Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

In silico Evaluation of Deleterious nsSNPs in SOD1 and TARDBP Genes Associated with Amyotrophic Lateral Sclerosis

Manjunath B Malshetty
Manipal Academy of Higher Education
Sandeep Mallya
Manipal Academy of Higher Education
Sudharshan Prabhu
sudharshan.prabhu@manipal.edu

Manipal Academy of Higher Education

Article

Keywords: Amyotrophic lateral sclerosis (ALS), SOD1, TARDBP (TDP-43), nsSNPs, protein stability, structural modeling.

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8978267/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 27th, 2026. See the published version at https://doi.org/10.1038/s41598-026-68712-2.

Abstract

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder involving mutations in the *SOD1* and *TARDBP* genes that significantly influence its development. Nonsynonymous SNPs in these genes can disrupt protein stability, folding, and regulatory functions, leading to motor neuron loss. This study performed an in-silico approach to assess the functional and structural effects of nsSNPs in *SOD1* and *TARDBP*. A total of 5,964 *SOD1* and 10,630 *TARDBP* variants were retrieved from public databases, filtered for coding region with a MAF ≥ 0.001, and prioritized using CADD. Multiple approaches, including pathogenicity predictions, stability analysis, structural modeling, post-translational modification assessment, and network-based functions, were combined. 12 nsSNPs per gene met the inclusion criteria. Notably, *SOD1* variants V15G (rs1202989817) and I19M (rs1182088847) consistently predicted as deleterious, showing decreased stability indicated by negative ΔΔG values and localized structural disruptions without global misfolding. Conversely, *TARDBP* variants G335D (rs80356729) and I222T (rs1570722030) suggested destabilization but yielded mixed predictions regarding disease association. Network and pathway analyses highlighted *SOD1* and *TARDBP* as key nodes in ALS-related mechanisms such as oxidative stress, RNA metabolism, proteostasis, and mitochondrial impairment. These findings prioritize structurally destabilizing variants with potential pathogenic relevance in ALS and provide a computational framework for downstream experimental validation.

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive and deadly neurodegenerative disorder characterized by the selective loss of upper and lower motor neurons (Vintilescu et al., 2016), leading to muscle weakness, spasticity, dysarthria, dysphagia, and eventually respiratory failure. Most patients die within 2–5 years of symptom appearance, although survival varies based on genetic factors, environmental exposures, and molecular influences. (Michael A Van Es, 2024; Yuan et al., 2024). ALS mainly affects individuals between 50–70 years old and includes both sporadic ALS (sALS), making up nearly 90% of cases, and familial ALS (fALS), accounting for approximately 10% (Goutman et al., 2022; Huang et al., 2024). Increasing evidence shows that ALS exists within a spectrum of neurodegenerative disorders, sharing key pathological mechanisms with frontotemporal dementia (FTD), primarily through the buildup of TAR DNA-binding protein 43 (TDP-43) inclusions (Balendra et al., 2025; Feneberg et al., 2025; Zeng et al., 2024).

More than 40 genes have been linked to ALS, involved in pathways related to proteostasis, oxidative stress responses, RNA metabolism, axonal transport, and mitochondrial maintenance (Goutman et al., 2022). Among these, *SOD1* (superoxide dismutase 1) and *TARDBP* (encoding TDP-43) are two of the most significant contributors to ALS pathology. Under normal physiological conditions, *SOD1* converts toxic superoxide radicals into hydrogen peroxide and oxygen, thereby protecting neurons from oxidative damage. Meanwhile, *TARDBP* (TDP-43) plays a role in regulating transcription, alternative splicing, mRNA transport, and microRNA processing. Mutations in *SOD1* account for approximately 15–20% of familial ALS (fALS) and 1–2% of sporadic ALS (sALS), making it one of the earliest and most well-studied ALS-related genes (Huang et al., 2024; Opie-Martin et al., 2022). Although *TARDBP* mutations are less common, TDP-43 pathology appears in over 95% of ALS cases, emphasizing its key role in both familial and sporadic forms (Balendra et al., 2025; Wang et al., 2024). Single-nucleotide polymorphisms (SNPs), particularly nonsynonymous SNPs (nsSNPs), cause amino acid substitutions that alter protein structure and function. In ALS, nsSNPs in *SOD1* and *TARDBP* alter protein stability, aggregation, and toxicity, thereby influencing disease onset and progression (Berdyński et al., 2022; Dash et al., 2022; Lépine et al., 2024).

Mutant *SOD1* proteins acquire toxic gain-of-function properties, involving abnormal folding, disrupted metallation, defective maturation, and an increased tendency to aggregate (Trist et al., 2021, 2022). These alterations promote the formation of insoluble aggregates, impair mitochondrial activity, disrupt axonal structure, and induce cellular stress responses that compromise motor neuron survival. Considerable heterogeneity exists among *SOD1* mutations, with

different variants influencing disease onset, progression, and survival (Berdyński et al., 2022, 2025; Opie-Martin et al., 2022). Specific mutations can alter polypeptide length or structural stability, modify aggregation kinetics, and contribute to variant-specific neurotoxic profiles (Berdyński et al., 2025). Pathological studies have further revealed abnormalities in *SOD1* maturation, metal imbalance, and oxidation states in ALS spinal cord tissue, highlighting their contribution to the selective vulnerability of motor neurons (Trist et al., 2022).

