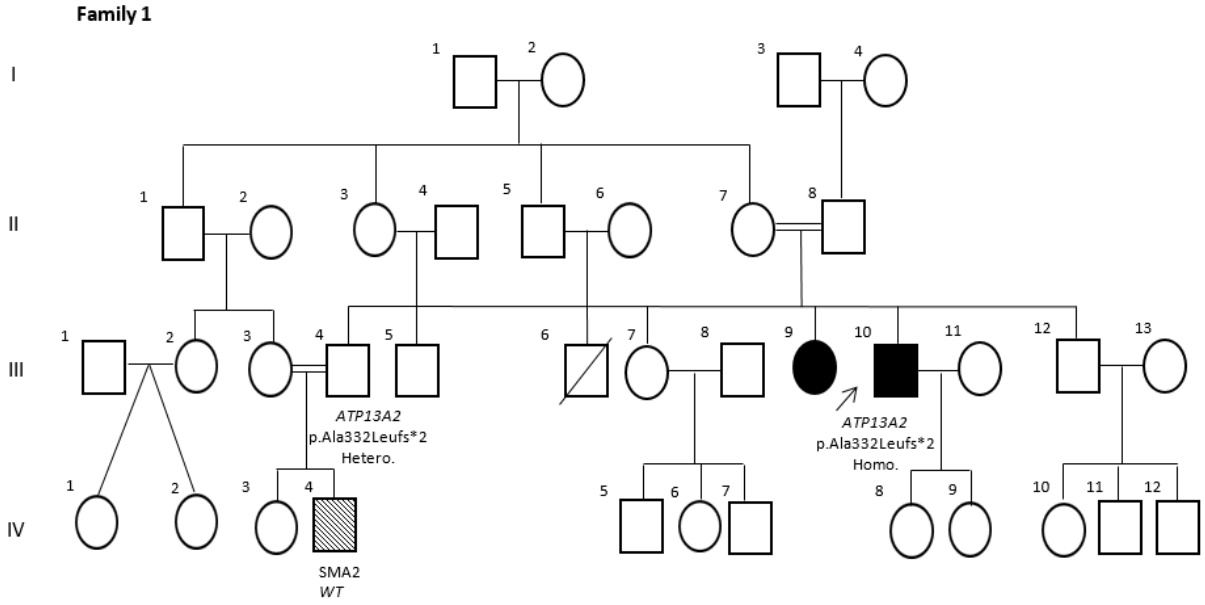
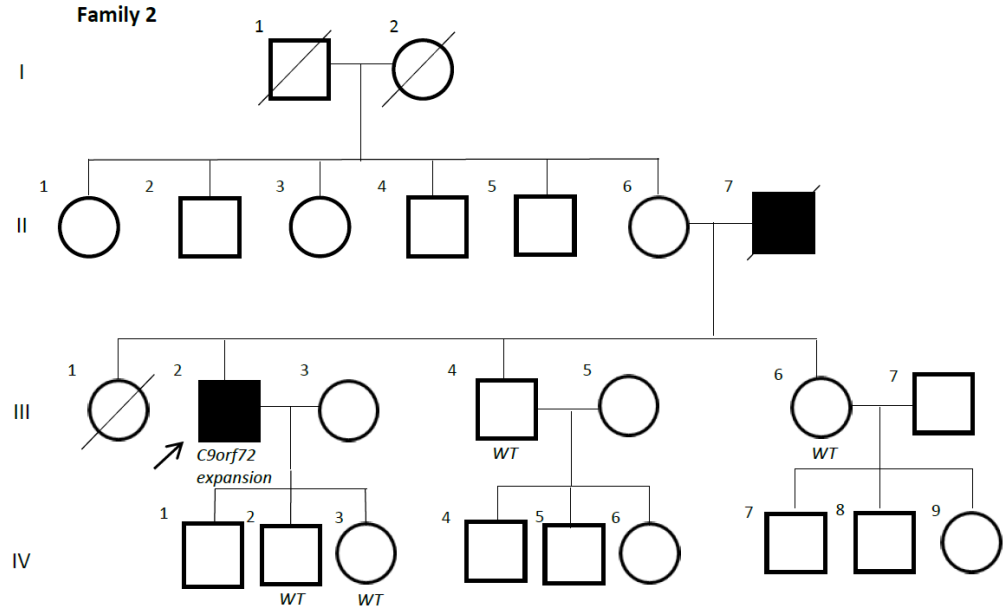
