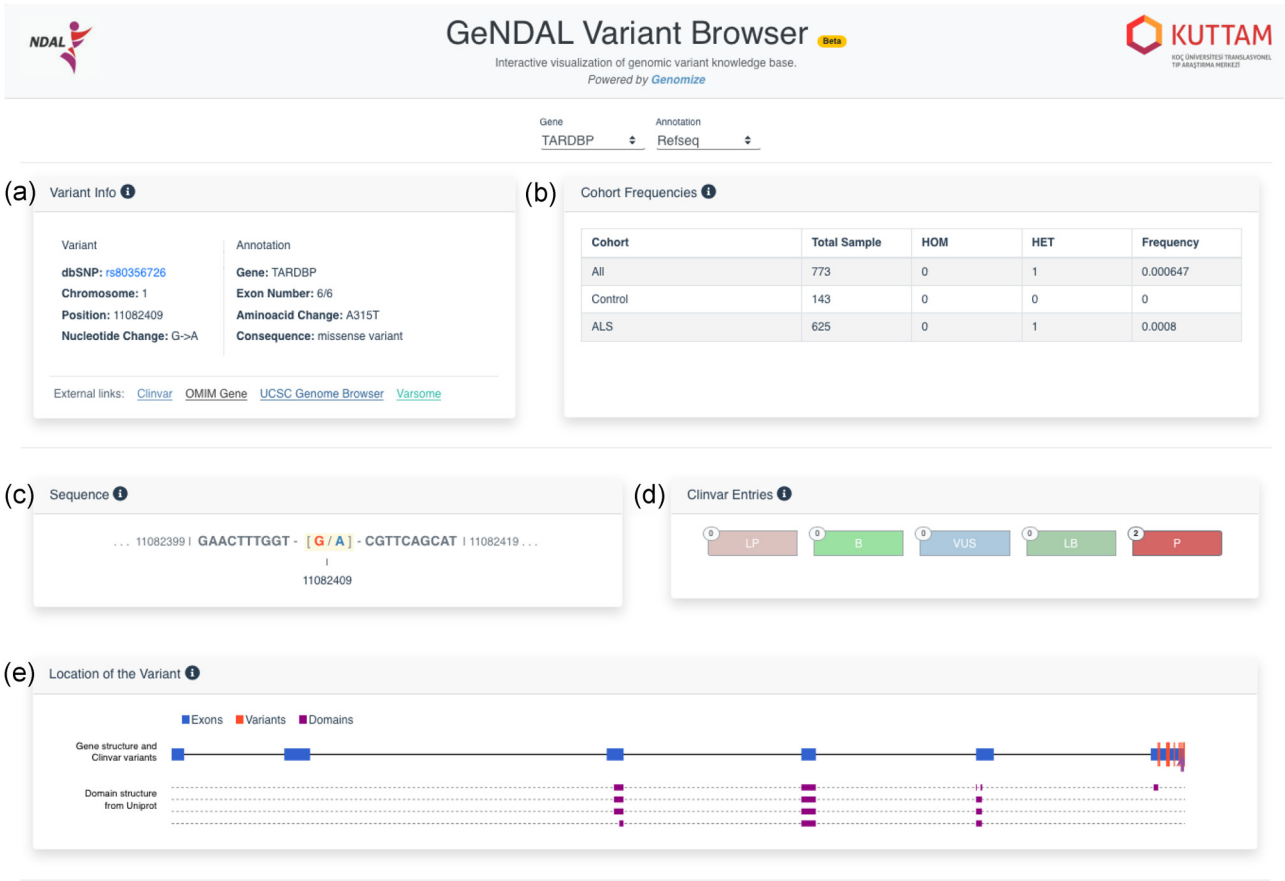

FIGURE 3 Representation of the GeNDAL variant database. (a) Information regarding the variant and its annotation with complementary links to external databases. (b) Phenotype-dependent frequencies of the variant of interest in the internal WGS cohort of NDAL. (c) Genomic location of the nucleotide change and its surrounding sequence. (d) The number of pathogenicity verdict and detailed information on the variant in ClinVar. (e) Display of pathogenic and likely pathogenic variants reported in ClinVar aligned to the current transcript and protein domains, allowing visualization of variational hotspots. WGS, whole genome sequencing

cohort, with *TARDBP* and *FUS* solving another 5%. NGS data of "apparently sporadic"/isolated cases, unraveled genomic variants in common ALS genes that contribute to 2.1% (*SOD1*: 1.3%, *TARDBP*: 0.2% and *FUS*: 0.6%), to total 6.1% with sporadic cases carrying the *C9orf72* expansion. These results not only support the evidence of incomplete penetrance but also the de novo occurrence of variants in these genes, leading to genetic misclassification of patients as sporadic. Thus, the distinction between the clinically indistinguishable familial and sporadic disease, becomes unclear and should be handled with care in diagnostic settings and particularly during genetic counseling.

The hexanucleotide repeat expansion in the *C9orf72* gene is the most common genetic cause of ALS worldwide, with an exception of Japan (Ogaki et al., 2012). Although lower in frequency compared to Northern European countries (50% of fALS and 20% of sALS; Majounie et al., 2012), this expansion is the most abundant genomic variation both in fALS (17%) and sALS (4%) also in the present Turkish cohort. *SOD1* variants, the second most frequent genetic causes of ALS in the Turkish cohort contribute to high allelic heterogeneity (Figure 2). Evidence for the reduced penetrance of *SOD1* variants was obvious in the Turkish population, such as (a) the genomic variants detected in sporadic patients, (b) the families with asymptomatic carriers, and (c) dual inheritance patterns observed in SOD1-p.(Asn86Ser), p.(Asp90Ala), p.(Leu117Val), and p.(Glu133Lys). This variability may stem from modifier genes/variants, an upcoming research field.

The fact that an ample number of people with Turkish origin migrated to Turkey from the Balkans many generations ago, rationalizes the predominance of the common Balkan variant, SOD1-p.(Leu144Phe) in our cohort. The average AO for p.(Leu144Phe) carriers is 52 years without any gender bias and all of them have limb-onset disease. This variation results in classical ALS and appears to be highly penetrant in large families consisting of several branches. The only exception is an apparently sporadic male patient (AO:57), with deceased parents who were not available for analysis (Table 2).

The biallelic p.(Asp90Ala) variation is the second most common pathogenic *SOD1* variant in our cohort. No affected individuals have been detected carrying the heterozygous variant and the presence of

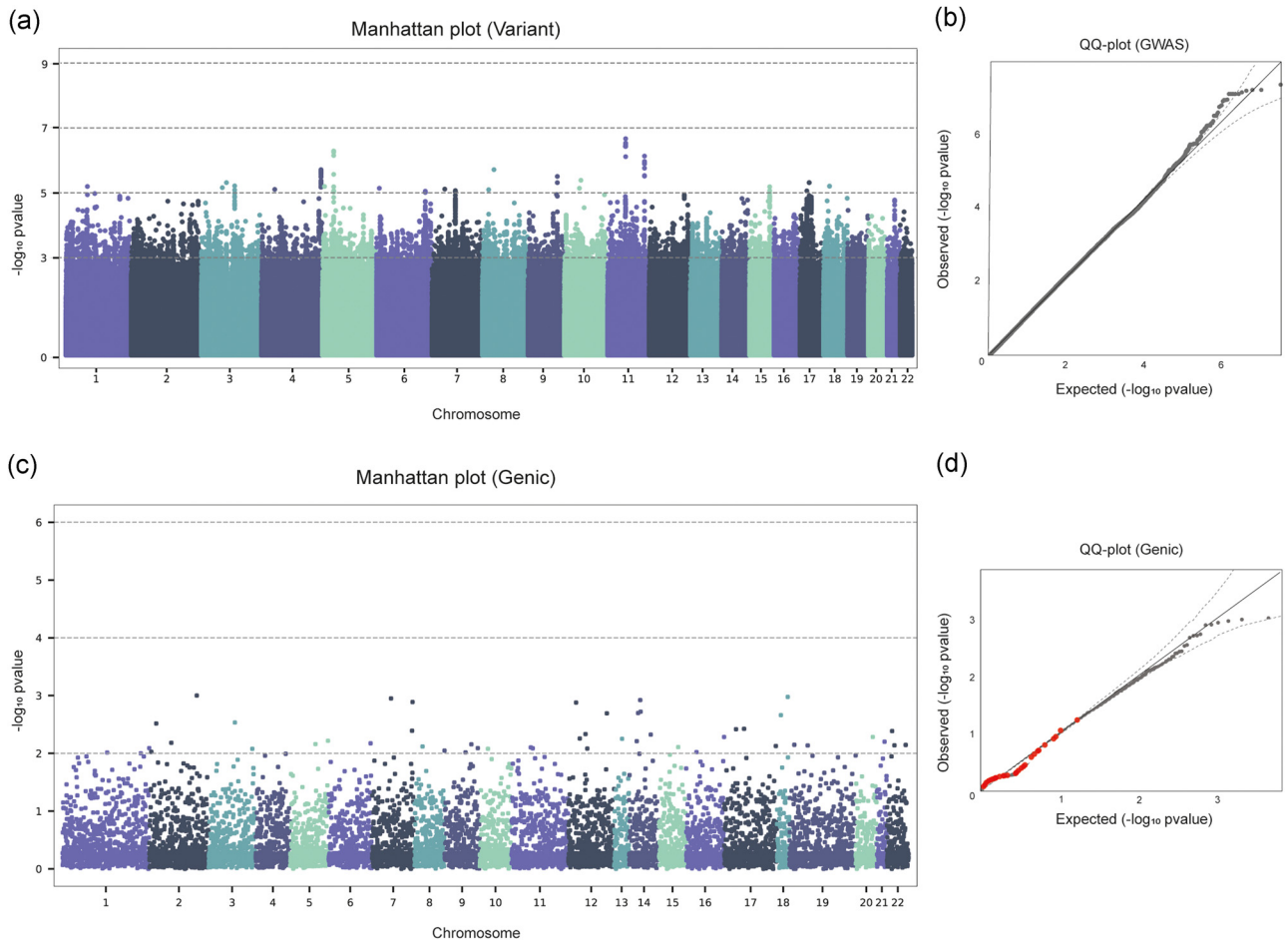

FIGURE 4 Manhattan plots and Quantile-quantile plots of GWAS and SKAT-O analysis using logistic regression. (a) Approximately six million SNPs (MAF ≥ 0.05) are displayed in the Manhattan plot. (b) Quantile-quantile plot of the GWAS *p* values. (c) *p* values derived from the gene-based SKAT-O analysis are displayed in the Manhattan plot. Each of ~10,000 genes in SKAT-O analysis is represented by a single dot. (d) Quantile-quantile plot of genic association *p* values. Known ALS-related genes are highlighted in red. Dashed curves correspond to the 95% confidence limits. ALS, amyotrophic lateral sclerosis; GWAS, Genome-wide association studies

the Scandinavian founder haplotype in Turkish recessive p.(Asp90Ala) cases was previously shown (Özoğuz et al., 2015). Except for one patient with a mixed site of onset, limb-onset disease predominates in p.(Asp90Ala) patients. The disease progression is very slow and in accordance with the stereotyped Scandinavian phenotype (Andersen et al., 1995). The average AO of the recessive p.(Asp90Ala) carriers is 10 years earlier than the p.(Leu144Phe) patients.

SOD1 is a ubiquitously expressed protein that acts as a superoxide radical scavenger in the cell. Its pathogenicity in ALS is explained by the misfolding of the mutant product which leads to accumulation within the cell in aggregates (Paré et al., 2018). There is not enough evidence in the literature to comment on the mechanism behind the phenotypic heterogeneity of different SOD1 variants as well as their acting mechanisms that show both dominant and recessive inheritance. Although most evidence supports the gain-of-function mechanism, loss-of-function of the mutant allele may still have a role in the presentation of *SOD1*-based disease. Reduction in overall activity of the mutant form has been shown in blood and fibroblast samples, but the two specific genomic variants, p.(Asp90Ala) and the heterozygous p.(Leu117Val), had only slight reductions in enzymatic activity. This might explain the milder phenotype in patients carrying these variants and the low penetrance observed in parents of homozygous individuals (Saccon, Bunton-Stasyshyn, Fisher, & Fratta, 2013). On the contrary, the homozygous p.(Leu117Val) genomic variation was reported to result in a more severe reduction in enzymatic activity than the heterozygous variant which is also concordant with the early AO (AO:24) and the fast disease progression of the patient with the biallelic genomic variation reported in this study (Table 2; Synofzik et al., 2012). Although the function of the mutant protein is not completely lost, the activity may be reduced by aggregation; thus, different SOD1 variants with different aggregation propensities may have variable enzymatic activity and this may act on the phenotypic representation of the disease.

ALS-associated TDP-43 and FUS variants are known to accumulate in the C-termini and although the pathogenicity behind these two RNA/DNA binding proteins is not yet clear, nuclear clearance and cytoplasmic accumulation of both proteins are observed in ALS.