Meanwhile, TDP-43 pathology represents another core hallmark of ALS. In healthy neurons, TDP-43 resides in the nucleus; however, in disease states, it mislocalizes to the cytoplasm and forms phosphorylated, ubiquitinated, and insoluble aggregates that impair RNA metabolism and destabilize neuronal function (Balendra et al., 2025; Zeng et al., 2024). TDP-43 proteinopathy is also observed in a broader range of neurodegenerative diseases and is characterized by prion-like propagation and progressive accumulation of misprocessed TDP-43 species (de Boer et al., 2021).

TARDBP mutations further worsen this pathology by disrupting RNA-binding activity and nucleocytoplasmic transport, ultimately impairing neuronal function (Balendra et al., 2025; Suk & Rousseaux, 2020). Studies using human iPSC-derived motor neurons show that ALS-linked *TARDBP* variants cause early electrophysiological problems and reduced synaptic activity before significant aggregation occurs (Lépine et al., 2024). Additional research demonstrates rapid mislocalization of endogenous TDP-43 in inducible models (Ganssauge et al., 2025), and region-specific proteomic analysis reveals alterations in ALS brain and spinal cord tissues (Feneberg et al., 2025). Moreover, TDP-43 buildup in peripheral tissues such as the retina may serve as an early biomarker of neurodegeneration (Glashutter et al., 2025)

Despite their distinct origins, *SOD1* and *TARDBP*-related mechanisms converge on common pathways involving protein misfolding, oxidative damage, mitochondrial dysfunction, impaired RNA metabolism, and deficits in proteostasis (Trist et al., 2021, 2022; Zeng et al., 2024). Understanding how these overlapping processes interact is essential for clarifying ALS pathogenesis and developing effective diagnostic and therapeutic strategies. However, several key questions still need to be answered, including the exact order of disease-causing events, why motor neurons are more vulnerable than other neuron types, and what factors contribute to the clinical differences in *SOD1* and *TARDBP*-related ALS. Solving these issues is crucial for developing a unified model of how ALS progresses based on gene-specific mechanisms. Since *SOD1* and *TARDBP* play significant roles in shaping the molecular structure of ALS, a thorough understanding of their functions is essential for discovering new biomarkers, improving disease models, and developing targeted treatments to slow down neurodegeneration caused by these proteins.

2. Methods

2.1 Ethical considerations

This study relied exclusively on computational analyses using publicly available genomic data sets and bioinformatics tools. As no human or animal subjects were involved, ethical approval and informed consent requirements did not apply to this study.

2.2 Sample approach and justification

A targeted gene-based computational workflow was employed to analyze *SOD1* and *TARDBP*, two genes closely associated with ALS. The overall workflow is shown in Fig. 1. Variant data were retrieved from the NCBI dbSNP database, which provides curated and standardized variant annotations suitable for in-silico analyses (Landrum et al., 2018).

2.3 Variant inclusion and exclusion criteria

Inclusion criteria

- SNPs located within coding regions of *SOD1* and *TARDBP*
- Nonsynonymous single-nucleotide polymorphisms (nsSNPs)
- Minor allele frequency (MAF) ≥ 0.001

Exclusion criteria

- Synonymous or intronic SNPs
- Variants with MAF < 0.001

2.4 Variant retrieval and Dataset Processing

Variant data for *SOD1* and *TARDBP* were retrieved from the NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp/) and UniProt databases. nsSNPs located within the coding regions were filtered and processed in R Studio (version 4.3.2) to remove duplicates, incomplete entries, and variants lacking functional annotations. The refined dataset included the rsID, genomic position, nucleotide substitution, and corresponding amino-acid alterations. Only variants consistently predicted as deleterious by both SIFT and PolyPhen-2 were retained for downstream analyses to ensure high confidence prioritization.

2.5 Functional impact prediction

The functional significance of nsSNPs was assessed using the Combined Annotation Dependent Depletion (CADD) scoring system (https://cadd.gs.washington.edu/), which integrates diverse genomic and functional features to estimate variant deleteriousness (Schubach et al., 2024). PHRED-like CADD scores were interpreted using standard thresholds, and only variants with scores ≥ 20 were retained for further analysis.

2.6 Pathogenicity assessment

Deleterious nsSNPs analysis using SIFT

The Sorting Intolerant From Tolerant (SIFT) tool (https://sift.bii.a-star.edu.sg/) was used to evaluate the impact of each amino acid substitutions based on evolutionary conservation (Sim et al., 2012). Variants with scores ≤ 0.05 were classified as deleterious, whereas those with scores **>** 0.05 were predicted to be tolerated.