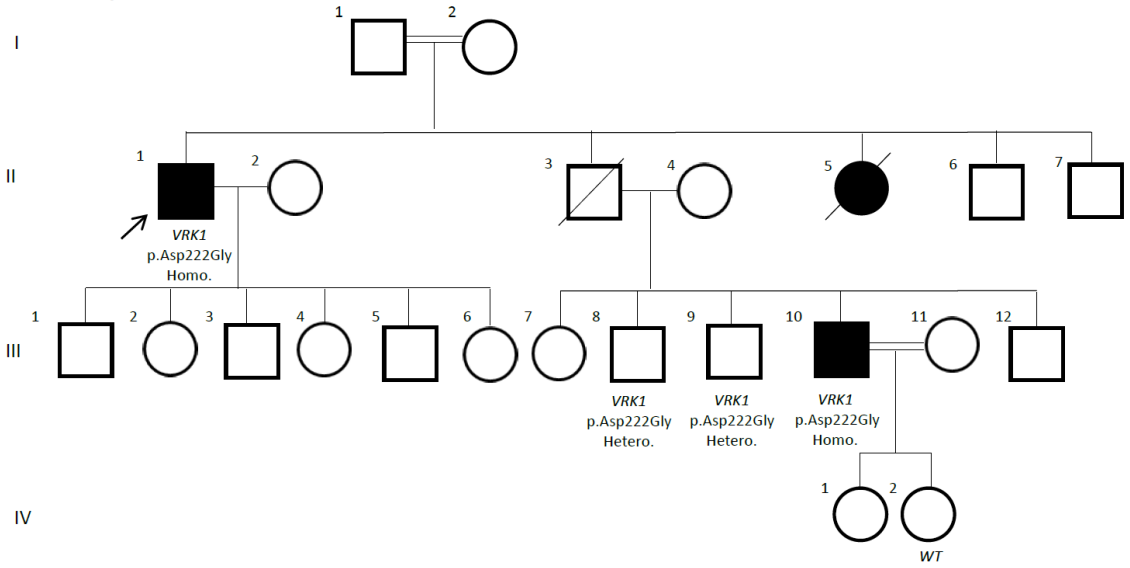
Supplementary information