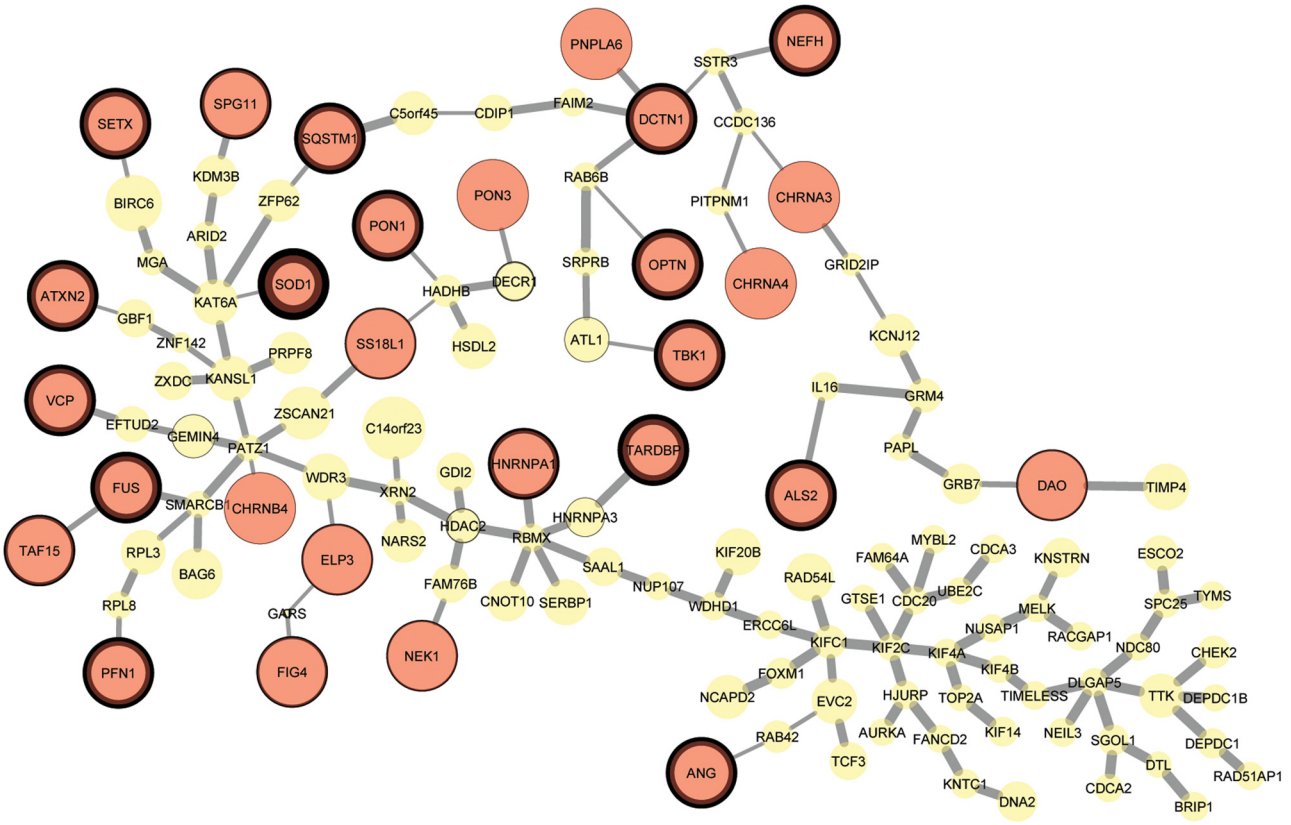

FIGURE 5 The predicted subnetwork of genes by ST-Steiner on ALS GWAS data. The predicted subnetwork contains 126 genes. Red denotes ground truth (terminal) genes ($n = 28$) used to build the network, and yellow denotes the newly predicted genes ($n = 98$). The node size represents ALS risk, based on −log10 transformed p values from SKAT-O analysis. The border thickness depicts the GeneRIF and DisGeNet scores for each gene, and edge thickness the strength of gene expression correlation between a pair of genes according to the BrainSpan database. ALS, amyotrophic lateral sclerosis; GWAS, Genome-wide association studies

In fact, TDP-43-positive cytoplasmic inclusions are a common hallmark of fALS and sALS, regardless of an ALS-associated genetic variation in patients. Unlike FUS variants gathered in the nuclear localization signal domain, TDP-43 variants are found in the prion-like domain of the protein (Figure 2). Our results regarding *TARDBP* and *FUS* variants are restricted to screening of C-terminal hotspots for these two genes with the exception of the heterozygous deletion in the N-terminus of the protein detected via WES. There are rare variants reported in the N-terminal region of TDP-43, like the p.Ala90Val, however with the lack of segregation analysis, the presence of the variant in healthy individuals and with mild abnormal cytoplasmic localization, the pathogenicity of the variant remains questionable (Winton et al., 2009; Wobst et al., 2017). Despite the importance of nuclear import and export signals on the transport of a protein, studies showed different cytoplasmic accumulation levels for different N-terminal FUS variants and also a critical role for C-terminal deletions in FUS in formation of stress granules, all suggesting a complicated mechanism for both proteins, ranging from loss of nuclear function to gain of toxic function through aggregates (Guerrero et al., 2016).

Juvenile ALS (AO < 25) was observed in 68 isolated/sporadic cases (7%) in our cohort. This form of ALS most frequently occurs due to consanguinity and has a rather slow disease progression compared to classical ALS. However, four de novo FUS cases with non-consanguineous parents, ages of onset ranging from 12 to 17, had an aggressive disease progression, resulting in the retirement of the children from all daily activities. Severe bulbar symptoms in addition to initial limb-onset disease, eventually lead to death almost within a year. De novo *FUS* gene variants are reported in juvenile cases in populations where consanguinity is not common. We also suggest the screening of *FUS* as the initial step in isolated juvenile patients with a fast disease progression and asymptomatic parents (Hübers et al., 2015; Leblond et al., 2016; Therrien, Dion, & Rouleau, 2016).

In the cohort under study, common ALS genes contribute to 35% of fALS, which increases to 45% with the addition of rare genes. According to this picture, more than 50% of Turkish fALS cases remain unsolved as compared to 30% in Caucasian populations (Ghasemi & Brown, 2018); this result points towards an expected higher locus heterogeneity in the Turkish population. The Turkish peninsula, geographically located at the intersection of many civilizations, has a heterogeneous ethnic and genetic background. This complexity in the population leads to the dilution of pathogenic variants in common ALS genes like *C9orf72* or *SOD1*. In this sense, the frequencies observed in Turkey are concordant with the common notion of decreasing north-south gradient for these genes (Andersen, 2006; Lamp et al., 2018; Smith et al., 2013). Novel coding variants, as well as

chromosomal changes (large rearrangements, copy number variations, repeat expansions, indels), and variants in regulatory, intronic and intergenic regions, not covered by WES, are expected to unravel the missing heritability in the present cohort.

6.3 | Clinical exome sequencing in the differential diagnosis of ALS and ALS-like phenotypes

The majority of inherited diseases are caused by genomic variations in protein-coding regions; thus, exome sequencing unravels the causative variants in a considerable number of cases allowing data interpretation, less dependent of the initial clinical diagnosis, which is to the benefit of both the clinician and the patient. This unbiased candidate variant prioritization approach allows to differentially diagnose cases with uncertain clinical phenotypes due to overlapping features between diseases and their progressive nature with absence of full-grown symptoms in juvenile cases. Some examples to this common problem are reported here in non-ALS MND genes like *ZFYVE26*, *DNAJB2*, *PLEKHG5*, *TRPV4*, and *FBXO38*, in which early disease-onset or intrafamilial phenotypic heterogeneity, ranging from neuropathy to motor neuron disease, lead to uncertain clinical diagnoses (Figure 6).

Identification of new genetic players implicated in ALS and ALS-like disease opens new opportunities for understanding the converging mechanisms in neurodegeneration and/or motor neuron loss. Moreover, these also contribute to the development of more specific, even personalized, therapeutic targets like gene-specific antisense oligonucleotides. Thus, today, it is important to define the genetic causes of even yet untreatable diseases, to drive pharmaceutical/gene-editing research and to offer hope to patients and their families. Clinical exome sequencing, for which the diagnosis success rate increased exponentially in clinical settings, includes several beneficial outcomes: (a) treatment of patients with syndromic diseases like enzyme deficiencies, (b) using genetic information for reproductive genetic counseling and family planning and (c) recruitment of patients with specific genomic variants into clinical trials. Most importantly, WES shortens the diagnostic delay of at least 1 year in ALS, which may include several invasive and expensive procedures, and these should be taken into account while considering its cost-effectiveness

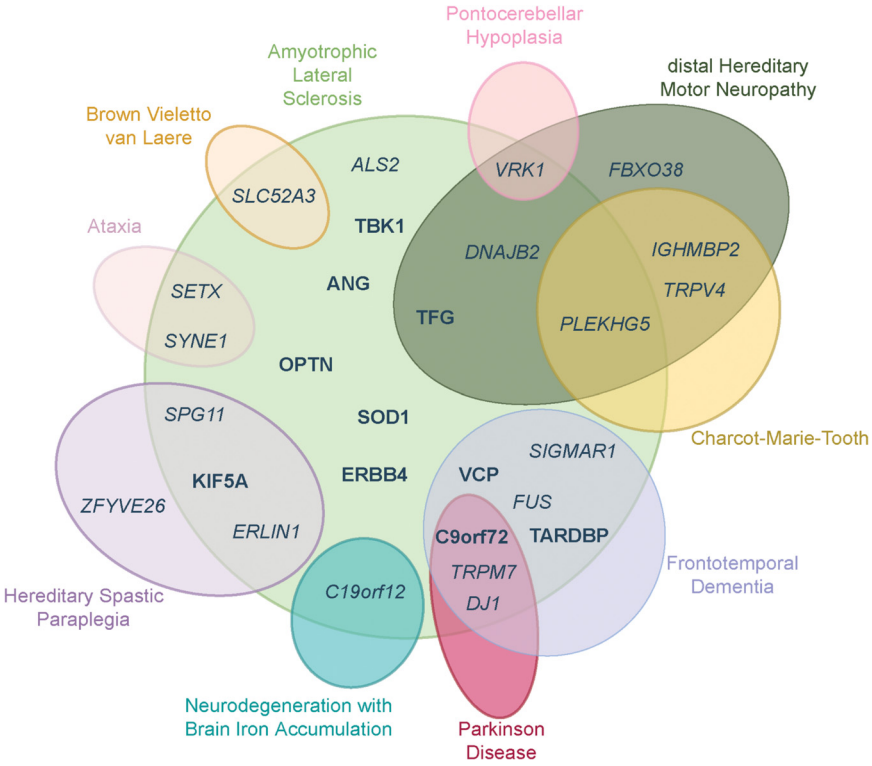

FIGURE 6 Genetic heterogeneity behind motor neruon diseases in our cohort. Individuals carrying the genomic variations represented in bold had clear/definitive initial diagnosis of classical ALS with an average age of onset of 46. Genes represented in italics are identified in patients presenting with juvenile motor neuron disease or nonclassical ALS-like disease with expanding phenotypes in the index or in affected family members (average AO: 13). This picture emphasizes pleiotropy in genes in addition to the expansion of phenotypes, calling for a gene-based disease classification. Venn diagram is drawn according to the disease-gene associations obtained from the literature (Akçimen et al., 2019; Al-Saif, Al-Mohanna, & Bohlega, 2011; Annesi et al., 2005; Brenner et al., 2018; Chen et al., 2004; Cirulli et al., 2015; Cottenie et al., 2014; Daoud et al., 2012; Deschauer et al., 2012; Frasquet, Va, & Sevilla, 2017; Greenway et al., 2006; Grohmann et al., 2001; Hermosura et al., 2005; Hughes et al., 2001; Ishiura et al., 2012; Iskender et al., 2015; Kimonis, Fulchiero, Vesa, & Watts, 2008; Maruyama et al., 2010; Maystadt et al., 2007; Nalini, Pandraud, Mok, & Houlden, 2013; H. P. Nguyen, Van Broeckhoven, & van der Zee, 2018; Stoll et al., 2016; Sumner et al., 2013; Synofzik et al., 2016; Takahashi et al., 2013; Tunca et al., 2018; Velilla et al., 2019; Yang et al., 2001). ALS, amyotrophic lateral sclerosis; AO, age of onset

(Fogel et al., 2014; Fogel, Satya-Murti, & Cohen, 2016; Trujillano et al., 2017).

6.4 | Project MinE to understand sporadic ALS: Impact of Turkish WGS data

Currently, the Turkish population is one of the highly represented cohorts in Project MinE in terms of sample size. Our results from screening for pathogenic exonic variants in the Turkish Project MinE cohort revealed marked incomplete penetrance for the three common ALS genes (*SOD1*, *TARDBP*, and *FUS*). In recent years, many studies showed that the frequency of ALS patients carrying more than one variant is higher than expected by chance, providing evidence that ALS may result from multiple rare variants with additive effects on disease development and presentation, for example, age of disease onset, progression and severity (van Blitterswijk et al., 2012). This "oligogenic model" of ALS may solve a portion of unexplained sporadic cases. The variations in *PON1* and *PON3* detected in our cohort might also act in such a manner (Table A5). The PON variants can lead to oligomerization of the native protein through N-terminal HDL particles, as previously reported, and further decrease its own hydrolytic activity (Josse et al., 2002). Disease pathology caused by paraoxonase genes, intensely studied for their role in ALS, may arise with reduced ability of PON enzymes, responsible of detoxifying organophosphates, which are neurotoxins associated with an increased ALS risk (Cronin, Greenway, Prehn, & Hardiman, 2007; Landers et al., 2008; Menini & Gugliucci, 2014; Merwin, Obis, Nunez, & Re, 2017; Ticozzi et al., 2010; Verde et al., 2019; Wills et al., 2009).

Genome-wide association study and gene-based burden testing for population-specific local signals with WGS data of 623 sporadic patients and 142 Turkish controls, a four times larger Turkish cohort than the one from van Rheenen et al. (2016), did not reveal any significant loci neither in variant-based GWAS, nor in gene-based disease burden analyses. Yet, the resulting network, which combined gene-based burden analysis with coexpression information, pointed to a significantly high enrichment (~15-fold) of cell cycle-related genes. Accordingly, changes in expression levels and subcellular localization of cell cycle proteins and their transcriptional regulators have been linked to neuronal death in the literature in terms of ALS and other neurodegenerative diseases (M. D. Nguyen et al., 2003; Ranganathan & Bowser, 2003).

Network analysis combined with literature text-mining suggested *DECR1*, *ATL1*, *HDAC2*, *GEMIN4*, and *HNRNPA3* as disease susceptibility candidates although these had no significant burden for ALS in the SKAT-O analysis. Previous studies show increased levels of the mitochondrion-related β-oxidation enzyme *DECR1* at disease onset in SOD1-G93A mice spinal cords (Q. Li et al., 2010; Pharaoh et al., 2019). *Atlastin-1* (*ATL1*), associated with upper motor neuron syndromes (De Bot et al., 2013), is a protein effective in structural and functional integrity of the endoplasmic reticulum (Muriel et al., 2009). *Histone deacetylase 2* (*HDAC2*) is important for the nervous system and was shown to be upregulated in ALS patients (Janssen et al., 2010). The increased HDAC activity in neurodegeneration and positive effects of HDAC inhibition, including HDAC2, on motor symptoms are reported in the literature (Lazo-Gómez, Ramírez-Jarquín, Tovar-y-Romo, & Tapia, 2013; Rossaert et al., 2019). Finally, GEMIN4 and HNRNPA3 are involved in RNA processing and interact closely with ALS-causative proteins in this machinery. GEMIN4 acts in the survival motor neuron complex formation, disrupted in lower motor neuron disease, and is in the FUS interactome together with HNRNPA3, which is detected in the spinal cords of *C9orf72*-positive ALS and FTD patients (Chi et al., 2018; Davidson et al., 2017; Fifita et al., 2017). Altogether, pathogenic variants in these genes, which are not yet detected in familial ALS, should be further investigated in larger cohorts for their possible contribution to sporadic ALS.

Although there is a wide variability in disease incidence and manifestation across populations and geographical regions, we acknowledge the shortcoming of our analyses to catch a significant population-specific signal considering the insufficient number of ALS patients and controls, which calls for far more Turkish samples to be sequenced. On the other hand, our WGS data with 900 samples and expanding, is expected to exemplify a unique population with a heterogeneous gene pool that will support to study the combinatorial effects of diverse SNPs in manifestation of sALS. In this respect, we are confident that this report will encourage local clinicians for recruitment of new patients. Only then, it will be possible to overcome power limitations, currently faced even by Project MinE with >9,000 DNA samples analyzed (Dekker et al., 2019; van Rheenen et al., 2018). The answer to the question of how to move forward at this point will be the collection of larger case and control cohorts in the framework of national and international collaborations and making all genomic data publicly available to increase the power of datasets. This will allow us to understand and interpret how a variant causes disease within the context of the larger population.