Damaging nsSNPs prediction using PolyPhen-2

Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/) was used to predict the structural and functional consequences of amino acid substitutions using evolutionary features, comparative modelling, and physicochemical descriptors (Adzhubei et al., 2013). Scores were interpreted based on standard classification thresholds, with values from 0.00 to 0.15 considered benign, 0.15 to 0.85 classified as possibly damaging, and 0.85 to 1.00 classified as probably damaging, respectively. Variants classified as probably or possibly damaging were retained.

PhD-SNP analysis

The pathogenic potential of nonsynonymous SNPs was evaluated using the Predictor of human deleterious Single Nucleotide Polymorphisms (PhD-SNP) classifier, a Support Vector Machine (SVM) based classifier that predicts whether amino acid substitutions are disease-associated or neutral (https://snps.biofold.org/phd-snp/phd-snp.html) (Al-Ayari et al., 2023). The tool analyzes sequence-based features, evolutionary conservation, and substitution patterns to classify each variant accurately. Each prediction is accompanied by a Reliability Index (RI; 0–10), indicating the

confidence level of each prediction; higher values indicate greater reliability. These Predictions were used to differentiate harmful mutations from functionally tolerated substitutions, thereby providing additional evidence for pathogenicity and aiding in variant prioritization.

SNPs&GO analysis

The Single Nucleotide Polymorphism & Gene Ontology (SNPs&GO) integrates protein sequence information, evolutionary conservation, and Gene Ontology annotations to predict disease relevance (https://snps-and-go.biocomp.unibo.it/snps-and-go/) (Pavithran & Kumavath, 2021).This tool also provides a Reliability Index (RI) ranging from 0 to 10, where a higher value indicates greater confidence in the predictions. Variants predicted as Disease with an RI of 10 were considered highly reliable and likely pathogenic, supported by strong sequence conservation and functional context. SNP&GO predictions were used to complement other pathogenicity tools and strengthen variant prioritization.

Clinical annotation using ClinVar

Clinical relevance of the prioritized nsSNPs were retrieved from the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/), a publicly accessible repository of variant interpretations submitted by clinical laboratories, researchers, and expert groups. ClinVar annotations, including pathogenic, likely pathogenic, conflicting interpretations or not reported, were retrieved for all variants with available rsIDs. ClinVar was used solely for clinical annotation and contextual interpretation and not as an *in-silico* pathogenicity prediction tool, as the database contains user-submitted clinical classifications rather than standardized algorithmic predictions.

2.7 Protein stability analysis

I-Mutant 2.0 analysis

Protein stability changes induced by amino acid substitutions were assessed using I-Mutant 2.0, a sequence-based prediction tool that estimates mutation-induced variations in Gibbs free energy (ΔΔG) (https://folding.biofold.org/i-mutant/i-mutant2.0.html). It classifies substitutions as stabilizing or destabilizing relative to the wild-type protein, with negative ΔΔG values indicating reduced thermodynamic stability and positive values suggesting stabilization (Al-nakhle & Khateb, 2023). Consistent with previously reported in silico variant characterization studies, I-Mutant 2.0 predictions were interpreted with consideration of model-specific performance behaviour across diverse mutation datasets.

MUpro analysis

Protein stability changes were further evaluated using MUpro, a machine-learning–based method that predicts mutation-induced free-energy differences (https://mupro.proteomics.ics.uci.edu). MUpro classifies substitutions as stabilizing or destabilizing on the basis of sequence-derived features and has been widely used in recent structural analyses of deleterious missense variants (Ali et al., 2024). Variants predicted to decrease stability by MUpro were considered likely to contribute to structural disruption.

DynaMut-2 analysis

The effect of missense variants on protein stability and conformational dynamics was evaluated using DynaMut-2 (https://biosig.lab.uq.edu.au/dynamut2/), which integrates normal mode analysis, graph-based signatures, and machine-learning algorithm. The tool predicts mutation-induced changes in Gibbs free energy (ΔΔG) and protein flexibility by assessing alterations in intramolecular interactions and atomic fluctuations. Negative ΔΔG values ($\Delta\Delta G < 0$) indicate destabilizing mutations, whereas positive values ($\Delta\Delta G > 0$) indicate stabilizing effects relative to the wild-

type protein (Kamal et al., 2024). Variants predicted to induce destabilization or marked flexibility changes were prioritized for downstream structural and functional analyses.

INPS-MD Analysis

The effects of missense variants on protein stability were further evaluated using Impact of Non-synonymous Protein mutations–Molecular Dynamics (INPS-MD), a computational tool which predicts mutation-induced stability changes by combining sequence, structural, and dynamic features (https://inpsmd.biocomp.unibo.it/), which estimates changes in Gibbs free energy (ΔΔG), where positive values (ΔΔG > 0) indicate stabilizing mutations and negative values (ΔΔG < 0) indicate destabilizing effects (Alam et al., 2025). The method combines residue environment, solvent accessibility, and interaction networks to evaluate mutation-driven stability alterations. INPS-MD predictions were used to complement other stability tools and support prioritization of variants with potential structural and energetic impacts.