First insights into genotype and phenotype of familial amyotrophic lateral sclerosis in Egypt: Early onset and high consanguinity

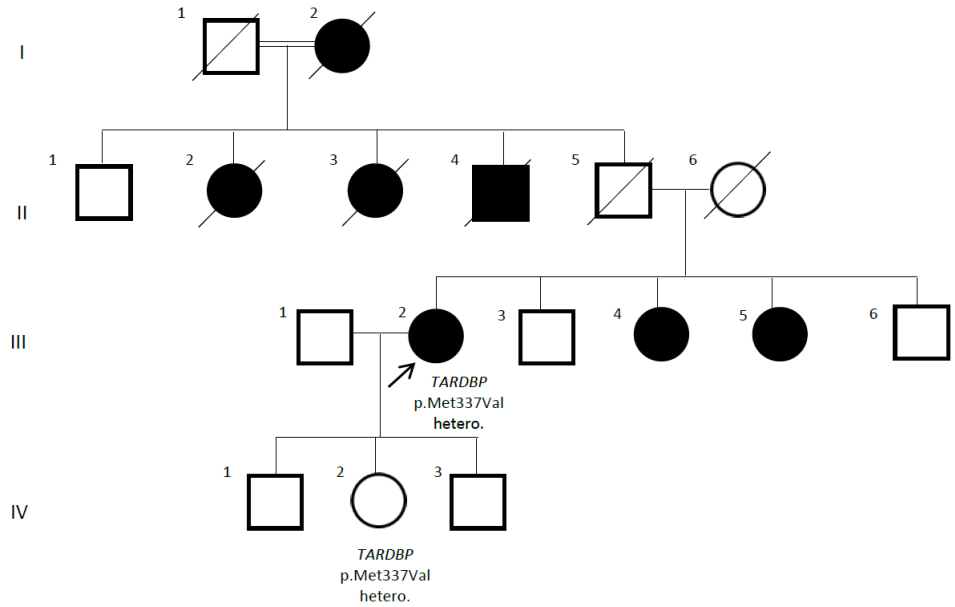




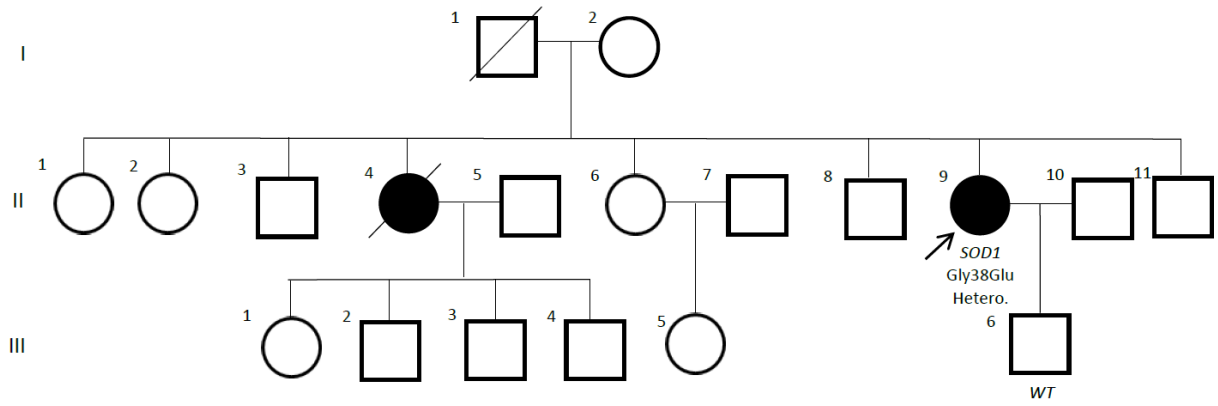
Family 3



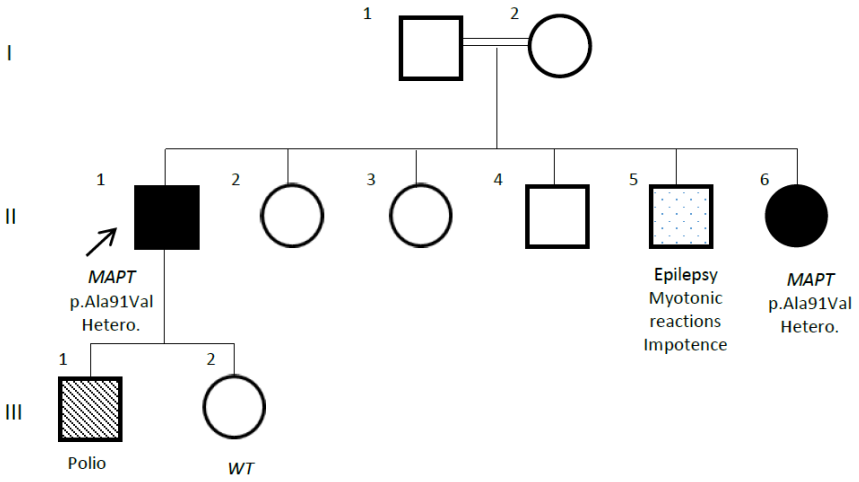
Family 4



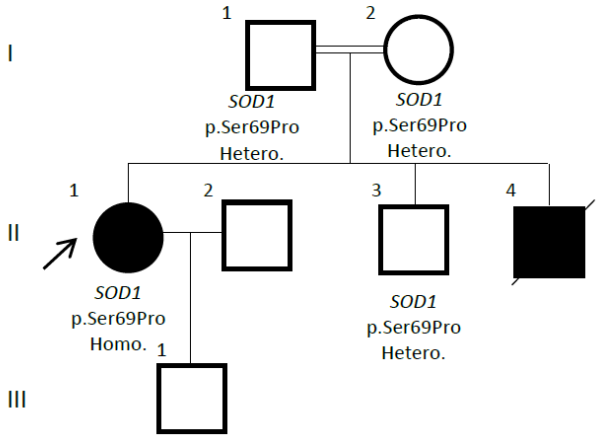
Family 5



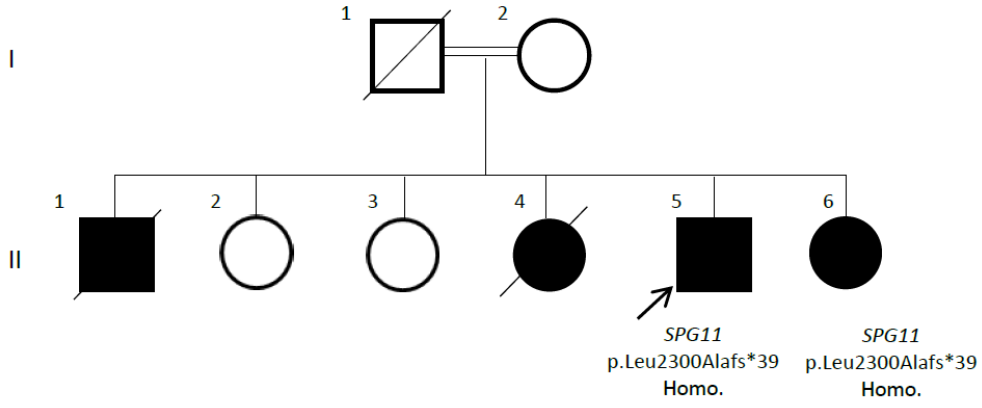
Family 6



Family 7



Family 8



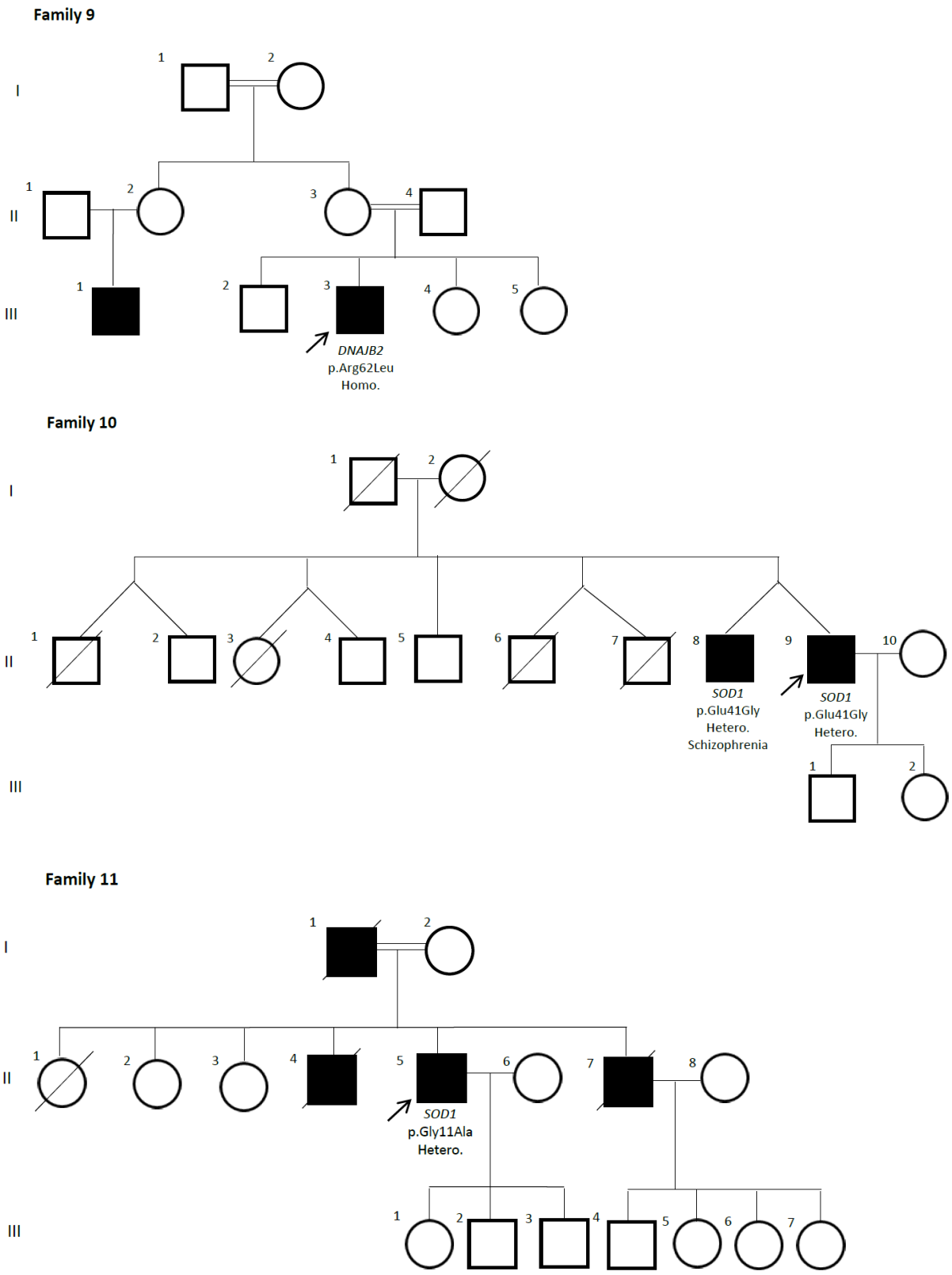


Fig. S1 ALS pedigrees for 11 families. Affected individuals are indicated with solid black symbols. Symbols with a diagonal line show the individual is deceased. Circles represent females. Squares represent males. The proband is pointed at with an arrow. Genetic variants and homo/heterozygosity distinction are displayed where information was available for individuals who have been tested. Other phenotypes are indicated with a symbol patter.

Table S1 Demographic and basic clinical characteristics of the whole Egyptian Cohort of fALS