7 | CONCLUSION

Genetics offers a means to dissect the heterogeneity of ALS and to understand the cellular mechanisms resulting in motor neuron degeneration. The recent genetic findings driven by the NGS technology have not only expanded our knowledge of the wealth of genes giving rise to motor neuron degeneration, but also on the pleiotropic effects and extensive phenotypic spectra associated with specific ALS genes. Since the road from family pedigrees to clinical interpretation of variants is challenging, deep phenotyping of the patient, comprehensive analysis of the candidate variants with advanced bioinformatic tools and most importantly a tight researcher-clinician relationship are indispensable parts of the whole process. Ultimately, the discovery of all ALS genes will help to better define the multifaceted nature of ALS, which is accepted no more as a monolithic disease, but recognized as a spectrum of diseases converging into common clinical features. This allows a subclassification of patients into more precise clinical categories in which a common genetic cause is more likely to be identified.

Data presented in this study compiles the molecular analysis results of ALS patients at NDAL in Istanbul, Turkey for 18 years and is an expanded update and upscale of our 2015 results from 443 to 1,200 probands (Özoğuz et al., 2015). The excessive number of rare and novel variants reported here once more show the power of clinical exome sequencing. GWAS conducted with 800 WGS samples and coupled with a coexpression network analysis identifies disease-risk gene candidates. The results on Turkey presented here will hopefully contribute to the diversity of genetic and mechanistic factors underlying ALS, further driving research. Moreover, they are expected to shed light on the multilevel heterogeneity of the disease, as an important factor in a precision medicine approach towards the development of molecular therapies for stratified patient subgroups. Finally, the GeNDAL variant browser, a novel tool to observe Turkish population-specific allele frequencies, is expected to be a unique and valuable resource for disease gene identification studies for the neurogenetics community.

ACKNOWLEDGEMENTS

We would like to thank Güneş Birdal, Doruk Savaş, Betül Uysal, Aslı Gündoğdu Eken and Irmak Şahbaz for excellent technical assistance. We thank all other NDAL members for their friendship and their encouragement. The Turkish ALS-MND Association and all patients and families are acknowledged for their cooperation. We cordially thank Prof. Jan Veldink and the Project MinE Sequencing Consortium for the collegial collaboration and support. We extend our deepest gratitude to our academic advisors Prof. Robert H. Brown (UMass Medical School, MA, USA) and to Prof. Jeffrey D. Macklis (Harvard Medical School, MA, USA) for supporting NDAL's growth and development over the years. Last but not least, we would like to express our heartfelt thanks to Suna, İnan and İpek Kıraç, none of our efforts would be possible without their vision, devotion, dedicated mentorship and sustained support. This article is funded by Suna and İnanKıraç Foundation, Grant/Award Number: 2005-2020; Boğaziçi University Research Funds, Grant/Award Number: 15B01P1; TÜBiTAK, Grant/Award Number: 109S075. The authors gratefully acknowledge use of the services and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Presidency of Strategy and Budget. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Presidency of Strategy and Budget.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

AUTHOR CONTRIBUTIONS

C. T., A. N. B., A. E. Ç., and E. K. designed the study. C. T., T. Ş., F. A., C. C., E. B., R. P., S. Z., C. K., E. K., N. E. Ş., H. H., A. O. E., U. N., O. K., and G. O. performed the experiments. T. A., H. D., E. Ş., A. Ç., A. B. G., G. B.Y., B. İ., K. U., H. H., B. B., N. T., F. A., M. E., C. B., D. K., H. İ., A. S., N. U. A., M. A. A., F. K., E. T., P. O., F. D., and Y. P. performed neurological examination on the patients and provided detailed clinical information. C. T. and A. N. B. wrote the manuscript. F. A., A. E. Ç., E. K., Ö. T., and Y. P. were instrumental in the compilation of parts of the manuscript.

DATA AVAILABILITY STATEMENT

Data in this paper is published within the paper and deposited to ClinVar public database (https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB7287039). Additionally, ALS-related variants identified in the WGS data are available in GeNDAL variant browser (www.gendal.org).

ORCID

Ceren Tunca http://orcid.org/0000-0002-0657-6348
Tuncay Şeker http://orcid.org/0000-0002-4567-627X
Fulya Akçimen http://orcid.org/0000-0003-0931-5247
Cemile Koçoğlu http://orcid.org/0000-0003-2055-6607
Ece Kartal http://orcid.org/0000-0002-7720-455X
Hamid Hamzeiy http://orcid.org/0000-0001-7990-2530
A. Nazlı Başak http://orcid.org/0000-0001-6977-2517

REFERENCES

Akçimen, F., Vural, A., Durmuş, H., Çakar, A., Houlden, H., Parman, Y. G., & Nazlı Başak, A. (2019). A novel homozygous FBXO38 variant causes an early-onset distal hereditary motor neuronopathy type IID. *Journal of Human Genetics*, *64*, 1141–1144. https://doi.org/10.1038/s10038-019-0652-y

Al-Saif, A., Al-Mohanna, F., & Bohlega, S. (2011). A mutation in sigma-1 receptor causes juvenile amyotrophic lateral sclerosis. *Annals of Neurology*, *70*(6), 913–919. https://doi.org/10.1002/ana.22534

Andersen, P. (2006). Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. *Current Neurology and Neuroscience Reports*, *6*(1), 37–46.

Andersen, P., Nilsson, P., Ala-Hurula, V., Keränen, M. L., Tarvainen, I., Haltia, T., … Marklund, S. L. (1995). Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. *Nature Genetics*, *10*, 61–66. https://doi.org/10.1038/ng0595-61

Annesi, G., Savettieri, G., Pugliese, P., D'Amelio, M., Tarantino, P., Ragonese, P., … Quattrone, A. (2005). DJ-1 mutations and parkinsonism-dementia-amyotrophic lateral sclerosis complex. *Annals of Neurology*, *58*, 803–807. https://doi.org/10.1002/ana.20666

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., … Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*, 68–74. https://doi.org/10.1038/nature15393

Battistini, S., Benigni, M., Ricci, C., & Rossi, A. (2013). SOD1 mutations in amyotrophic lateral sclerosis. *European Neurological Journal*, *252*(7), 782–788. https://doi.org/10.1007/s00415-005-0742-y

Brenner, D., Yilmaz, R., Müller, K., Grehl, T., Petri, S., Meyer, T., … Kassubek, J. (2018). Hot-spot KIF5A mutations cause familial ALS. *Brain*, *141*(3), 688–697. https://doi.org/10.1093/brain/awx370

Brown, R. H., & Al-Chalabi, A. (2017). Amyotrophic lateral sclerosis. *New England Journal of Medicine*, *377*(2), 162–172. https://doi.org/10.1056/NEJMra1603471

Chen, Y.-Z., Bennett, C. L., Huynh, H. M., Blair, I. P., Puls, I., Irobi, J., … Chance, P. F. (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *The American Journal of Human Genetics*, *74*(6), 1128–1135. https://doi.org/10.1086/421054

Chi, B., O'Connell, J. D., Yamazaki, T., Gangopadhyay, J., Gygi, S. P., & Reed, R. (2018). Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and

contains multiple ALS/SMA-causative proteins. *Scientific Reports*, *8*(8755), 8755. https://doi.org/10.1038/s41598-018-27136-3

Çiçek, A. E. (2017). k-shell decomposition reveals structural properties of the gene coexpression network for neurodevelopment. *Turkish Journal of Biology*, *41*(2), 333–341. https://doi.org/10.3906/biy-1608-30

Cirulli, E. T., Lasseigne, B. N., Petrovski, S., Sapp, P. C., Dion, P. a, Leblond, C. S., … Goldstein, D. B. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science*, *347*(6229), 1436–1441.

Cottenie, E., Kochanski, A., Jordanova, A., Bansagi, B., Zimon, M., Horga, A., … Houlden, H. (2014). Truncating and missense mutations in IGHMBP2 cause Charcot-Marie tooth disease type 2. *American Journal of Human Genetics*, *95*, 590–601. https://doi.org/10.1016/j.ajhg.2014.10.002

Cronin, S., Greenway, M. J., Prehn, J. H. M., & Hardiman, O. (2007). Paraoxonase promoter and intronic variants modify risk of sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, *78*(9), 984–986. https://doi.org/10.1136/jnnp.2006.112581

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daoud, H., Zhou, S., Noreau, A., Sabbagh, M., Belzil, V., Dionne-laporte, A., … Rouleau, G. A. (2012). Exome sequencing reveals SPG11 mutations causing juvenile ALS. *NBA*, *33*(4), 839.e5–839.e9. https://doi.org/10.1016/j.neurobiolaging.2011.11.012

Davidson, Y. S., Flood, L., Robinson, A. C., Nihei, Y., Mori, K., Rollinson, S., … Mann, D. M. A. (2017). Heterogeneous ribonuclear protein A3 (hnRNP A3) is present in dipeptide repeat protein containing inclusions in Frontotemporal Lobar Degeneration and Motor Neurone disease associated with expansions in C9orf72 gene. *Acta Neuropathologica Communications*, *5*(31), 1–10. https://doi.org/10.1186/s40478-017-0437-5

De Bot, S. T., Veldink, J. H., Vermeer, S., Mensenkamp, A. R., Brugman, F., Scheffer, H., … Van De Warrenburg, B. P. (2013). ATL1 and REEP1 mutations in hereditary and sporadic upper motor neuron syndromes. *Journal of Neurology*, *260*(3), 869–875. https://doi.org/10.1007/s00415-012-6723-z

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., … Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215. https://doi.org/10.1038/nature13772

Dekker, A. M., Diekstra, F. P., Pulit, S. L., Tazelaar, G. H. P., van der Spek, R. A., van Rheenen, W., … Veldink, J. H. (2019). Exome array analysis of rare and low frequency variants in amyotrophic lateral sclerosis. *Scientific Reports*, *9*(1), 3–10. https://doi.org/10.1038/s41598-019-42091-3

Deschauer, M., Gaul, C., Behrmann, C., Prokisch, H., Zierz, S., & Haack, T. B. (2012). C19orf12 mutations in neurodegeneration with brain iron accumulation mimicking juvenile amyotrophic lateral sclerosis. *Journal of Neurology*, *259*(11), 2434–2439. https://doi.org/10.1007/s00415-012-6521-7

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., & Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics*, *24*, 2125–2137. https://doi.org/10.1093/hmg/ddu733

Elden, A. C., Kim, H. J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., … Gitler, A. D. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, *466*, 1069–1075. https://doi.org/10.1038/nature09320

EVS. (2014). Exome variant server. NHLBI GO Exome Sequencing Project (ESP).

Exome Aggregate Consortium. (2016). ExAC Browser.

Fifita, J. A., Zhang, K. Y., Galper, J., Williams, K. L., McCann, E. P., Hogan, A. L., … Blair, I. P. (2017). Genetic and pathological assessment of hnRNPA1, hnRNPA2/B1, and hnRNPA3 in familial and sporadic amyotrophic lateral sclerosis. *Neurodegenerative Diseases*, *17*(6), 304–312. https://doi.org/10.1159/000481258

Fogel, B. L., Lee, H., Deignan, J. L., Strom, S. P., Kantarci, S., Wang, X., … Nelson, S. F. (2014). Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurology*, *71*(10), 1–10. https://doi.org/10.1001/jamaneurol.2014.1944

Fogel, B. L., Satya-Murti, S., & Cohen, B. H. (2016). Clinical exome sequencing in neurologic disease: AAN model coverage policy. *Neurology: Clinical Practice*, *6*(2), 164–176. https://doi.org/10.1212/CPJ.0000000000000239

Frasquet, M., Va, J. F., & Sevilla, T. (2017). The role of DNAJB2 in amyotrophic lateral sclerosis. *Brain*, *139*(10), e57.

Ghasemi, M., & Brown, R. H. (2018). Genetics of amyotrophic lateral sclerosis. *Cold Spring Harbor Perspectives in Medicine*, *8*(5), 1–38.

Greenway, M. J., Andersen, P. M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., … Green, A. (2006). ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nature Genetics*, *38*(4), 2005–2007. https://doi.org/10.1038/ng1742

Grohmann, K., Schuelke, M., Diers, A., Hoffmann, K., Lucke, B., Adams, C., … Hübner, C. (2001). Mutations in the gene encoding immunoglobulin μ-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nature Genetics*, *29*, 75–77. https://doi.org/10.1038/ng703

Guerrero, E. N., Wang, H., Mitra, J., Hegde, P. M., Stowell, S. E., Liachko, N. F., … Hegde, M. L. (2016). TDP-43/FUS in motor neuron disease: Complexity and challenges. *Progress in Neurobiology*, *145–146*, 78–97. https://doi.org/10.1016/j.pneurobio.2016.09.004

Hamzeiy, H., Savaş, D., Tunca, C., Şen, N. E., Gündoğdu Eken, A., Şahbaz, I., … Başak, A. N. (2018). Elevated global DNA methylation is not exclusive to amyotrophic lateral sclerosis and is also observed in spinocerebellar ataxia types 1 and 2. *Neurodegenerative Diseases*, *18*(1), 38–48. https://doi.org/10.1159/000486201

Hardiman, O., Al-Chalabi, A., Brayne, C., Beghi, E., Van Den Berg, L. H., Chio, A., … Rooney, J. (2017). The changing picture of amyotrophic lateral sclerosis: Lessons from European registers. *Journal of Neurology, Neurosurgery and Psychiatry*, *88*(7), 557–563. https://doi.org/10.1136/jnnp-2016-314495

Hermosura, M. C., Nayakanti, H., Dorovkov, M. V., Calderon, F. R., Ryazanov, A. G., Haymer, D. S., & Garruto, R. M. (2005). A TRPM7 variant shows altered sensitivity to magnesium that may contribute to the pathogenesis of two Guamanian neurodegenerative disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 11510–11515. https://doi.org/10.1073/pnas.0505149102

Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., … Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, *8*(9), R183. https://doi.org/10.1186/gb-2007-8-9-r183

Hübers, A., Just, W., Rosenbohm, A., Müller, K., Marroquin, N., Goebel, I., … Volk, A. E. (2015). De novo FUS mutations are the most frequent genetic cause in early-onset German ALS patients. *Neurobiology of Aging*, *36*(11), 3117.e1–3117.e6. https://doi.org/10.1016/j.neurobiolaging.2015.08.005