2.8 Post-translational modification analysis

Post-translational modification (PTM) alterations were evaluated using the Group-based Prediction System-Methylation sited prediction (GPS-MSP) platform (http://gps.biocuckoo.cn/online.php), which predicts kinase-specific phosphorylation sites based on sequence patterns and motif signatures. This machine-learning–based approach follows current PTM prediction practices and enables the identification of mutations that may disrupt regulatory modification sites (Meng et al., 2022), thereby providing additional insight into how variants may alter protein regulation, signalling, or stability particularly for proteins involved in neurodegenerative pathways.

2.9 Secondary structure analysis

GOR4 (Garnier-Osguthorpe-Robson)

Protein secondary structure was predicted using the GOR4 (Garnier–Osguthorpe–Robson) algorithm, which assigns α-helices, β-strands, and random coils based on information-theoretic analysis of amino acid composition and positional context within the sequence (https://npsa-prabi.ibcp.fr/NPSA/npsa_gor4.html). GOR4 evaluates residue propensities and neighboring interactions to estimate secondary structure probabilities, enabling assessment of mutation-induced local structural organization (Xia et al., 2011). Predicted secondary structure profiles of wild-type and mutant proteins were compared to identify conformational changes associated with nonsynonymous substitutions.

SOPMA

Secondary structure elements were also predicted using Self-Optimized Prediction Method with Alignment (SOPMA), which estimates α-helices, β-strands, turns, and coils based on residue propensity scores (https://npsa-pbil.ibcp.fr/NPSA/npsa_sopma.html). SOPMA enhances prediction accuracy by incorporating multiple sequence alignments and iterative refinement, enabling accurate detection of mutation-induced shifts in secondary structure composition (Mukherjee et al., 2022). The secondary structure profiles of wild-type and mutant sequences were compared to identify local conformational changes.

2.10 Three-dimensional structural modeling

SWISS-MODEL

Three-dimensional protein structures of selected variants were generated using the SWISS-MODEL homology modeling platform (https://swissmodel.expasy.org/) through template identification, alignment, and energy-optimized model construction (Waterhouse et al., 2018). Models with the highest Global Model Quality Estimation (GMQE) scores were selected. And structural validation was performed using Ramachandran plot analysis and MolProbity evaluation,

confirming acceptable stereochemical value, favorable MolProbity scores, and minimal steric clashes, supporting the reliability of the predicted models for downstream functional analysis.

2.11 Expression, interaction, and functional analysis

Human Protein Atlas (HPA)

Protein and transcript expression patterns of the target genes were analyzed using the Human Protein Atlas (HPA) database across multiple human brain regions. HPA provides region-specific RNA expression and immunohistochemistry-based protein expression data, enabling evaluation of spatial expression patterns within the central nervous system (https://www.proteinatlas.org/) (Mohamed et al., 2025). Expression profiles across different brain regions were analyzed to evaluate regional variability and support the biological relevance of the analyzed genes in neurological contexts. This strategy aligns with established in-silico methodologies that integrate variant analyses with tissue-specific expression data to strengthen functional interpretation.

STRING

Protein–protein interaction (PPI) networks for *SOD1* and *TARDBP* were constructed using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database, which integrates known and predicted interactions from experimental evidence, computational predictions, co-expression, and curated pathway databases (https://string-db.org/) (Crosara et al., 2018). STRING was used to identify direct and indirect functional associations and to visualize interaction networks based on confidence scores. Also, it incorporates scoring algorithms and expanded protein–protein association datasets were utilized to support comprehensive interaction mapping relevant to disease mechanisms (Szklarczyk et al., 2023).

KEGG and ShinyGO

Functional pathway and enrichment analyses were performed using the KEGG database (https://www.kegg.jp) and the ShinyGO platform (https://bioinformatics.sdstate.edu/go/). KEGG was used to map the identified gene set to curate metabolic and signaling pathways (Kanehisa et al., 2023). To complement this analysis, the curated gene list was uploaded to ShinyGO for Gene Ontology (GO) enrichment and pathway interpretation, which identifies significantly enriched biological processes, molecular functions, cellular components, and pathway categories, including KEGG pathways (Karunakara et al., 2024) .

Cytoscape

Cytoscape analysis was done for detailed network visualization and topological analysis of ALS-associated genes. Cytoscape was used to visualize the interaction structure and to calculate key network parameters, including degree centrality, betweenness centrality, closeness centrality, and clustering coefficient, which facilitate the identification of hub nodes and highly influential proteins within the network. These topological metrics enable a systematic assessment of the network's organization and the functional significance of individual proteins (Majeed & Mukhtar, 2023).

Gene Mania

GeneMANIA was used to construct a gene-gene interaction and functional association network for *SOD1* and *TARDBP* with ALS-associated genes (https://genemania.org/). Gene-MANIA combines various data types, including co-expression, physical interactions, genetic interactions, co-localization, pathways, and shared protein domains, to find genes that are functionally related to the query. The tool was applied to expand the interaction map and identify

additional genes that may be involved in shared biological processes or disease pathways (Irfan et al., 2022). This method enabled the functional contextualization of the genes by highlighting biologically relevant interaction partners.