Family	Affected members	Genetic test	Sex	Age at onset (years)	Age at diagnosis (years)	Consang-uinity	El Escorial criteria/ Other*	Site of onset	ALS-FRS R at diagnosis	Unaffected Members genetically tested
1	Proband III10	Yes	M	24	36	Parents	Definite	LL	25	III4, IV4
	Sister III9	No	F	27	38	first cousins	Definite	LL	19	
2	Proband III2	Yes	M	41	48	No	Definite	LL	29	III4, III6
	Father II7	No	M	59	?		Other	?	?	IV2, IV3
3	Proband II1	Yes	M	43	64	Parents	Probable	LL	38	III8, III9
	Sister II5	No	F	?	?	first	Other	?	?	
	Nephew III10	Yes	M	32	?	cousins	Other	?	?	
4	Proband III2	Yes	F	55	66	Grand-	Definite	LL	35	IV2
	Grandmother I2	No	F	?	?	parents	Other	?	?	
	Uncle II4	No	M	65	?	first	Other	?	?	
	Aunt II2	No	F	63	?	cousins	Other	?	?	
	Aunt II3	No	F	66	?		Other	?	?	
	Sister III4	No	F	49	?		Other	?	?	
	Sister III5	No	F	52	?		Other	?	?	
5	Proband II9	Yes	F	33	38	No	Definite	UL	34	III6
	Sister II4	No	F	30	?		Other	?	?	
6	Proband II1 [5]	Yes	M	61	64	Parents	Definite	UL	24	III2
	Sister II6	Yes	F	53	?	first cousins	Other	?	?	
7	Proband II1	Yes	F	26	26	Parents	Definite	LL	38	I1, I2, II3
	Brother II4	No	M	10	12	first cousins	Definite	LL	?	