Hughes, C. A., Byrne, P. C., Webb, S., McMonagle, P., Patterson, V., Hutchinson, M., & Parfrey, N. A. (2001). SPG15, a new locus for autosomal recessive complicated HSP on chromosome 14q. *Neurology*, *56*, 1230–1233. https://doi.org/10.1212/WNL.56.9.1230

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., … Sieh, W. (2016). REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *American*

Journal of Human Genetics, *99*(4), 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Human Genetics*, *92*(6), 841–853. https://doi.org/10.1016/j.ajhg.2013.04.015

Ishiura, H., Sako, W., Yoshida, M., Kawarai, T., Tanabe, O., Goto, J., … Tsuji, S. (2012). The TRK-fused gene is mutated in hereditary motor and sensory neuropathy with proximal dominant involvement. *American Journal of Human Genetics*, *91*(2), 320–329. https://doi.org/10.1016/j.ajhg.2012.07.014

Iskender, C., Kartal, E., Akcimen, F., Kocoglu, C., Kotan, D., Eraksoy, M., … Başak, A. N. (2015). Turkish families with juvenile motor neuron disease broaden the phenotypic spectrum of SPG11. *Neurology: Genetics*, *1*(3), e25. https://doi.org/10.1212/NXG.0000000000000025

Janssen, C., Schmalbach, S., Boeselt, S., Sarlette, A., Dengler, R., & Petri, S. (2010). Differential histone deacetylase mRNA expression patterns in amyotrophic lateral sclerosis. *Journal of Neuropathology and Experimental Neurology*, *96*(6), 573–581. https://doi.org/10.1097/NEN.0b013e3181ddd404

Jimeno-Yepes, A. J., Sticco, J. C., Mork, J. G., & Aronson, A. R. (2013). GeneRIF indexing: Sentence selection based on machine learning. *BMC Bioinformatics*, *14*(1), 171. https://doi.org/10.1186/1471-2105-14-171

Josse, D., Ebel, C., Stroebel, D., Fontaine, A., Borges, F., Echalier, A., … Masson, P. (2002). Oligomeric states of the detergent-solubilized human serum paraoxonase (PON1). *Journal of Biological Chemistry*, *277*, 33386–33397. https://doi.org/10.1074/jbc.M200108200

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., … MacArthur, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*, 531210. https://doi.org/10.1101/531210

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., … Zoing, M. C. (2011). Amyotrophic lateral sclerosis. *The Lancet*, *377*(9769), 942–955. https://doi.org/10.1016/S0140-6736(10)61156-7

Kimonis, V. E., Fulchiero, E., Vesa, J., & Watts, G. (2008). VCP disease associated with myopathy, Paget disease of bone and frontotemporal dementia: Review of a unique disorder. *Biochimica et Biophysica Acta*, *1782*(12), 744–748. https://doi.org/10.1016/j.bbadis.2008.09.003

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Aguilera, M. A., Meyer, R., & Massouras, A. (2019). VarSome: The human genomic variant search engine. *Bioinformatics*, *35*(11), 1978–1980. https://doi.org/10.1093/bioinformatics/bty897

Lahut, S., Ömür, Ö., Uyan, Ö., Ağim, Z. S., Özoğuz, A., Parman, Y., … Başak, A. N. (2012). ATXN2 and its neighbouring gene SH2B3 are associated with increased ALS risk in the Turkish population. *PLoS One*, *7*(8), e42956. https://doi.org/10.1371/journal.pone.0042956

Lamp, M., Origone, P., Geroldi, A., Verdiani, S., Gotta, F., Caponnetto, C., … Mandich, P. (2018). Twenty years of molecular analyses in amyotrophic lateral sclerosis: Genetic landscape of Italian patients. *Neurobiology of Aging*, *66*, 179.e5–179.e16. https://doi.org/10.1016/j.neurobiolaging.2018.01.013

Landers, J. E., Shi, L., Cho, T. J., Glass, J. D., Shaw, C. E., Leigh, P. N., … Brown, R. H. (2008). A common haplotype within the PON1 promoter region is associated with sporadic ALS. *Amyotrophic Lateral Sclerosis*, *9*(5), 306–314. https://doi.org/10.1080/17482960802233177

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*, D980–D985. https://doi.org/10.1093/nar/gkt1113

Lazo-Gómez, R., Ramírez-Jarquín, U. N., Tovar-y-Romo, L. B., & Tapia, R. (2013). Histone deacetylases and their role in motor neuron degeneration. *Frontiers in Cellular Neuroscience*, *7*(243), 1–7. https://doi.org/10.3389/fncel.2013.00243

Leblond, C. S., Webber, A., Gan-Or, Z., Moore, F., Dagher, A., Dion, P. A., & Rouleau, G. A. (2016). De novo *FUS* P525L mutation in Juvenile amyotrophic lateral sclerosis with dysphonia and diplopia. *Neurology: Genetics*, *2*(2), e63. https://doi.org/10.1212/NXG.0000000000000063

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Q., Velde, C., Vande, Israelson, A., Xie, J., Bailey, A. O., Dong, M. Q., … Miller, T. M. (2010). ALS-linked mutant superoxide dismutase 1 (SOD1) alters mitochondrial protein composition and decreases protein import. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(49), 21146–21151. https://doi.org/10.1073/pnas.1014862107

Liu, L., Lei, J., & Roeder, K. (2015). Network assisted analysis to reveal the genetic basis of autism. *Annals of Applied Statistics*, *9*(3), 1571–1600. https://doi.org/10.1214/15-AOAS844

Liu, L., Lei, J., Sanders, S. J., Willsey, A. J., Kou, Y., Cicek, A. E., … Roeder, K. (2014). DAWN: A framework to identify autism genes and subnetworks using gene expression and genetics. *Molecular Autism*, *5*, 1–22. https://doi.org/10.1186/2040-2392-5-22

Majounie, E., Renton, A. E., Mok, K., Dopper, E. G. P., Waite, A., Rollinson, S., … Traynor, B. J. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *The Lancet Neurology*, *11*(4), 323–330. https://doi.org/10.1016/S1474-4422(12)70043-1

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., … Kawakami, H. (2010). Mutations of optineurin in amyotrophic lateral sclerosis. *Nature*, *465*(7295), 223–226. https://doi.org/10.1038/nature08971

Maystadt, I., Rezsöhazy, R., Barkats, M., Duque, S., Vannuffel, P., Remacle, S., … Verellen-Dumoulin, C. (2007). The nuclear factor κB–activator gene PLEKHG5 is mutated in a form of autosomal recessive lower motor neuron disease with childhood onset. *The American Journal of Human Genetics*, *81*(1), 67–76. https://doi.org/10.1086/518900

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biology*, *17*(122), 1–14. https://doi.org/10.1186/s13059-016-0974-4

Menini, T., & Gugliucci, A. (2014). Paraoxonase 1 in neurological disorders. *Redox Report*, *19*(2), 49–58. https://doi.org/10.1179/1351000213Y.0000000071

Merwin, S. J., Obis, T., Nunez, Y., & Re, D. B. (2017). Organophosphate neurotoxicity to the voluntary motor system on the trail of environment-caused amyotrophic lateral sclerosis: The known, the misknown, and the unknown. *Archives of Toxicology*, *91*(8), 2939–2952. https://doi.org/10.1007/s00204-016-1926-1

Muriel, M. P., Dauphin, A., Namekawa, M., Gervais, A., Brice, A., & Ruberg, M. (2009). Atlastin-1, the dynamin-like GTPase responsible for spastic paraplegia SPG3A, remodels lipid membranes and may form tubules and vesicles in the endoplasmic reticulum. *Journal of Neurochemistry*, *110*(5), 1607–1616. https://doi.org/10.1111/j.1471-4159.2009.06258.x

Nalini, A., Pandraud, A., Mok, K., & Houlden, H. (2013). Madras motor neuron disease (MMND) is distinct from the riboflavin transporter genetic defects that cause Brown-Vialetto-Van Laere syndrome. *Journal of the Neurological Sciences*, *334*(1–2), 119–122. https://doi.org/10.1016/j.jns.2013.08.003

Nguyen, H. P., Van Broeckhoven, C., & van der Zee, J. (2018). ALS genes in the genomic era and their implications for FTD. *Trends in Genetics*, *34*(6), 404–423. https://doi.org/10.1016/j.tig.2018.03.001

Nguyen, M. D., Boudreau, M., Kriz, J., Couillard-Després, S., Kaplan, D. R., & Julien, J. P. (2003). Cell cycle regulators in the neuronal death pathway of amyotrophic lateral sclerosis caused by mutant superoxide dismutase 1. *Journal of Neuroscience*, *23*(6), 2131–2140.

Nicolas, A., Kenna, K. P., Renton, A. E., Ticozzi, N., Faghri, F., Chia, R., … Traynor, B. J. (2018). Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron*, *97*(6), 1268–1283.e6. https://doi.org/10.1016/j.neuron.2018.02.027

Norman, U., & Cicek, A. E. (2019). ST-Steiner: A spatio-temporal gene discovery algorithm. *Bioinformatics*, *35*(18), 3433–3440. https://doi.org/10.1093/bioinformatics/btz110

Ogaki, K., Li, Y., Atsuta, N., Tomiyama, H., Funayama, M., Watanabe, H., … Sobue, G. (2012). Analysis of C9orf72 repeat expansion in 563 Japanese patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*, *33*(10), 11–16. https://doi.org/10.1016/j.neurobiolaging.2012.05.011

Özoğuz, A., Uyan, Ö., Birdal, G., Iskender, C., Kartal, E., Lahut, S., … Başak, A. N. (2015). The distinct genetic pattern of ALS in Turkey and novel mutations. *Neurobiology of Aging*, *36*(4), 1764.e9–1764.e18. https://doi.org/10.1016/j.neurobiolaging.2014.12.032

Paré, B., Lehmann, M., Beaudin, M., Nordström, U., Saikali, S., Julien, J. P., … Gros-Louis, F. (2018). Misfolded SOD1 pathology in sporadic amyotrophic lateral sclerosis. *Scientific Reports*, *8*(14223), https://doi.org/10.1038/s41598-018-31773-z

Pharaoh, G., Sataranatarajan, K., Street, K., Hill, S., Gregston, J., Ahn, B., … Van Remmen, H. (2019). Metabolic and stress response changes precede disease onset in the spinal cord of mutant SOD1 ALS mice. *Frontiers in Neuroscience*, *13*(487), https://doi.org/10.3389/fnins.2019.00487

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., … Furlong, L. I. (2017). DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, *45*(D1), D833–D839. https://doi.org/10.1093/nar/gkw943

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Raczy, C., Petrovski, R., Saunders, C. T., Chorny, I., Kruglyak, S., Margulies, E. H., … Tanner, S. W. (2013). Isaac: Ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*, *29*(16), 2041–2043. https://doi.org/10.1093/bioinformatics/btt314

Ranganathan, S., & Bowser, R. (2003). Alterations in G1 to S phase cell-cycle regulators during amyotrophic lateral sclerosis. *American Journal of Pathology*, *162*(3), 823–835. https://doi.org/10.1016/S0002-9440(10)63879-5

Rossaert, E., Pollari, E., Jaspers, T., Van Helleputte, L., Jarpe, M., Van Damme, P., … Van Den Bosch, L. (2019). Restoration of histone acetylation ameliorates disease and metabolic abnormalities in a FUS mouse model. *Acta Neuropathologica Communications*, *7*(107), 1–19. https://doi.org/10.1186/s40478-019-0750-2

Saccon, R. A., Bunton-Stasyshyn, R. K. A., Fisher, E. M. C., & Fratta, P. (2013). Is SOD1 loss of function involved in amyotrophic lateral sclerosis? *Brain*, *136*(8), 2342–2358. https://doi.org/10.1093/brain/awt097

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., … State, M. W. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Smith, B. N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., … Shaw, C. E. (2013). The C9ORF72 expansion mutation is a common cause of ALS+/−FTD in Europe and has a single founder. *European Journal of Human Genetics*, *21*(1), 102–108. https://doi.org/10.1038/ejhg.2012.98

Smith, B. N., Topp, S. D., Fallini, C., Shibata, H., Chen, H. J., Troakes, C., … Shaw, C. E. (2017). Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. *Science Translational Medicine*, *9*(388), eaad9157. Retrieved from https://csic.gtbib.net/menu_usuario.php?centro=CSIC&p=l0dUZFhOMVlYSnBiMTl3WlhScFkybHZibVZ6TG5Cb2NBPT0=&campo=PETICION&texto=13709134&export=mendeley

Stoll, M., Teoh, H., Lee, J., Reddel, S., Zhu, Y., Buckley, M., … Nicholson, G. (2016). Novel motor phenotypes in patients with VRK1 mutations without pontocerebellar hypoplasia. *Neurology*, *87*(1), 65–70. https://doi.org/10.1212/WNL.0000000000002813

Sumner, C. J., D'Ydewalle, C., Wooley, J., Fawcett, K. A., Hernandez, D., Gardiner, A. R., … Houlden, H. (2013). A dominant mutation in FBXO38 causes distal spinal muscular atrophy with calf predominance. *American Journal of Human Genetics*, *93*(5), 976–983. https://doi.org/10.1016/j.ajhg.2013.10.006

Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., … Dang, C. (2013). Allen Brain Atlas: An integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Research*, *41*(Database issue), D996–D1008. https://doi.org/10.1093/nar/gks1042

Synofzik, M., Ronchi, D., Keskin, I., Basak, A. N., Wilhelm, C., Gobbi, C., … Andersen, P. M. (2012). Mutant superoxide dismutase-1 indistinguishable from wild-type causes ALS. *Human Molecular Genetics*, *21*(16), 3568–3574. https://doi.org/10.1093/hmg/dds188

Synofzik, M., Smets, K., Mallaret, M., Di Bella, D., Gallenmüller, C., Baets, J., … Bauer, P. (2016). SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: A large scale multi-centre study. *Brain*, *138*(5), 1378–1393. https://doi.org/10.1093/brain/aww079

Takahashi, Y., Fukuda, Y., Yoshimura, J., Toyoda, A., Kurppa, K., Moritoyo, H., … Tsuji, S. (2013). ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis Type 19. *American Journal of Human Genetics*, *93*(5), 900–905. https://doi.org/10.1016/j.ajhg.2013.09.008

Teer, J. K., Green, E. D., Mullikin, J. C., & Biesecker, L. G. (2012). VarSifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics*, *28*(4), 599–600. https://doi.org/10.1093/bioinformatics/btr711

Therrien, M., Dion, P. A., & Rouleau, G. A. (2016). ALS: Recent developments from genetics studies. *Current Neurology and Neuroscience Reports*, *16*(6), 59. https://doi.org/10.1007/s11910-016-0658-1

Ticozzi, N., LeClerc, A. L., Keagle, P. J., Glass, J. D., Wills, A. M., Van Blitterswijk, M., … Landers, J. E. (2010). Paraoxonase gene mutations in amyotrophic lateral sclerosis. *Annals of Neurology*, *68*(1), 102–107. https://doi.org/10.1002/ana.21993