3. Results

3.1 Identification and prioritization of nsSNPs in *SOD1* and *TARDBP*

A comprehensive database mining of the dbSNP repository yielded 5,964 variants in *SOD1* and 10,630 variants in *TARDBP* genes. Among these, 226 *SOD1* and 742 *TARDBP* variants were categorized as missense (nonsynonymous) SNPs, suggesting possible amino acid substitutions with functional significance. After applying stringent inclusion criteria-coding region localization, nonsynonymous effect, and minor allele frequency (MAF ≥ 0.001) refined the dataset to 12 nsSNPs per gene **(Table S1)**. These variants were prioritized using CADD and subsequently assessed using SIFT, PolyPhen-2, PhD-SNP, and SNPs&GO for in-silico functional and pathogenicity analyses.

3.2 CADD functional impact assessment

CADD analysis revealed multiple *SOD1* and *TARDBP* variants associated with ALS showing high deleterious potential, characterised by PHRED scores ≥ 20. In *SOD1*, three substitution events corresponding to two unique SNPs (rs1202989817 and rs1182088847) were analyzed. The rs1202989817 variant carried two alternate alleles (T > C and T > G), which were analyzed separately due to distinct CADD scores. These substitutions exhibited PHRED scores ranging from 23.0 to 28.0, indicating moderate to strong predicted functional impact. In *TARDBP*, six substitutions exhibited PHRED scores between 22.6 to 34.0, with the *T > G* substitution displaying the highest score, suggesting a pronounced functional effect **(**Table 1**)**. Overall, the CADD results indicate that prioritized variants in both genes are likely to exert substantial functional constraints, supporting their potential relevance in ALS-associated molecular mechanisms

Table 1
High-impact variants with CADD PHRED scores ≥ 20.

Gene	rsIDs	Ref	Alt	Raw Score	PHRED
SOD1	rs1202989817	T	C	4.504	25.5
	rs1202989817	T	G	5.010	28.0
	rs1182088847	C	G	3.601	23.0
TARDBP	rs1228733743	T	C	4.371	25.1
	rs1228733743	T	G	6.461	34.0
	rs80356715	C	G	3.478	22.6
	rs80356715	C	T	3.654	23.2
	rs1570722030	T	C	5.125	28.7
	rs80356729	G	A	3.849	23.6

3.3 *In silico* evaluation of functional impact and disease association of prioritized variants

Functional impact of prioritized *SOD1* and *TARDBP* variants was evaluated using SIFT and PolyPhen-2. Several variants in both genes were predicted as deleterious by SIFT and as probably damaging by PolyPhen-2, indicating potential effects on protein structure and function **(**Table 2**).**

Disease relevance was further assessed using PhD-SNP and SNPs&GO. In *SOD1*, the variants rs1202989817 (V15G) and rs1182088847 (I19M) were consistently classified as deleterious (0.02 and 0.01) by SIFT and probably damaging (1.000 and 0.985) by PolyPhen-2. Both variants were additionally classified as disease-associated by PhD-SNP and by SNPs&GO. In contrast, *TARDBP* variants rs80356729 (G335D) and rs1570722030 (I222T) were classified as deleterious (0.00) by SIFT and probably damaging (0.999-1.000) by PolyPhen-2 however were predicted as neutral by both PhD-SNP and SNPs&GO **(**Table 2**).**

Table 2
Consensus pathogenicity predictions for high-confidence variants.

Gene	RSID	Mutation	Tool	ΔΔG (kcal/mol)	Prediction (per tool)
SOD1	rs1202989817	V15G	SIFT	0.02	Deleterious
			PolyPhen-2	1.000	Probably damaging
			PhD-SNP	8	Disease
			SNPs&GO	10	Disease
	rs1182088847	I19M	SIFT	0.01	Deleterious
			PolyPhen-2	0.985	Probably damaging
			PhD-SNP	5	Disease
			SNPs&GO	10	Disease
TARDBP	rs80356729	G335D	SIFT	0.00	Deleterious
			PolyPhen-2	0.999	Probably damaging
			PhD-SNP	3	Neutral
			SNPs&GO	9	Neutral
	rs1570722030	I222T	SIFT	0.00	Deleterious
			PolyPhen-2	1.000	Probably damaging
			PhD-SNP	1	Neutral
			SNPs&GO	5	Neutral

3.4 ClinVar-based clinical annotation of prioritized variants

The clinical annotations for prioritized *SOD1* and *TARDBP* variants were obtained from the ClinVar database, which compiles user-submitted variant interpretations. Among the evaluated *SOD1* variants, the T > G substitution of rs1202989817 (p.Val15Gly) was classified as likely pathogenic and associated with ALS, based on a germline submission from a clinical diagnostic laboratory. In contrast, the T > C alternate allele of rs1202989817 and rs1182088847 (C > G) were not reported in ClinVar **(Table S2)**, indicating the absence of current clinical classification for these substitutions.

For *TARDBP*, most prioritized variants, including alternate alleles of rs1228733743, rs80356715, and rs1570722030, were not reported in ClinVar. However, the G > A substitution of rs80356729 exhibited conflicting interpretations of pathogenicity in ClinVar **(Table S2)**, with clinical submissions reporting associations with ALS but lacking a consensus classification.