8	Proband II5	Yes	M	20	27	Parents first	Probable-lab-supported	UL	32	None
	Sister II6	Yes	F	28	32	cousins	Probable	UL	35	
	Sister II4	No	F	21	?		Other	?	?	
	Brother II5	No	M	20	?		Other	?	?	
9	Proband III3	Yes	M	31	32	Parents	Definite	Bulbar	44	None
	Cousin III1	No	M	?	?	first cousins	Other	?	?	
10	Proband II9	Yes	M	57	59	No	Definite	LL	38	None
	Twin brother II8 (dizygotic)	Yes	M	57	59		Definite	LL	44	
11	Proband II5	Yes	M	57	59	Parents	Definite	LL	44	None
	Father I1	No	M	60	?	first	Other	?	?	
	Brother II4	No	M	22	?	cousins	Other	?	?	
	Brother II7	No	M	35	?		Other	?	?	
	Total	15 tested 17 untested	18 M 14 F	Mean (±SD) 41.28 ±16.9	Mean (±SD) 44 ±17	8/11 72.7%		UL 67% LL 27% Bulbar 7%	34.21 ±7.49	14 tested

UL: upper limbs. LL: lower limbs. "?": information could not be provided by the patient, or was unavailable because proband's relatives did not present for diagnosis. *Other: phenotype reported by family members.

[5]: This previously reported family has been included here for comparison in terms of genotype-phenotype correlation.