Torkamani, A., Dean, B., Schork, N. J., & Thomas, E. A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Research*, *20*(4), 403–412. https://doi.org/10.1101/gr.101956.109

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E., Köster, J., Marais, A., … Abou Jamra, R. (2017). Clinical exome sequencing: Results from 2819 samples reflecting 1000 families. *European Journal of Human Genetics*, *25*(2), 176–182. https://doi.org/10.1038/ejhg.2016.146

Tunca, C., Akçimen, F., Coşkun, C., Gündoğdu-Eken, A., Kocoglu, C., Çevik, B., … Başak, A. N. (2018). ERLIN1 mutations cause teenage-onset slowly progressive ALS in a large Turkish pedigree. *European*

Journal of Human Genetics, *26*(5), 745–748. https://doi.org/10.1038/s41431-018-0107-5

van Blitterswijk, M., van Es, M., Hennekam, E. M., Dooijes, D., van Rheenen, W., Medic, J., … van den Berg, L. H. (2012). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human Molecular Genetics*, *21*(17), 3776–3784. https://doi.org/10.1093/hmg/dds199

van der Spek, R. A. A., van Rheenen, W., Pulit, S. L., Kenna, K. P., McLaughlin, R. L., Moisse, M., … Veldink, J. H. (2019). The Project MinE databrowser: Bringing large-scale whole-genome sequencing in ALS to researchers and the public. *BioRxiv*, *20*(5–6), 432–440. https://doi.org/10.1101/377911

van Es, M. A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R. J., Veldink, J. H., & van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *The Lancet*, *4*(390), 2084–2098. https://doi.org/10.1016/S0140-6736(17)31287-4

van Rheenen, W., Pulit, S. L., Dekker, A. M., Al Khleifat, A., Brands, W. J., Iacoangeli, A., … Project MinE ALS Sequencing Consortium (2018). Project MinE: Study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *European Journal of Human Genetics*, *26*(10), 1537–1546. https://doi.org/10.1038/s41431-018-0177-4

van Rheenen, W., Shatunov, A., Dekker, A. M., McLaughlin, R. L., Diekstra, F. P., Pulit, S. L., … Veldink, J. H. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nature Genetics*, *48*(9), 1043–1048. https://doi.org/10.1038/ng.3622

Velilla, J., Marchetti, M. M., Toth-Petroczy, A., Grosgogeat, C., Bennett, A. H., Carmichael, N., … Gupta, V. A. (2019). Homozygous TRPV4 mutation causes congenital distal spinal muscular atrophy and arthrogryposis. *Neurology: Genetics*, *5*(2), 312. https://doi.org/10.1212/NXG.0000000000000312

Verde, F., Tiloca, C., Morelli, C., Doretti, A., Poletti, B., Maderna, L., … Ticozzi, N. (2019). PON1 is a disease modifier gene in amyotrophic lateral sclerosis: Association of the Q192R polymorphism with bulbar onset and reduced survival. *Neurological Sciences*, *40*, 1469–1473. https://doi.org/10.1007/s10072-019-03834-2

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), 1–7. https://doi.org/10.1093/nar/gkq603

Wills, A. M., Cronin, S., Slowik, A., Kasperaviciute, D., Van Es, M. A., Morahan, J. M., … Brown, R. H. (2009). A large-scale international meta-analysis of paraoxonase gene polymorphisms in sporadic ALS. *Neurology*, *73*(1), 16–24. https://doi.org/10.1212/WNL.0b013e3181a18674

Winton, M. J., Van Deerlin, V. M., Kwong, L. K., Yuan, W., Wood, E. M., Yu, C., … Lee, M. (2009). A90V TDP-43 variant results in the aberrant localization of TDP-43 in vitro. *FEBS Letters*, *582*(15), 2252–2256.

Wobst, H. J., Wesolowski, S. S., Chadchankar, J., Delsing, L., Jacobsen, S., Mukherjee, J., … Moss, S. J. (2017). Cytoplasmic relocalization of TAR DNA-binding protein 43 is not sufficient to reproduce cellular pathologies associated with ALS in vitro. *Frontiers in Molecular Neuroscience*, *10*, 1–13. https://doi.org/10.3389/fnmol.2017.00046

Yang, Y., Hentati, A., Deng, H. X., Dabbagh, O., Sasaki, T., Hirano, M., … Siddique, T. (2001). The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nature Genetics*, *29*(2), 160–165. https://doi.org/10.1038/ng1001-160

How to cite this article: Tunca C, Şeker T, Akçimen F, et al. Revisiting the complex architecture of ALS in Turkey: Expanding genotypes, shared phenotypes, molecular networks, and a public variant database. *Human Mutation*. 2020;41:e7–e45. https://doi.org/10.1002/humu.24055

APPENDIX

Figure A1. Table A1–A10.

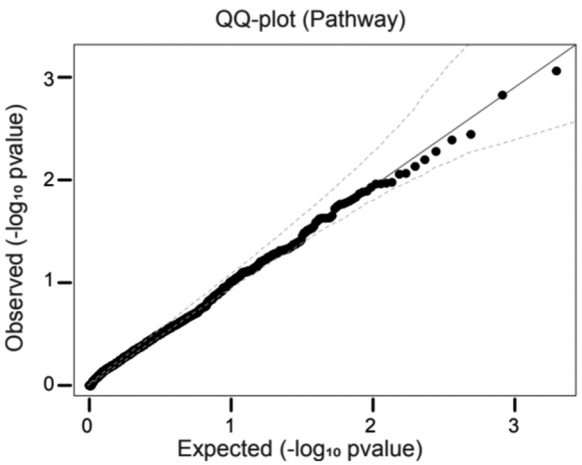

FIGURE A1 The quantile-quantile plot of pathway based SKAT-O *p* values. Each dot represents a pathway entry in gene set enrichment analysis

TABLE A1 Clinical data of cases investigated with WES

Family #	Gender	Age of onset	Consanguinity	Family history	Inheritance	Initial diagnosis
Fam1	M	50	No	Yes	AD	ALS
	M	45	No			
Fam2	M	20	No	No	Inconclusive	Juvenile ALS
Fam3	M	17	No	No	Inconclusive	Juvenile ALS
Fam4	M	23	Yes	Yes	AR	Early onset ALS
Fam5	M	14	Yes	Yes	AR	Juvenile ALS
Fam6	F	17	Yes	Yes	AR	Juvenile ALS
Fam7	F	15	Yes	Yes	AR	Juvenile ALS
	M	49	Yes			
	M	32	Yes			
Fam8	F	17	Yes	No	AR	Juvenile MND
Fam9	M	22	Yes	No	AR	Juvenile ALS
Fam10	M	17	Yes	No	AR	Juvenile ALS
Fam11	F	37	Yes	No	AR	ALS
Fam12	F	22	Yes	Yes	AR	Juvenile ALS
Fam13	F	47	No	Yes	AD	ALS/sensory neuropathy
Fam14	M	21	No	No	Inconclusive	MND
Fam15	M	40	Yes	No	AR	ALS
Fam16	F	45	Yes	No	AR	ALS
Fam17	M	27	Yes	No	AR	ALS
Fam18	M	25	Yes	No	AR	ALS
Fam19	F	37	Yes	No	AR	ALS
Fam20	M	14	Yes	No	AR	Juvenile MND
Fam21	M	17	Yes	No	AR	Juvenile MND
Fam22	M	61	Yes	No	AR	ALS
Fam23	M	60	Yes	Yes	AR	ALS
Fam24	F	49	No	Yes	AD	ALS
Fam25	M	44	Yes	Yes	AR	ALS
Fam26	F	17	Yes	No	AR	Juvenile ALS/disferlinopathy
Fam27	M	60	No	Yes	AD	ALS
Fam28	F	52	Yes	Yes	Inconclusive	ALS
	F	32	Yes		AR	
Fam29	M	23	Yes	No	AR	MND
Fam30	M	26	Yes	No	AR	ALS
Fam31	F	31	Yes	No	AR	MND
Fam32	M	9	Yes	No	AR	Juvenile ALS
Fam33	F	10	Yes	No	AR	Juvenile ALS
Fam34	M	21	Yes	Yes	AR	Juvenile ALS
Fam35	F	1	Yes	Yes	AR	Juvenile MND
	M	41	Yes			ALS
Fam36	F	20	Yes	Yes	AR	Juvenile ALS

(Continues)

TABLE A1 (Continued)

Family #	Gender	Age of onset	Consanguinity	Family history	Inheritance	Initial diagnosis
	M	13	Yes			SBMA
	F	20	Yes			Juvenile ALS
Fam37	F	25	Yes	No	AR	MND
Fam38	M	17	Yes	No	AR	Juvenile MND
Fam39	M	20	Yes	No	AR	Juvenile MND
Fam40	M	2	Yes	No	AR	Juvenile ALS
Fam41	F	Childhood	No	Yes	AD	CMT
	F		No			Juvenile MND
Fam42	M	52	No	No	Inconclusive	ALS
Fam43	F	60	Yes	Yes	Inconclusive	ALS/FTD
	F	60	Yes			
Fam44	F	48	No	Yes	AD	ALS
	F	48	No			
	M	47	No			
Fam45	F	16	No	Inconclusive	Inconclusive	ALS/Madras MND
Fam46	M	17	Yes	No	Inconclusive	ALS
Fam47	F	10	Yes	Yes	AR	Juvenile MND
	F		Yes			
Fam48	M	20	Yes	No	AR	MND
Fam49	M	35	Yes	No	AR	MND
Fam50	M	25	Yes	No	AR	Early onset ALS
Fam51	F	~3 months	Yes	Yes	AR	Juvenile MND
	F		Yes			
Fam52	M	25	No	No	Inconclusive	MND
Fam53	F	57	Yes	Yes	AR	ALS
	M	44	Yes			
Fam54	M	29	Yes	No	AR	MND
Fam55	F	58	No	Yes	AD	ALS
Fam56	F	76	No	Inconclusive	Inconclusive	ALS
Fam57	M	51	No	Yes	AD	ALS
	F	NA	No			
Fam58	F	40	No	Yes	AR	ALS
	M	45	No			
	F	NA	No			
Fam59	M	46	No	Yes	Inconclusive	ALS
Fam60	M	40	No	Yes	AD	ALS
	F	67	No			
Fam61	M	52	No	Yes	AD	ALS
Fam62	M	46	No	Yes	AD	ALS
Fam63	M	65	No	Yes	Inconclusive	ALS
Fam64	M	41	Yes	No	AR	ALS
Fam65	M	24	Yes	No	AR	Early onset ALS

TABLE A1 (Continued)

Family #	Gender	Age of onset	Consanguinity	Family history	Inheritance	Initial diagnosis
Fam66	M	6	Yes	Yes	AR	Juvenile MND
Fam67	M	14	Yes	No	AR	MND
Fam68	F	22	Yes	No	AR	Juvenile MND
Fam69	M	22	Yes	Yes	AR	ALS
Fam70	F	20	Yes	No	AR	MND
Fam71	F	Congenital	Yes	No	AR	MND
Fam72	M	14	Yes	Yes	AR	Juvenile ALS
Fam73	F	16	Yes	Yes	AR	ALS
Fam74	F	8	Yes	Yes	AR	Juvenile MND
	F	8	Yes			
	F	9	Yes			
Fam75	F	54	No	Yes	AD	ALS
	F	58	No			
	F	54	No			
Fam76	F	58	No	Yes	Inconclusive	ALS
Fam77	F	16	No	No	Inconclusive	ALS/Madras MND
Fam78	M	12	Yes	Yes	AR	Juvenile ALS
Fam79	M	39	No	Yes	Inconclusive	MND
	F	24	No			
Fam80	M	66	No	No	Inconclusive	ALS
Fam81	M	29	Yes	No	AR	ALS
Fam82	F	45	No	Yes	AD	ALS
Fam83	M	14	No	No	Inconclusive	Juvenile ALS
Fam84	F	32	No	Yes	AD	ALS
Fam85	M	67	No	Yes	AD	ALS
	F	42	No			
Fam86	M	58	No	Yes	AD	ALS
	F	59	No			
Fam87	F	46	No	Yes	AD	ALS
Fam88	M	44	Yes	Yes	AR	ALS
Fam89	F	37	Yes	No	AR	ALS
Fam90	M	57	Yes	No	AR	ALS
Fam91	M	24	Yes	No	AR	ALS
Fam92	M	51	No	Yes	AD	ALS
	F	NA	No			
Fam93	F	11	No	Yes	AD	Juvenile ALS
Fam94	F	38	Yes	No	AR	ALS
Fam95	M	31	No	No	AR	MND
Fam96	M	27	No	No	Inconclusive	ALS
Fam97	M	20	No	Yes	AR	MND
Fam98	F	52	Yes	No	AR	ALS

(Continues)

TABLE A1 (Continued)

Family #	Gender	Age of onset	Consanguinity	Family history	Inheritance	Initial diagnosis
Fam99	M F	61 65	No No	Yes	AR	ALS
Fam100	M	28	No	No	Inconclusive	MND
Fam101	F	31	Yes	Yes	AD	ALS
Fam102	F	54	No	Yes	AD	ALS
Fam103	F	49	Yes	No	AR	ALS
Fam104	M	43	Yes	No	AR	ALS
Fam105	F	29	No	No	Inconclusive	ALS
Fam106	M	45	Yes	No	AR	ALS
Fam107	F	1	Yes	No	AR	Juvenile ALS
Fam108	M M	Teenage Teenage	No No	Yes	Inconclusive	Juvenile ALS
Fam109	M M	33 36	Yes Yes	Yes	AR	ALS
Fam110	M	21	Yes	Yes	AR	Juvenile ALS
Fam111	F	58	No	Yes	AD	ALS
Fam112	F	32	Yes	Yes	AR	ALS
Fam113	F	57	Yes	No	AR	ALS
Fam114	M	19	Yes	Yes	Inconclusive	ALS
Fam115	F	58	No	Yes	AD	ALS
Fam116	M F	42 43	Yes Yes	Yes	AR	ALS
Fam117	M	26	No	Yes	AD	ALS
Fam118	F	9	Yes	No	AR	ALS
Fam119	M M M	37 NA 42	No No No	Yes	AD	ALS
Fam120	F	21	No	No	Inconclusive	ALS
Fam121	F	38	Yes	No	AR	ALS
Fam122	M	58	Yes	No	AR	ALS
Fam123	M	1	Yes	No	AR	Juvenile ALS
Fam124	M	1	Yes	Yes	AR	Juvenile ALS
Fam125	F	2,5	Yes	No	AR	Juvenile ALS
Fam126	M	16	Yes	No	AR	ALS
Fam127	M	24	Yes	Yes	AR	ALS-Parkinson-Dementia Complex

Abbreviations: ALS, amyotrophic lateral sclerosis; MND, motor neuron disease.