3.5 Protein stability prediction

Protein stability analysis using I-Mutant 2.0, MUpro, DynaMut-2, and INPS-MD consistently demonstrated destabilizing effects for the *SOD1* variants (V15G and I19M) (Table 3). All tools produced negative ΔΔG values for both substitutions, indicating reduced structural stability. For *TARDBP*, the rs1570722030 (I222T) variant also exhibited consistent destabilization across all prediction tools, returning negative ΔΔG values (Table 3). In contrast, rs80356729 (G335D) demonstrated mixed predictions; MUpro, DynaMut-2, and INPS-MD predicted destabilization, whereas I-Mutant 2.0 predicted marginal stabilization (ΔΔG = + 0.15 kcal/mol), suggesting a limited or context-dependent stability effect.

Overall, the stability predictions indicate a consistent loss of stability for the *SOD1* variants and the *TARDBP* I222T substitution, while the G335D variant exhibits variable stability outcomes across tools (Table 3).

Table 3
Protein stability predictions for consensus variants.

Gene	RSID	Mutation	Tool	ΔΔG (kcal/mol)	Prediction (per tool)	Consensus Interpretation
SOD1	rs1202989817	V15G	MUpro	-1.149	Decrease Stability	Strong destabilizing effect (all tools consistently predict loss of stability)
			I-Mutant 2.0	-2.56	Decrease Stability	
			DynaMut-2	-2.7	Destabilizing	
			INPS-MD	-3.55	Destabilizing	
	rs1182088847	I19M	MUpro	-0.866	Decrease Stability	Consistently destabilizing across all tools
			I-Mutant 2.0	-0.38	Decrease Stability	
			DynaMut-2	-0.14	Destabilizing	
			INPS-MD	-0.89	Weakly destabilizing	
TARDBP	rs80356729	G335D	MUpro	-0.951	Decrease Stability	Predominantly destabilizing, but I-Mutant 2.0 suggests slight stabilization.
			I-Mutant 2.0	0.15	Increase Stability	
			DynaMut-2	-0.09	Destabilizing	
			INPS-MD	-0.66	Weakly destabilizing	
	rs1570722030	I222T	MUpro	-2.132	Decrease Stability	Strong destabilizing effect (all tools consistently predict loss of stability)
			I-Mutant 2.0	-3.82	Decrease Stability	
			DynaMut-2	-1.62	Destabilizing	
			INPS-MD	-2.66	Destabilizing	

3.6 Post-translational modification analysis

Post-translational modification analysis using GPS-MSP 6.0 server identified several conserved phosphorylation residues in both *SOD1* and *TARDBP* **(Table S3)**. In *SOD1*, the variants rs1202989817 (V15G) and rs1182088847 (I19M) retained all predicted phosphorylation sites observed in the wild-type protein, including T3, T59, S60, and T89, which were primarily associated with CK1 and TKL kinases. Phosphorylation scores and residue positions were preserved across wild-type and variant sequences. A similar pattern was observed for *TARDBP*. Both rs80356729 (G335D) and rs1570722030 (I222T) exhibited conserved phosphorylation sites distributed across serine and tyrosine residues. High-confidence tyrosine sites (Y4, Y73, Y123, and Y374) and multiple serine residues (S92, S273, S393, and S403) were consistently predicted across all *TARDBP* variants. No gain or loss of major phosphorylation sites was observed relative to the wild-type protein **(Table S3)**. Overall, PTM analysis indicated that the prioritized *SOD1* and *TARDBP* variants do not alter predicted phosphorylation motifs, with conserved kinase recognition patterns maintained across wild-type and mutant sequences.

3.7 Secondary structure impact assessment

Secondary structure prediction using GOR4 and SOPMA revealed that prioritized variants in *SOD1* and *TARDBP* introduced only minor, localized alterations without disrupting global secondary-structure organization **(**Table 4**)**. In *SOD1*, the V15G variant showed a moderate increase in α-helical content with a corresponding decrease in extended strands, while I19M exhibited smaller shifts of similar direction.

In *TARDBP*, G335D showed no detectable changes in secondary-structure elements, whereas I222T demonstrated minor shifts involving slight reductions in α-helix and marginal increases in extended strand and coil regions **(**Table 4**)**. Complementary SOPMA analysis supported these findings, confirming that both *SOD1* and *TARDBP* variants largely preserve overall secondary structure profiles **(Figure S2)**. In *SOD1* (V15G) variant showed a modest increase in α-helix and extended strand, while (I19M) displayed small increases in α-helix, extended strand, and random coil. Meanwhile *TARDBP* variants G335D and I222T showed only negligible changes (≤ 0.25%) across structural elements **(Figures S1-S2)**

Table 4
Predicted impact of nsSNPs on the secondary structure of the *SOD1* and *TARDBP* proteins compared to wild-type.