Table S2 ALS gene panel analysed in this study, ordered alphabetically [1]

Gene symbol	Gene name	Transcript
ALS2	Alsin Rho guanine nucleotide exchange factor ALS2	NM_020919.4
ANG	Angiogenin	NM_001145.4
ANXA11	Annexin A11	NM_145869.1
ATXN2	Ataxin 2	NM_002973.4
C9orf72	C9orf72-SMCR8 complex subunit	NM_001256054.3
C21orf2	Chromosome 21 open reading frame 2	NM_004928.2
CCNF	Cyclin F	NM_001761.3
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain containing 10	NM_001301339.2
CHMP2B	Charged multivesicular body protein 2B	NM_014043.4
DAO	D-amino-acid oxidase	NM_001917.4
DCTN1	Dynactin subunit 1	NM_004082.5
ERBB4	Erb-b2 receptor tyrosine kinase 4	NM_005235.3
FIG4	FIG4 phosphoinositide 5-phosphatase	NM_014845.6
FUS	FUS RNA binding protein	NM_004960.4
GLE1	GLE1 RNA export mediator	NM_001003722.2
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	NM_031157.4
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	NM_031243.3
MAPT	Microtubule associated protein tau	NM_001123066.4
MATR3	Matrin 3	NM_018834.6
NEFH	Neurofilament heavy chain	NM_021076.4
NEK1	NIMA related kinase 1	NM_001199397.3
OPTN	Optineurin	NM_001008211.1
PFN1	Profilin 1	NM_005022.4
SETX	Senataxin	NM_015046.7
SIGMAR1	Sigma non-opioid intracellular receptor 1	NM_005866.4
SOD1	Superoxide dismutase 1	NM_000454.5
SPG11	SPG11 vesicle trafficking associated, spatacsin	NM_025137.4
SQSTM1	Sequestosome 1	NM_003900.5
TARDBP	TAR DNA binding protein	NM_007375.4
TBK1	TANK binding kinase 1	NM_013254.4
TIA1	TIA1 cytotoxic granule-associated RNA binding protein	NM_022173.2
TUBA4A	Tubulin alpha 4a	NM_006000.3
UBQLN2	Ubiquilin 2	NM_013444.4
VAPB	VAMP associated protein B and C	NM_004738.5
VCP	Valosin containing protein	NM_007126.5

Table S3 Summary of the variant analysis and ACMG classification

Gene	Pedigree	Variant			MAF GnomAD	ClinVar	LOVD	HGMD	Mutation Taster (Tree vote Deleterious | benign)	SNP&Go (Reliability Index	SIFT	FATHMM-XF score	Phast Cons 46way	Phylo-P 46way	ACMG Classific-ation
		cDNA	AA Change	Homo/ Hetero-zygous											
SOD1	7	c.205T>C	p.Ser69Pro	Homo.	not listed	Likely pathogenic	not listed	DCM	87|13	Disease 10	Damaging 0.01	Pathogenic 0.808	0.335	0.066	Likely pathogenic
	5	c.113G>A	p.Gly38Glu	Hetero.	not listed	not listed	not listed	DCM	98|2	Disease 10	Damaging 0.00	Pathogenic 0.942	0.99	5.16	Likely pathogenic
	11	c.32G>C	p.Gly11Ala	Hetero.	not listed	not listed	pathogenic	DCM	90|10	Disease 10	Damaging 0.01	Pathogenic 0.930	1.00	4.42	Likely pathogenic
	10	c.122A>G	p.Glu41Gly	Hetero.	not listed	Pathogenic/ VUS	not listed	DCM	92|8	Disease 10	Damaging 0.03	Pathogenic 0.910	0.78	1.79	Likely Pathogenic
TARDBP	4	c.1009A>G	p.Met337Val	Hetero.	0.000008	Pathogenic	Pathogenic	DCM	94|6	Neutral 9	Tolerated 0.18	Pathogenic 0.951	1.00	5.13	Pathogenic
MAPT	6	c.272C>T	p.Ala91Val	Hetero.	0.0001022	VUS	not listed	DCM	0|100	Disease 5	Damaging 0.01	Pathogenic 0.552	1.00	3.14	VUS
SPG11	8	c.6898_6899del	p.Leu2300Alafs*39	Homo.	0.000004	Pathogenic	pathogenic	DCM	200|0	—	—	—	0.97	3.77	Pathogenic
ATP13A2	1	c.984_993dup	p.Ala332Leufs*2	Homo.	not listed	not listed	not listed	not listed	185|15	—	Tolerated 0.86	—	0.98	2.42	Pathogenic
VRK1	3	c.665A>G	p.Asp222Gly	Homo.	not listed	not listed	not listed	not listed	58|42	Neutral 1	Damaging 0.03	Pathogenic 0.895	1.00	5.17	Likely pathogenic
DNAJB2	9	c.185G>T	p.Arg62Leu	Homo.	not listed	not listed	not listed	not listed	79|21	Disease 6	Damaging 0.00	Pathogenic 0.924	1.00	5.65	Likely pathogenic