TABLE A2 Gene-based scores from GeneRIF and DisGeNet databases

	Gene	Count
GeneRIF Scores	*SOD1*	216
	TARDBP	106
	C9ORF72	91
	FUS	66
	ATXN2	19
	OPTN	18
	PFN1, VAPB	17
	ANG, TBK1	15
	VCP	13
	VEGFA, UBQLN2	11
	IGFALS, PON1	10
	HFE, NEFH, CHCHD10	8
	SQSTM1	7
	APOE, CHGB, GRN, NEFL, MATR3, UNC13A, ZNF512B	6
	DPP6, SIGMAR1, SETX, FGGY	5
	APP, DCTN1, IGF1, MMP9, MTHFR, KIFAP3, CHMP2B, ALS2	4
	CCNF, CST3, HNRNPA1, MAPT, SLC1A2, TIA1, TUBA4A, FIG4, TREM2, RTN4, LOC643387	3
	AR, CASP3, CYBB, DAO, ELAVL1, ERBB4, EWSR1, GPX3, GRIA2, GSK3B, IGFBP3, ITPR2, LGALS3, NFE2L2, NOS1, P2RX7, P4HB, PON2, PRPH, CCL2, SMN1, SMN2, TLR4, TNF, XRCC1, TAF15, RAB29, GPNMB, PPARGC1A, PARK7, SS18L1, KCNIP1, VPS54, SUSD2, CAMK1G, ARHGEF28, BTBD10, LRRK2, NIPA1, RBM45, MIR206	2
	ADARB1, ADCYAP1, ADCYAP1R1, ADH5, ADORA2A, PARP1, AGER, ANXA11, APEX1, ATP5F1D, BDNF, BRCA1, BSG, C3, CHGA, CHI3L1, CHI3L2, CHRNA4, TBCB, CNTF, CSF2, CSPG4, CTSD, CX3CR1, CYP27A1, ACE, DDIT3, EPHA3, EPO, FXN, FTL, GABRA1, GJA1, GLE1, GRIA1, GSK3A, GSTP1, HTT, HDAC2, NRG1, HIF1A, HK1, HLA-DRA, HLA-DRB5, HLA-F, HMGB1, HNMT, HNRNPA2B1, HNRNPH1, HSF1, HTR2B, IDI1, IFNG, IGF1R, IGF2, IGFBP2, IGHMBP2, IL6, IL10, IL18, ITIH4, KARS, KCNA2, KDR, KIF5C, KIFC3, KIR3DL2, KRT18, LAMC1, LOX, SMAD2, SMAD3, MAP1B, MBL2, MEF2C, MEF2D, MET, MMP2, MYO6, MYOD1, NAIP, NEK1, NGFR, SLC11A2, OGG1, PAX7, PGR, ABCB1, SERPINA1, PIN1, PLA2G4A, PON3, POU5F1, PPIA, MAP2K5, PSEN1, PTGS2, PURA, RAB1A, RAB5A, RET, RRAD, SLC16A2, SNCA, SNRPC, SOX2, SOX5, STAT3, TERT, TFAM, TIAM1, TLR2, TNFRSF1A, TNFRSF1B, TNNT2, TP53, TXNRD1, VDAC1, VDR, VSNL1, SLC30A3, TNFSF14, AIFM1, VAPA, ABCG2, KEAP1, CCS, HDAC6, SH2B3, TNIP1, YWHAQ, KIF3A, DCTN3, SIRT2, KIF1B, FAM120A, SIRT3, SIRT1, CABIN1, HSPB8, PABPC1, COQ2, STK39, HTRA2, UBQLN1, CHCHD2, IL23A, BCL11A, EQTN, OXR1, ELP3, RHOT1, MFN1, SLC30A6, PDGFC, UBQLN4, GJD2, SORCS2, PINK1, FTO, HDAC11, IFT74, SPG11, POLDIP3, TTBK1, IDI2, NLRP3, FBXO32, GRIN3B, TTBK2, NEWENTRY, HNRNPA3, ALS2CL, MIR125B1, MIR125B2, MIR141, MIR200A, MIR338, MIR424, CCR2, WASHC1, MIR1825	1
DisGeNet Scores	*SOD1*	1,291
	TARDBP	396
	C9ORF72	282
	FUS	199
	IGFALS	55
	VCP	54
	ATXN2	46
	OPTN	41
	ALS2	36
	ANG	35
	VEGFA, UBQLN2	32
	SLC1A2	29
	RBMS3, VAPB, PTBP1, SRRM2	28
	PNO1	27
	PFN1, SQSTM1	23
	SETX	21
	GRN	18
	PRPH	17
	GDNF, NEFH	16
	SMN2, SMN1	14

(Continues)

TABLE A2 (Continued)

Gene	Count
NEFL, *MATR3*, *CSF2*, *TBK1*	13
PON1, *UNC13A*, *SIGMAR1*, *DCTN1*, *LAMC2*	12
NUP62, *DCTN4*, *MAPT*, *GFAP*, *KHDRBS1*, *STMN1*, *GTF2H1*, *HFE*	11
BCL2, *SNRPN*, *SNURF*, *GRIA2*	10
NFE2L2, *MAK16*, *APOE*	9
TREM2, *SPG11*, *CASP3*, *CHCHD10*, *RIMS2*, *HNRNPA1*, *GABPA*, *CNTF*	8
DPP6, *GRM2*, *PTGS2*, *PPARGC1A*, *AR*, *IGF1*, *TNF*, *APP*, *CHMP2B*	7
TP53, *SOD2*, *CCL2*, *TAF15*, *CST3*	6
NUP98, *ADARB1*, *C3*	5
ROPN1L, *ASPM*, *APEX1*, *RNASE4*, *A1CF*, *P2RX7*, *ASPA*, *CAT*, *GPNMB*, *CNR2*, *ATG5*, *TUBA4A*, *CDK5*, *RAN*, *MOB3B*, *ASIP*, *EPHA4*, *KIFAP3*, *EPO*, *CASP1*, *NTF3*, *MST1*, *XBP1*, *SIRT1*, *ADA*	4
ZNF569, *COX2*, *CYP27A1*, *TMEM189-UBE2V1*, *DECR1*, *DES*, *S100A6*, *GRM5*, *BDNF*, *FIG4*, *TNFRSF1B*, *LIF*, *ELP3*, *GARS*, *TRPM7*, *NOS2*, *NOS1*, *NFKB1*, *PPARG*, *TMEM106B*, *TMPRSS13*, *RREB1*, *FMR1*, *ADAR*, *ZNF436*, *MMP9*, *THY1*, *XBP1P1*, *TLR4*, *PARP1*, *XIAP*, *ALAD*, *MSMB*, *HSPA4*, *P2RX4*, *MIR206*, *ZNF253*, *HSF1*, *BCL2L1,TMEM189*, *CCS*, *ZNF763*, *LMLN*, *CHGB*, *TNPO1*, *NEK1*, *UBE2V1*, *ZNF629*, *CYBB*	3
CAST, *PLA2G4A*, *PRNP*, *RBM8A*, *DNMT1*, *DAO*, *CUX1*, *CTSD*, *PTPA*, *RTN4*, *VDR*, *PIN1*, *VGF*, *PDC*, *VPS54*, *CREBBP*, *NELFE*, *FMO1*, *KIF1B*, *P2RX2*, *ATF6*, *HDAC6*, *OXR1*, *CTF1*, *P2RY1*, *TANK*, *RNF19A*, *SARM1*, *SS18L1*, *SUMO1*, *PHGDH*, *P2RY2*, *P2RX1*, *PRKN*, *P2RX3*, *C5AR1*, *P2RX5*, *TXN*, *MAPK8*, *TRPM2*, *OSBP*, *NIPA1*, *KCNA3*, *KHSRP*, *AKT1*, *POMC*, *AKT3*, *CS*, *EPHA3*, *FGF2*, *CASP12*, *PSEN1*, *TIAM1*, *IL2RB*, *AQP4*, *SRR*, *TUSC2*, *NRG1*, *GRIK1*, *SUGP1*, *KCNJ10*, *SIGLEC7*, *NGF*, *TIA1*, *APRT*, *TGM2*, *TFAM*, *NEFM*, *NAIP*, *HNRNPA2B1*, *CCNF*, *GRIA3*, *GRIA1*, *LSM2*, *GPX3*, *LGALS1*, *HNRNPH2*, *LGALS3*, *ATXN3*, *BSCL2*, *INA*, *MOBP*, *PSIP1*, *UBQLN1*, *HCLS1*, *HTT*, *PIGL*, *P2RX6*, *HNRNPA1P10*, *KEAP1*, *CHGA*, *ATXN2-AS*, *HMOX1*, *RAG2*, *HNRNPH1*, *RNASE1*, *SLC3A2*, *HSPB1*, *PINK1*, *RRAD*, *IFNG*	2
RBM45, *COX8A*, *PPARGC1B*, *SLCO6A1*, *MAP3K8*, *CNR1*, *TPPP*, *COL4A2*, *DBX1*, *LRRK2*, *CRP*, *CP*, *SH2B3*, *CRYAB*, *CRYM*, *MAPK14*, *CSF3*, *TTBK2*, *RBFOX3*, *LGALS16*, *CTNND2*, *CPNE4*, *RIPK3*, *YWHAQ*, *PNPLA6*, *CLU*, *SERPINA3*, *MECP2*, *ARAP2*, *MYBPH*, *MAP1B*, *MB*, *MDH1*, *CHAF1A*, *MAP3K5*, *MET*, *KITLG*, *MMP14*, *MPZ*, *MIR338*, *MT1A*, *MYH6*, *GRIN3B*, *MYO6*, *MAOB*, *DDX19A*, *ANKRD1*, *LAT*, *EGLN3*, *SLC52A3*, *NLRP3*, *CARD16*, *FBXO32*, *ADCYAP1*, *MIR1825*, *SPAG11B*, *MIR4299*, *GJB1*, *DDIT3*, *SIRT3*, *GRIP1*, *ZFYVE26*, *CABIN1*, *FOS*, *FPR2*, *SMUG1*, *DDX58*, *SLC7A11*, *ALOX5*, *MTOR*, *ABL1*, *ALS3*, *GABRA1*, *SIRT5*, *SUN3*, *RAB3GAP2*, *ALS2CL*, *GAP43*, *GAPDH*, *TIPARP*, *OSBPL3*, *GART*, *SPAG8*, *CLVS1*, *CLEC10A*, *OCLN*, *CDKN2A*, *AFF1-AS1*, *ERVK-2*, *MIR4649*, *ERVK-12*, *ERVK-22*, *ERVK-11*, *LINC00351*, *PPIF*, *CDK2*, *ALYREF*, *OLIG2*, *KCNMB2*, *NES*, *CHL1*, *MASP2*, *SIX2*, *OLFM4*, *ADARB2*, *CX3CR1*, *KCNMB2-AS1*, *CFDP1*, *KDR*, *SEMA3A*, *CDKN2D*, *TNIP1*, *MARCKS*, *XRN1*, *MIR23A*, *JAG2*, *JUND*, *IGF1R*, *HLA-DRB1*, *HLA-B*, *ANXA11*, *TNFRSF21*, *AMPH*, *HPGDS*, *GLE1*, *EIF3K*, *UBE2S*, *GLG1*, *HTRA2*, *ADGRD1*, *KRT18P55*, *SLC9A9*, *GPX1*, *ANK3*, *GRIA4*, *ANPEP*, *GRIN2A*, *GRM1*, *GRM3*, *GSR*, *GSTP1*, *JAK3*, *ITPR2*, *MIR155*, *ITGA9*, *HSPB8*, *HNRNPDL*, *OTOG*, *HSPB2*, *HSPA8*, *DNAJB2*, *HRES1*, *PRMT1*, *IDE*, *HNRNPK*, *HNRNPC*, *APAF1*, *HMOX2*, *HES1*, *TGM6*, *IGF2R*, *HLA-F*, *IGFBP7*, *FAS*, *IL1B*, *IL4*, *IL10*, *IL13*, *IL17A*, *CXCL10*, *UBN1*, *SLC40A1*, *HDAC2*, *HEXA*, *PFN2*, *PKLR*, *PLCD1*, *ATP7A*, *TMED9*, *PPIA*, *PPID*, *FGGY*, *RHOT1*, *NGFR*, *NFIL3*, *NEUROG1*, *NCAM1*, *LAMA3*, *LAMC1*, *LCN2*, *LDLR*, *LGALS4*, *LIG4*, *FADS3*, *LMNB1*, *LTBR*, *LUM*, *LY6E*, *MIR146A*, *PENK*, *SERPINF1*, *IL23A*, *NTF4*, *HGF*, *PIK3R4*, *HIF1A*, *UBE2K*, *GSTK1*, *MNX1*, *KCNK2*, *KIF3C*, *STK36*, *NOTCH1*, *NRF1*, *OPCML*, *PDGFA*, *P4HB*, *GEMIN4*, *PAWR*, *PCBP1*, *FOXP3*, *ATL1*, *APH1A*, *DBR1*, *PCSK2*, *SLC25A37*, *PRRX2*, *ANKRD2*, *DCTN5*, *GBX2*, *CDK5R1*, *PROM1*, *HSP90B2P*, *TRN-GTT2–1*, *TLE3*, *TTR*, *TNFRSF10B*, *CCR2*, *TXNRD1*, *RAB11A*, *UCP2*, *UCP3*, *UGCG*, *TNFSF14*, *RIPK1*, *VDAC1*, *ELP1*, *MAP4K3*, *VIM*, *INPP4B*, *GTPBP1*, *POLDIP3*, *DDX46*, *RARA*, *RASGRF1*, *ABCG1*, *TP73*, *ARHGEF7*, *TFPI2*, *TNS1*, *MSTN*, *IFT74*, *BRD3*, *TTBK1*, *CASR*, *PABPN1*, *CAMK4*, *DHX16*, *TRIM8*, *CAPN1*, *CDC7*, *ARHGAP24*, *BTBD10*, *C19ORF12*, *CASP9*, *XRCC1*, *WNT7A*, *CLIP2*, *VSNL1*, *RAB29*, *TLR2*, *TMSB4X*, *CLDN5*, *BCL10*, *TNFRSF1A*, *REG1A*, *PRMT8*, *RELA*, *RANGAP1*, *SUSD2*, *MFN2*, *SORT1*, *PKN1*, *RXRA,TDP1*, *SLC12A5*, *PTPRN*, *MAPK1*, *MAPK3*, *CLIP1*, *CD40LG*, *RRM2*, *PRDX6*, *MAPK9*, *MAP2K5*, *EIF2AK2*, *S100A8*, *S100A9*, *KIAA0513*, *ZNF704*, *ATXN1*, *PTPRZ1*, *BCYRN1*, *CCDC88A*, *GDA*, *CAMK1G*, *NUP153*, *CD59*, *UPF1*, *WASHC5*, *PTS*, *MIR524*, *CD68*, *HDAC4*, *TRIM27*, *RET*, *CTR9*, *ZNF512B*, *BAX*, *RAB5A*, *PVALB*, *KIAA0040*, *RAB1A*, *PTEN*, *BAG1*, *SLC17A6*, *PTGER2*, *PSMD2*, *RALGDS*, *LXN*, *BBS2*, *JPH3*, *HDAC11*, *ATAT1*, *CALB2*, *SIRT2*, *CBLL2*, *AAK1*, *FCGR3B*, *FCGR3A*, *PTK2B*, *F9*, *EWSR1*, *ETS2*, *FGF6*, *PRF1*, *A1BG*, *SLC30A6*, *SNAP25*, *SNCA*, *VAPA*, *SLC33A1*, *SOS1*, *SPAST*, *BRCA1*, *SST*, *STAT5A*, *SULT1E1*, *UNC119*, *BSG*, *CALB1*, *FNDC3A*, *SCFD1*, *TAC1*, *CAMTA1*, *GDI1*, *PABPC1*, *SUN2*, *ICE1*, *PSD3*, *BICD2*, *GSX2*, *DHFR*, *DNM2*, *DNMT3A*, *DPYD*, *EDN1*, *EGF*, *CELSR3*, *EGFR*, *EIF4G2*, *ELAVL1*, *ENG*, *EPHA1*, *EPHB2*, *AK4*, *ERBB4*, *EREG*, *MGRN1*, *FMO3*, *SYT1*, *SYP*, *HSPB3*, *SI*, *TACR1*, *ST3GAL3*, *BCL11B*, *SLC1A1*, *NRXN1*, *WNK1*, *CD163*, *SLC6A1*, *SLC6A3*, *SPAG11A*, *SLC12A2*, *CD14*, *SLC12A4*, *MSC*, *CHST2*, *SCD*, *TIMP3*, *MGAM*, *GEMIN2*, *PHF5A*, *CFAP410*, *SLC30A3*, *CACNA1S*, *SCG2*, *DERL1*, *TIMP2*, *MTHFSD*, *GORASP1*, *SUMO2*, *TAF12*, *TAT*, *MIR582*, *PRDX2*, *MFN1*, *TEAD1*, *TGFB1*, *TGFB2*, *TSPO*, *SHC1*, *SNAI1*, *TIMP1*, *SUMO3*, *ARHGEF28*, *SGK1*, *SIL1*, *CD7*, *COP1*, *LOC643387*, *SFPQ*, *NPEPPS*, *SELPLG*, *PTGES*, *SCN8A*	1