Gene	Structure Element	GOR4 Wild-Type	GOR4 Variant	GOR4 Change (%)	SOPMA Wild-Type	SOPMA Variant	SOPMA Change (%)
SOD1 (rs1202989817) V15G	α-helix	0(0.00%)	6(3.90%)	3.90	0(0.0%)	7(4.55%)	4.55
	Extended strand	55(35.71%)	49(31.82%)	-3.89	46(29.87%)	47(30.52%)	0.65
	Random coil	99(64.29%)	99(64.29%)	0	100(64.94%)	100(64.94%)	0
SOD1 (rs1182088847) I19M	α-helix	0(0.00%)	3(1.95%)	1.95	0(0.0%	5(3.25%)	3.25
	Extended strand	55(35.71%)	52(33.77%)	-1.94	46(29.87%)	47(30.52%)	0.65
	Random coil	99(64.29%)	99(64.29%)	0	100(64.94%)	102(66.23%)	1.29
TARDBP (rs80356729) G335D	α-helix	77(18.60%)	77(18.60%)	0	54(13.04%)	55(13.29%)	0.25
	Extended strand	96(23.19%)	96(23.19%)	0	62(14.98)	63(15.22%)	0.24
	Random coil	241(58.21%)	241(58.21%)	0	298(71.98%)	296(71.50%)	-0.48
TARDBP (rs1570722030) I222T	α-helix	77(18.60%)	72(17.39%)	-1.21	54(13.04%)	54(13.04%)	0
	Extended strand	96(23.19%)	100(24.15)	0.96	62(14.98)	61(14.73%)	-0.25
	Random coil	241(58.21%)	242(58.45%)	0.24	298(71.98%)	299(72.22%)	0.24

3.8 Three-dimensional structural modeling

SWISS MODEL

Homology-based structural modeling using SWISS-MODEL generated high-quality three-dimensional models for all variants **(**Fig. 2**)**. The *SOD1* (V15G and I19M) variants showed low MolProbity scores (0.62–0.71), zero clash scores, and > 96% residues in favored Ramachandran regions **(**Fig. 3**)**, indicating high stereochemical accuracy **(Table S4)**.

In contrast, *TARDBP* (G335D and I222T) variants exhibited higher MolProbity scores (1.73–1.76) and reduced percentages of favored Ramachandran residues (~ 74%) **(Table S4)**, with 12% outliers, suggesting localized conformational strain. Despite these differences, the global architecture of all models remained preserved **(**Fig. 2**)**.

3.9 Expression profiling and interaction network analysis

Analysis of the Human Protein Atlas (HPA) brain RNA-seq dataset showed that *SOD1* and *TARDBP* are broadly expressed across major human brain regions. *SOD1* exhibited higher transcript abundance in the cerebral cortex, hippocampus, thalamus, hypothalamus, cerebellum, and pons **(Figure S3.A)**. Similarly, *TARDBP* exhibited a similarly widespread expression pattern, with relatively higher levels in the cerebral cortex, hippocampus, cerebellum, and white matter **(Figure S3.B)**. These expression profiles indicate that both genes are constantly expressed in brain regions associated with motor coordination, synaptic regulation and cognition.

4. Functional enrichment and network analysis

Functional enrichment analysis performed using STRING revealed that interaction partners of *SOD1* and *TARDBP* are significantly enriched in biological processes significant to neurodegeneration. The protein–protein interaction (PPI) map formed a dense and highly interconnected cluster, reflecting strong functional coupling among ALS-associated genes **(**Fig. 4A**)**. The network enrichment was highly significant (FDR < 1.0×10^{-16}), indicating non-random biological associations.

Gene Ontology biological process (GO-BP) analysis demonstrated enrichment for pathways involved in intracellular ion regulation, calcium-dependent signaling, and cytoskeletal organization. Highly enriched terms included regulation of calcium ion transport, voltage-gated calcium channel activity, calcineurin-NAFT signaling, and neurofilament organization **(**Fig. 4B**)** The strongest enrichments corresponded to processes with large gene representation and low .

Gene Ontology molecular function (GO-MF) analysis highlighted enrichment of protein–protein interaction domains and enzymatic activities, including calcium-dependent phosphatase activity, calmodulin binding, and structural components of the intermediate filament cytoskeleton **(**Fig. 4C**)**.

Gene Ontology cellular component (GO-CC) analysis revealed that interacting proteins localize predominantly to synapses, membrane microdomains, transcriptional regulatory complexes, and Wnt signalosome **(**Fig. 4D**)**.

4.1 KEGG pathway integration and disease-level signatures

KEGG pathway analysis utilizing ShinyGO extended the STRING-derived functional enrichment into a disease-specific framework. The ALS pathway appeared as the most significantly enriched module, with numerous genes from the *SOD1-TARDBP* interaction network mapping directly to key pathogenic nodes **(Figure S4.A)**. Highlighted genes included *SOD1, TARDBP, FUS, VCP, OPTN, C9orf72, CHCHD10, and NEK1*, which were spread across diverse ALS-related pathways involving excitotoxicity, mitochondrial dysfunction, proteostasis disruption, axonal transport, and RNA-binding protein aggregation.