Methods

Study cohort

Eleven ALS families were recruited at the Neuromuscular Unit, Ain Shams University, Cairo, Egypt from 2021 to 2023, with a total of 32 subjects with ALS phenotype. For each family, ALS diagnosis was based on the revised El Escorial criteria for the probands and their affected relatives who presented to the Neuromuscular Unit. For the other family members, the diagnosis was determined by reported phenotypes consistent with ALS by the respective family relatives (indicated by "other" in Table S1).

A total of 30 participants, including 15 affected and 15 unaffected family members were subjected to genetic testing (Table S1). In addition, 49 healthy control subjects were included to assess the distribution of *ATXN2* alleles in the population. All participants provided written informed consent for genetic testing, following the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. The study was approved by the Research Ethics Committee of Ain Shams Faculty of Medicine (ID: FMASU MS 298/2021).

Genetic analysis

Genomic DNA was extracted from EDTA-collected venous blood. C9orf72 hexanucleotide repeat expansion (HRE) and ATXN-2 polyQ expansion were analysed using conventional and repeat-primed PCR (RP-PCR)-based fragment analysis [2], with further analysed using Southern blot for C9orf72 repeat size determination [3]. Whole exome sequencing was performed using Illumina's NexteraTM Exome Kit, generating an average of 10 gigabases of sequence. The average depth was 125×, with at least 95% coverage of target regions at a minimum depth of 20×. For samples not meeting quality criteria, the TWISTTM Exome Kit was used. GeneMapper V 3.0 software was used for fragment data analysis.

Variant analysis

A virtual panel comprising 42 known ALS-genes (Table S2) [1] was applied to the exome data of all families, using the varvis® software platform (Limbus Technologies GmbH, Rostock). If no class 3 (uncertain significance), 4 (likely pathogenic), or 5 (pathogenic) variants according to ACMG recommendations [4], were found in these genes, the data were further analyzed on exome level with no pre-determined gene panel. Non-synonymous variants (SNVs, indels) in protein-coding regions, including exon-intron borders, were considered for analysis if their minor allele frequency in GnomAD/ExAC and dsSNP was ≤0.1%. Variant relevance, mode of inheritance, and pathogenicity were assessed using standard databases: OMIM, ClinVar, LOVD, HGMD professional, DECIPHER, and literature. The potential functional consequences of missense variants were evaluated using four in silico predictive algorithms: SIFT, Mutation Taster, FATHMM-XF, and SNPs & GO. Frameshift variants and indels were assessed using SIFT and Mutation Taster for functional impact. VarSEAK was used to test splicing effects. Conservation scores were obtained using PhastCons and PhyloP tools in USCS with 46 species assemblies. Variants' pathogenicity was reported based on the American College of Medical Genetics and Genomics (ACMG) criteria. For variants of unknown significance (VUS), pathogenicity was further predicted using AlphaMissense [5].

References

1. Müller K, Brenner D, Weydt P, Meyer T, Grehl T, Petri S, et al. Comprehensive analysis of the mutation spectrum in 301 German ALS families. J Neurol Neurosurg Psychiatry. 2018;89: 817–827.

2. Warner JP, Barron LH, Goudie D, Kelly K, Dow D, Fitzpatrick DR, et al. A general method for the detection of large CAG repeat expansions by fluorescent PCR. J Med Genet. 1996;33: 1022–1026.

3. Hübers A, Marroquin N, Schmoll B, Vielhaber S, Just M, Mayer B, et al. Polymerase chain reaction and Southern blot-based analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. Neurobiol Aging. 2014;35: 1214.e1–6.

4. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17: 405–424.

5. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381: eadg7492.