TABLE A3 Clinical data of patients with C9orf72 repeat expansion

	Total ALS	fALS	sALS
#			
Probands	80	42	38
Family members	8 + 7 assymptomatic	8	–
Male	45	23	22
Female	35	19	16
Male:female	1.3	1.2	1.4
AO			
Juvenile (<25 yrs)	1	1	–
Middle (25–45 yrs)	9	5	4
Late (>45 yrs)	66	36	32
Range (years)	32–80	32–80	40–71
MAO			
Total ±*SD*	54.5 ± 9.9	54.0 ± 11.0	55.0 ± 7.8
Male ±*SD*	52.7 ± 11.2	51.14 ± 13.3	54.4 ± 8.1
Female ±*SD*	56.7 ± 7.5	57.3 ± 7.5	56.0 ± 7.3
	9	5	4
Dementia SO			
Limb	50	26	24
Bulbar	18	12	6
Limb + bulbar	5	2	3
Not available	5	1	4

Abbreviations: #, numbers; ALS, amyotrophic lateral sclerosis; AO, age of onset; fALS, familial ALS; MAO, mean age of onset; sALS, sporadic ALS; *SD*, standard deviation; SO, site of onset.

TABLE A4 Information on variants identified by WES and WGS

Gene	Chromosomal location	Amino acid change	dbSNP ID	GERP++ score (version 2010)	DANN score (version 2014)	ACMG verdict	ExAC
ERBB4	2:212251725	p.(Arg1096Cys)	rs144311212	5.59	0.99	Likely pathogenic	–
KIF5A	12:57976397	p.(Asp1002Gly)	–	5.47	0.99	Likely pathogenic	–
TBK1	12:64875731	p.(Arg308*)	rs1284582102	5.15	0.99	Pathogenic	–
	12:64882360	p.(Val479Glufs*4)	rs876657405	5.25	–	Likely pathogenic	–
VCP	9:35065361	p.(Arg155Cys)	rs121909330	6.02	0.99	Likely pathogenic	–
	9:35065349	p.(Arg159Cys)	rs387906789	6.07	0.99	Likely pathogenic	–
	9:35065252	p.(Arg191Pro)	–	5.64	0.99	Likely pathogenic	–
UBQLN2	X:56591822	p.(Pro506Ser)	rs387906711	4.17	0.93	VUS	–
	X:56591879	p.(Pro525Ser)	rs369947678	3.92	0.84	VUS	–
TFG	3:100467026	p.(Pro285Leu)	rs207482230	6.16	1	Likely pathogenic	–
ANG	14:21161931	p.(Ile70Val)	rs121909541	4.73	0.5	VUS	0.0006095
CHCHD10	22:24109646	p.(Ser59Leu)	–	3.66	0.99	Likely pathogenic	–
FBXO38	5:147796726	p.(Arg526Gln)	rs376255193	5.51	0.99	Likely pathogenic	0.07485
TRPV4	12:110236628	p.(Arg315Trp)	rs267607143	0.22	0.99	Likely pathogenic	–
TRPM7	15:50878630	p.(Thr1482Ile)	rs8042919	5.25	0.98	VUS	0.08703
SETX	9:135172384	p.(Ala1947Thr)	rs141440621	5.53	0.99	VUS	0.00008318
ERLIN1	10:101937913	p.(Val94Ala)	–	5.23	1	Pathogenic	–
SPG11	15:44943713	p.(Gln478*)	–	6.05	0.99	Pathogenic	–
	15:44943713	p.(Lys656Valfs*11)	–	5.71	–	Likely pathogenic	–
	15:44914992	p.(Phe750Leufs*3)	–	2.64	–	Pathogenic	–
	15:44855496	p.(Tyr2385*)	rs778305085	3.09	0.99	Pathogenic	–
OPTN	10:13151192	p.(His26Thrfs*19)	rs766608795	5.44	–	Pathogenic	–
	10:13164479	p.(Glu293Glyfs*19)	–	5.82	–	Likely pathogenic	–
	10:13167494	p.(Lys360Valfs*18)	–	5.44	–	Likely pathogenic	–
	10:13168014	p.(Thr406Lysfs*5)	–	5.57	–	Pathogenic	–
ALS2	2:202617888	p.(Ala573Glu)	–	5.16	0.99	Likely pathogenic	–
	2:202593315	p.(Arg921*)	rs587777132	5.82	0.99	Pathogenic	–
	2:202572614	p.(Arg1461*)	rs374047961	5.92	0.99	Pathogenic	–
	2:202569207	p.(Pro1603Leu)	–	5.53	0.99	Likely pathogenic	–
C19orf12	19:30199322	p.(Thr11Met)	rs397514477	−11.2	0.89	Likely pathogenic	0.00000831
	19:30193884	p.(Gly65Val)	–	4.57	0.99	Likely pathogenic	0.0000165

TABLE A4 (Continued)

Gene	Chromosomal location	Amino acid change	dbSNP ID	GERP++ score (version 2010)	DANN score (version 2014)	ACMG verdict	ExAC
SYNE1	6:152527392	p.(Gln7644*)	–	5.51	1	Pathogenic	–
	6:152497632	p.(Arg7842*)	rs775935265	5.75	1		0.000008237
ZFYVE26	14:68264904	p.(Leu692Serfs*52)	–	5.71	–	Pathogenic	–
	14:68257427	p.(Arg872Hisfs*17)	–	5.05	–	Likely pathogenic	–
DNAJB2	2:220144569	p.(Tyr5Cys)	rs730882140	2.34	1	Likely pathogenic	–
	2:220149491	p.(Glu253Lys)	–	4.48	0.99	Likely pathogenic	–
PLEKHG5	1:6530920	p.(Gln550*)	–	4.1	0.99	pathogenic	–
	1:6529648	p.(Pro707His)	–	5.67	0.97	Likely pathogenic	–
SIGMAR1	9:34635853	p.(Glu119Lys)	rs757260058	4.32	0.99	Likely pathogenic	–
	9:34635850	p.(Thr120Ala)	–	4.67	0.99	Likely pathogenic	–
VRK1	14:97326965	p.(Arg321Cys)	rs772731615	5.13	0.99	Likely pathogenic	0.0002071
	14:97342428	p.(Gln379Aspfs*23)	–	5.4	–	Pathogenic	–
DJ1	1:8025426	p.(Gln45*)	–	5.61	0.99	Pathogenic	–
IGHMBP2	11:68678998	p.(His213Arg)	rs137852666	4.7	0.99	Likely pathogenic	–
SLC52A3	20:744413	p.(Arg268Trp)	rs145498634	4.6	0.99	Likely pathogenic	0.00004945

Note: DANN score ranges is from 0 to 1, 1 predicted to be the most damaging, GERP++ score ranges from −12.3 to 6.17, 6.17 being the most conserved.
Abbreviations: VUS, variant of unknown significance; WES, whole exome sequencing; WGS, whole genome sequencing.

TABLE A5 ALS gene variants with unknown significance identified in WGS

Chromosome location	Gene	Transcript	DNA change	Protein change	Zygosity	rsID	No. of controls	No. of cases
chr9:35067907	*VCP*	NM_007126.5	c.283C>T	p.(Arg95Cys)	het	rs121909332	0	1
chr9:35068336	*VCP*	NM_007126.5	c.41C>T	p.(Thr14Ile)	het	–	1	0
chr2:212295795	*ERBB4*	NM_005235.3	c.2518G>A	p.(Val840Ile)	het	rs369248674	0	1
chr2:212989553	*ERBB4*	NM_005235.3	c.158A>G	p.(Tyr53Cys)	het	rs756650586	0	1
chr2:212589887	*ERBB4*	NM_005235.3	c.655G>A	p.(Gly219Ser)	het	rs757597004	0	1
chr2:212812268	*ERBB4*	NM_005235.3	c.308G>A	p.(Arg103His)	het	rs754487821	0	1
chr14:21162046	*ANG*	NM_001145.4	c.324dupT	p.(Gly109Trpfs*24)	het	–	0	1
chr9:135163699	*SETX*	NM_015046.7	c.6248G>T	p.(Arg2083Ile)	het	rs751252138	0	1
chr9:135172398	*SETX*	NM_015046.7	c.5825T>C	p.(Ile1942Thr)	het	rs773379832	1	0
chr4:170428877	*NEK1*	NM_012224.3	c.1816G>T	p.(Glu606*)	het	–	0	1
chr5:179251221	*SQSTM1*	NM_001142298.2	c.319G>A	p.(Gly107Arg)	het	rs781478225	0	1
chr17:4851653	*PFN1*	NM_005022.4	c.37G>A	p.(Ala13Thr)	hom	rs763837842	0	2
chr12:109278828	*DAO*	NM_001917.5	c.46G>A	p.(Ala16Thr)	het	rs778735604	0	1
chr12:109281243	*DAO*	NM_001917.5	c.212C>T	p.(Thr71Ile)	het	rs138277420	0	1
chr12:109286795	*DAO*	NM_001917.5	c.490G>T	p.(Val164Leu)	het	–	1	0
chr12:109294229	*DAO*	NM_001917.5	c.962G>A	p.(Gly321Glu)	het	–	0	1
chr2:74593112	*DCTN1*	NM_004082.4	c.2794C>T	p.(Arg932Cys)	het	rs373818927	1	1
chr2:74593597	*DCTN1*	NM_004082.4	c.2617G>A	p.(Ala873Thr)	het	rs764492372	1	0
chr2:74594488	*DCTN1*	NM_004082.4	c.2244C>G	p.(Asp748Glu)	het	rs751069902	0	1
chr2:74594495	*DCTN1*	NM_004082.4	c.2237T>C	p.(Leu746Pro)	het	–	0	1
chr2:74605312	*DCTN1*	NM_004082.4	c.94C>T	p.(Arg32Cys)	het	rs751177222	0	1
chr8:28017873	*ELP3*	NM_018091.6	c.1385G>A	p.(Arg462His)	hom	rs190129217	0	1
chr6:110036336	*FIG4*	NM_014845.5	c.122T>C	p.(Ile41Thr)	het	rs121908287	0	1
chr6:110056402	*FIG4*	NM_014845.5	c.547C>T	p.(Arg183*)	het	rs121908288	0	1
chr6:110086229	*FIG4*	NM_014845.5	c.1448G>A	p.(Arg483Gln)	het	rs749233172	0	1
chr6:110106223	*FIG4*	NM_014845.5	c.1940A>G	p.(Tyr647Cys)	het	rs150301327	1	0
chr19:7605129	*PNPLA6*	NM_006702.5	c.532G>T	p.(Gly178Cys)	het	–	0	1
chr19:7620584	*PNPLA6*	NM_006702.5	c.2914G>A	p.(Gly972Arg)	het	rs768107851	0	1
chr19:7626175	*PNPLA6*	NM_006702.5	c.365del	p.(Ala122Glnfs*3)	het	–	1	0
chr7:94935670	*PON1*	NM_000446.7	c.707A>G	p.(Tyr236Cys)	het	rs755475189	0	2
chr7:95024007	*PON3*	NM_000940.3	c.94C>T	p.(Arg32*)	het	rs147006695	0	5
chr12:49689173	*PRPH*	NM_006262.4	c.190C>T	p.(Arg64*)	het	–	0	2
chr15:78894232	*CHRNA3*	NM_000743.5	c.752C>G	p.(Pro251 Arg)	het	–	0	1
chr15:78894258	*CHRNA3*	NM_000743.5	c.725del	p.(Leu242Cysfs*32)	het	–	1	2
chr15:78894275	*CHRNA3*	NM_000743.5	c.708_709insG	p.(Ile237Aspfs*35)	het	–	1	2
chr15:78910978	*CHRNA3*	NM_000743.5	c.247_248insG	p.(Thr83Serfs*11)	het	–	1	0
chr15:78911138	*CHRNA3*	NM_000743.5	c.1A>G	p.?	het	rs745905590	0	3
chr20:61981766	*CHRNA4*	NM_000744.6	c.997C>T	p.(Arg333Cys)	het	rs761631713	1	0
chr20:61981784	*CHRNA4*	NM_000744.6	c.979G>A	p.(Val327Met)	het	rs201841018	0	1

TABLE A5 (Continued)

Chromosome location	Gene	Transcript	DNA change	Protein change	Zygosity	rsID	No. of controls	No. of cases
chr20:61982107	*CHRNA4*	NM_000744.6	c.656A>C	p.(Asn219Thr)	het	rs201645533	0	1
chr20:61982321	*CHRNA4*	NM_000744.6	c.442C>T	p.(Arg148Trp)	het	rs121912243	0	1
chr15:78921343	*CHRNB4*	NM_000750.5	c.1304C>T	p.(Ala435Val)	het	rs56317523	1	1
chr15:78922149	*CHRNB4*	NM_000750.5	c.498C>G	p.(Asn166Lys)	het	rs148540431	1	1
chr15:78927869	*CHRNB4*	NM_000750.5	c.116G>T	p.(Arg39Leu)	het	–	0	1

Abbreviations: ALS, amyotrophic lateral sclerosis; WGS, whole genome sequencing.

TABLE A6 The number of variants from WGS filtered according to minor allele frequency in gnomAD

	Total	<0.1% in TR cohort	<0.5% in TR cohort	<1% in TR cohort	<5% in TR cohort	≥5% in TR cohort
No. of novel variants	47971649	23339705	10652389	2166448	3648946	8164161
No. of existing variants in gnomAD	135851	70808	35459	6164	8483	14937

Abbreviation: WGS, whole genome sequencing.

TABLE A7 The top ranked variants from GWAS

Variant	Odds ratio	*p* value	dbSNP ID(s)
Odds ratio (case/control) > 1			
chr9:115956718:T:A	2.589	3E−06	rs10817455
chr18:19728805:T:A	2.414	6.1E−06	rs72879076
chr18:19728806:C:A	2.414	6.1E−06	rs62092220
chr5:4478547:A:C	2.150	6.3E−06	rs10066908
chr7:67216370:T:A	2.977	8.4E−06	rs6977845
chr17:20215352:T:G	1.984	8.6E−06	rs7223476
chr6:152561045:C:T	2.012	8.8E−06	rs1830820
chr7:67228631:A:G	2.068	9E−06	rs13307299
chr7:67231257:A:G	2.068	9E−06	rs9691826
chr6:152560478:C:A	2.010	9.3E−06	rs4869757
chr6:152559389:C:A	2.009	9.6E−06	rs9397089
chr6:152559605:A:G	2.009	9.6E−06	rs6928675
chr6:152559803:G:A	2.009	9.6E−06	rs6905741
chr7:67233851:T:C	2.027	1.2E−05	rs34912838
chr9:115974044:T:G	2.437	1.3E−05	rs4989078
chr17:39695278:C:T	2.100	1.3E−05	rs7212439
chr1:240672803:T:G	4.385	1.4E−05	rs61832588
chr7:67235252:T:G	2.068	1.5E−05	rs6946808
chr7:67235264:A:G	2.068	1.5E−05	rs6942794
chr17:39691789:G:A	2.087	1.5E−05	rs963478
chr17:19657847:C:G	2.194	1.5E−05	rs56191443

(Continues)

TABLE A7 (Continued)

Variant	Odds ratio	*p* value	dbSNP ID(s)
chr7:67236543:T:C	2.072	1.5E−05	rs35629450
chr7:67237915:A:C	2.059	1.6E−05	rs6964106
chr17:39677699:T:C	2.079	1.6E−05	rs56389952
chr17:39684410:G:A	2.079	1.6E−05	rs11550883
Odds ratio (case/control) < 1			
chr11:56460941:TTTG:T	0.403	2.1E−07	rs146003833, rs3071452, rs567274730
chr11:56466099:C:T	0.409	2.9E−07	rs1397048
chr11:56467085:T:C	0.409	2.9E−07	rs10896513
chr11:56465305:T:A	0.410	3E−07	rs1509995
chr11:56451517:A:T	0.411	3.3E−07	rs4340069
chr11:56441125:T:C	0.412	3.6E−07	rs7130569
chr11:56453126:G:T	0.412	3.6E−07	rs7109249
chr11:56453323:C:T	0.412	3.6E−07	rs1588387
chr11:56454288:G:A	0.412	3.6E−07	rs7113794
chr11:56457140:T:C	0.412	3.6E−07	rs55848395
chr5:32087228:C:T	0.472	5.1E−07	rs157494
chr5:32087802:A:G	0.473	5.1E−07	rs157497
chr5:32101168:A:T	0.475	5.1E−07	rs2279232
chr5:32094654:A:G	0.465	5.5E−07	rs245154
chr5:32103496:C:T	0.480	7E−07	rs4867419
chr11:121301819:C:A	0.429	7.2E−07	rs17125331
chr11:56455965:TAA:T	0.415	7.5E−07	rs3071436, rs71058026, rs762762596
chr11:121301299:TA:T	0.435	1.1E−06	rs34153159
chr11:121306462:G:A	0.439	1.1E−06	rs4146874
chr11:121302634:T:A	0.441	1.3E−06	rs55736743
chr11:121302646:A:G	0.441	1.3E−06	rs17125333
chr11:121303206:G:A	0.441	1.3E−06	rs17125336
chr11:121303952:CAAA:C	0.441	1.3E−06	rs10616182
chr11:121303574:T:G	0.446	1.7E−06	rs12417885
chr8:24329635:C:CA	0.281	1.9E−06	rs370799783

Abbreviation: GWAS, Genome-wide association studies.

TABLE A8 The top ranked genes from SKAT-O analysis

Symbol	Gene ID	# of controls (# of variants)	# of cases (# of variants)	Odds ratio	*p* value
Odds ratio > 1					
RALGAPA1	ENSG00000174373	3 (3)	15 (15)	1.14	.0020094
C16orf3	ENSG00000221819	0 (0)	2 (6)	–	.0052194
C14orf23	ENSG00000186960	3 (3)	32 (32)	2.43	.0061734
RET	ENSG00000165731	2 (2)	13 (13)	2.48	.0083422
DENND2C	ENSG00000175984	1 (1)	7 (7)	1.6	.0096899
RPL10L	ENSG00000165496	0 (0)	2 (2)	–	.0101959
C17orf96	ENSG00000179294	1 (1)	6 (6)	1.37	.0120467
ACSM5	ENSG00000183549	0 (0)	3 (3)	–	.0139554
BIRC6	ENSG00000115760	1 (1)	7 (7)	1.6	.014459
TRIM49	ENSG00000168930	0 (0)	2 (4)	–	.015285
Odds ratio < 1					
IKZF2	ENSG00000030419	7 (7)	12 (12)	0.39	.0009985
RTTN	ENSG00000176225	6 (7)	6 (6)	0.23	.0010553
CDK14	ENSG00000058091	2 (2)	1 (1)	0.11	.001121
ATP5S	ENSG00000125375	2 (2)	1 (1)	0.11	.001198
WDR86	ENSG00000187260	2 (2)	3 (4)	0.34	.0012955
GPRC5D	ENSG00000111291	1 (1)	3 (3)	0.68	.0013205
FRMD6	ENSG00000139926	3 (3)	2 (2)	0.15	.0019035
TPCN1	ENSG00000186815	1 (1)	4 (4)	0.91	.0020293
KCTD1	ENSG00000134504	4 (4)	3 (4)	0.17	.0021836
CCDC80	ENSG00000091986	2 (2)	0 (0)	0	.0029323

TABLE A9 Enriched gene clusters based on coexpression network analysis

Category/term	Count	List total	Fold enrichment	%	*p* value	Bonferroni	Benjamini	FDR	Genes
Cell Cycle-Related Cluster-Enrichment Score: 15.03									
UP_KEYWORDS/cell cycle	28	92	9	30	7.66E−19	1.26E−16	1.26E−16	9.29E−16	ENSG00000010292, ENSG00000161800, ENSG00000123485, ENSG00000184661, ENSG00000129195, ENSG00000237649, ENSG00000111665, ENSG00000144554, ENSG00000111602, ENSG00000171320, ENSG00000137804, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000183765, ENSG00000111206, ENSG00000099956, ENSG00000138182, ENSG00000080986, ENSG00000126787, ENSG00000142945, ENSG00000117399, ENSG00000186871, ENSG00000115760, ENSG00000128944, ENSG00000165304, ENSG00000175063, ENSG00000087586
UP_KEYWORDS/mitosis	20	92	17	21	2.79E−18	4.57E−16	2.28E−16	3.38E−15	ENSG00000010292, ENSG00000184661, ENSG00000138182, ENSG00000129195, ENSG00000237649, ENSG00000080986, ENSG00000111665, ENSG00000142945, ENSG00000117399, ENSG00000186871, ENSG00000115760, ENSG00000111602, ENSG00000128944, ENSG00000137804, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000175063, ENSG00000087586, ENSG00000183765
UP_KEYWORDS/cell division	21	92	11	22	5.53E−16	9.10E−14	3.03E−14	6.77E−13	ENSG00000010292, ENSG00000161800, ENSG00000184661, ENSG00000138182, ENSG00000129195, ENSG00000237649, ENSG00000080986, ENSG00000111665, ENSG00000142945, ENSG00000117399, ENSG00000186871, ENSG00000115760, ENSG00000111602, ENSG00000128944, ENSG00000137804, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000175063, ENSG00000087586, ENSG00000183765
GOTERM_BP_DIRECT/GO:0051301~cell division	20	88	11	21	4.59E−15	2.36E−12	2.36E−12	6.59E−12	ENSG00000010292, ENSG00000118193, ENSG00000184661, ENSG00000138182, ENSG00000129195, ENSG00000237649, ENSG00000080986, ENSG00000111665, ENSG00000142945,

TABLE A9 (Continued)

Category/term	Count	List total	Fold enrichment	%	*p* value	Bonferroni	Benjamini	FDR	Genes
									ENSG00000117399, ENSG00000186871, ENSG00000115760, ENSG00000111602, ENSG00000128944, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000175063, ENSG00000087586, ENSG00000183765
GOTERM_BP_DIRECT/GO:0007067~mitotic nuclear division	14	88	11	15	1.41E−10	7.32E−08	3.66E−08	2.04E−07	ENSG00000184661, ENSG00000138182, ENSG00000129195, ENSG00000080986, ENSG00000111665, ENSG00000142945, ENSG00000117399, ENSG00000186871, ENSG00000115760, ENSG00000111602, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000087586
Chromosome Structure Related Cluster-Enrichment Score: 6,8									
GOTERM_CC_DIRECT/ GO:0000777~condensed chromosome kinetochore	10	90	22	11	4.44E−10	7.41E−08	2.47E−08	5.40E−07	ENSG00000120071, ENSG00000123485, ENSG00000128944, ENSG00000111581, ENSG00000080986, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000186871, ENSG00000142945
UP_KEYWORDS/chromosome	15	92	9	16	8.64E−10	1.42E−07	2.36E−08	1.05E−06	ENSG00000010292, ENSG00000120071, ENSG00000123485, ENSG00000090889, ENSG00000080986, ENSG00000142945, ENSG00000186871, ENSG00000143476, ENSG00000128944, ENSG00000111581, ENSG00000171320, ENSG00000137804, ENSG00000152253, ENSG00000129810, ENSG00000184445
UP_KEYWORDS/centromere	10	92	17	11	4.17E−09	6.84E−07	9.77E−08	5.06E−06	ENSG00000120071, ENSG00000123485, ENSG00000128944, ENSG00000111581, ENSG00000080986, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000186871, ENSG00000142945
UP_KEYWORDS/kinetochore	9	92	19	10	1.49E−08	2.45E−06	3.06E−07	1.81E−05	ENSG00000120071, ENSG00000128944, ENSG00000111581, ENSG00000080986, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000186871, ENSG00000142945
GOTERM_BP_DIRECT/ GO:0007059~chromosome segregation	8	88	17	9	2.87E−07	1.49E−04	4.97E−05	4.16E−04	ENSG00000131747, ENSG00000123485, ENSG00000184661, ENSG00000128944, ENSG00000080986,

(Continues)

TABLE A9 (Continued)

Category/term	Count	List total	Fold enrichment	%	*p* value	Bonferroni	Benjamini	FDR	Genes
									ENSG00000171320, ENSG00000152253, ENSG00000129810
GOTERM_BP_DIRECT/GO:0007062~sister chromatid cohesion	8	88	16	9	5.79E−07	3.01E−04	7.52E−05	8.39E−04	ENSG00000111581, ENSG00000080986, ENSG00000152253, ENSG00000129810, ENSG00000184445, ENSG00000186871, ENSG00000117399, ENSG00000142945
GOTERM_CC_DIRECT/ GO:0000776~kinetochore	6	90	13	6	8.29E−05	0.013743	0.002764	0.100739	ENSG00000128944, ENSG00000111581, ENSG00000080986, ENSG00000112742, ENSG00000129810, ENSG00000142945
GOTERM_CC_DIRECT/ GO:0000775~chromosome, centromeric region	4	90	18	4	0.001218	0.184197	0.020152	1.471814	ENSG00000123485, ENSG00000080986, ENSG00000129810, ENSG00000142945

TABLE A10 The top ranked pathways in SKAT-O analysis

Pathway	# of variant (control)	# of variant (case)	Odds ratio	*p* value (SKAT-O)
Odds ratio > 1				
BIOCARTA_FEEDER_PATHWAY	15	54	0.82	.0008682
REACTOME_FGFR4_LIGAND_BINDING_AND_ACTIVATION	16	58	0.83	.0014919
KEGG_GALACTOSE_METABOLISM	60	200	0.76	.0035946
BIOCARTA_CARDIACEGF_PATHWAY	7	23	0.75	.0040825
KEGG_ETHER_LIPID_METABOLISM	12	34	0.65	.0052787
REACTOME_FGFR2C_LIGAND_BINDING_AND_ACTIVATION	4	6	0.34	.0063526
REACTOME_PI_METABOLISM	41	121	0.67	.0086292
REACTOME_POL_SWITCHING	2	1	0.11	.0087933
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES	19	54	0.65	.0105438
REACTOME_LAGGING_STRAND_SYNTHESIS	4	11	0.63	.0107531
Odds ratio < 1				
REACTOME_GLYCOGEN_BREAKDOWN_GLYCOGENOLYSIS	19	125	1.50	.0073818
KEGG_INOSITOL_PHOSPHATE_METABOLISM	47	235	1.14	.0160813
PID_RET_PATHWAY	7	62	2.02	.0166805
KEGG_DNA_REPLICATION	5	30	1.37	.0222791
KEGG_ARGININE_AND_PROLINE_METABOLISM	25	130	1.19	.0234903
REACTOME_GLUCOSE_METABOLISM	65	324	1.14	.0322741
SA_B_CELL_RECEPTOR_COMPLEXES	4	22	1.25	.0327217
REACTOME_DNA_REPAIR	36	179	1.13	.0396380
PID_NETRIN_PATHWAY	15	74	1.12	.0404352
PID_BCR_5PATHWAY	6	43	1.63	.0405759