Complementary KEGG over-representation analysis identified ALS as the top-enriched pathway, exhibiting the highest fold enrichment and statistical significance ($-\log_{10}P > 20$) **(Figure S4.B)**. Pathways associated with neurodegeneration were also significantly enriched, alongside neuronal signaling processes such as VEGF signaling, calcium signaling, glutamatergic synapse regulation, axon guidance, MAPK signaling, and Wnt signaling. Additional enriched pathways included autophagy-animal, mitophagy, mRNA surveillance, ferroptosis, peroxisome biology, and spliceosome-related processes.

Overall, KEGG analyses indicated that the *SOD1-TARDBP* interaction network aligns with established ALS mechanisms while intersecting with broader neurodegenerative pathways.

4.2 Network construction and topological analysis

Network topology analysis using Cytoscape was performed to quantify the structural significance of *SOD1* and *TARDBP* within the ALS-associated interaction network (Fig. 5**)**. The resulting network revealed a highly connected structures centered on two major hubs, *TARDBP* and *SOD1*, with several secondary nodes forming tightly clustered ALS modules. The topology parameters are summarized in (**Table S5)**. *TARDBP* emerged as the most influential node, exhibiting the highest degree (20), lowest average shortest path length (1.048), and highest closeness centrality (0.955). *SOD1* ranked as the second major hub, with a degree of 19, high closeness centrality (0.913), and notable betweenness centrality (0.227), indicating its prominent position within the network.

A secondary tier of highly connected nodes, including *C9orf72, FUS, NEFH, and SETX*, demonstrated substantial centrality (degree: 11–14) and high clustering coefficients (0.648–0.891), indicating densely interconnected ALS-associated sub-networks. Additional clustered nodes, such as *ALS2, DCTN1*, Fig. 4, *OPTN, and VAPB* (degree = 10), exhibited clustering coefficient of 1.0, consistent with tightly linked modules. In contrast, nodes such as *TMEM106B, RNF19A, PRNP, SMN1*, and *GSC* exhibited lower degree values and centrality values, indicating more peripheral network positions.

4.3 Gene-Gene Interaction Network Analysis

Gene–gene interaction analysis using GeneMANIA was performed to further characterize the functional landscape surrounding *SOD1* and *TARDBP*. The integrated network comprised 35 nodes connected by 666 functional links, reflecting strong connectivity among ALS-associated genes **(Figure S5)**,with most edges contributed by co-expression data, followed by physical interactions, shared pathways, and genetic interactions.

The network exhibited a clustered architecture with several coordinated biological modules. Key RNA-binding proteins, including *TARDBP, FUS, SETX*, and *TIA1*, clustered centrally. In parallel autophagy and proteostasis-related genes such as *OPTN, TBK1*, *VCP*, and *UBQLN2* formed a second major hub. Additional submodules included genes associated with mitochondrial function (*CHCHD10, NEK1*) and vesicular trafficking (*C9orf72, VAPB*).

Hub gene identification

Analysis of node degree and centrality metrics identified several highly connected hubs within the network. Genes with the greatest connectivity included *UBB/UBC, OPTN, VCP, TBK1, TARDBP*, and *SOD1*, indicating their prominent positions within the ALS-associated interaction network **(Table S6)**.

4.4 Functional enrichment and gene prioritization

Functional enrichment of the *SOD1-TARDBP* network using GeneMANIA revealed several core biological pathways strongly aligned with ALS progression. Dominant functional categories include autophagy and protein degradation (*OPTN, TBK1, VCP, UBQLN2*), RNA processing and splicing (*TARDBP, FUS, SETX, TIA1*), oxidative stress response (*SOD1, FTH1*), vesicular trafficking (*VAPB*, *GRB7*), and mitochondrial organization (*CHCHD10, NEK1*) **(Table S7)**.

Gene prioritization based on weighted GeneMANIA highest scores identified *C9orf72* (0.756), *NEK1* (0.727), *CHCHD10* (0.697), *VAPB* (0.671), and *SETX* (0.669) as the highest-ranking genes within the network **(Table S8)**. Additional centrally positioned genes included *TIA1*, *VCP, ANG, FUS, OPTN*, and *TBK1*, while lower-ranked genes such as *SNX6* and *IRAK2* contributed meaning interactions. Overall, enrichment and prioritization analyses indicate that the *SOD1-TARDBP* interaction network is organized around interconnected modules related to proteostasis, RNA regulation, mitochondrial biology, and vesicular processes.

Discussion

This in-silico study systematically evaluated nonsynonymous SNPs in *SOD1* and *TARDBP*, identifying high-confidence deleterious variants with the potential to influence ALS pathogenesis. Where two *SOD1* variants (rs1202989817 and rs1182088847) exhibited strong concordance for pathogenic predictions, strong destabilization, and localized secondary structure shift, whereas *TARDBP* variants showed partial variance across disease-association classifiers, with variant (I222T) demonstrated notable destabilizing potential. Structural modeling indicated preservation of global fold for all variants but revealed localized conformational changes. Network analyses placed *SOD1* and *TARDBP* as central nodes within ALS-associated regulatory hubs, linking RNA metabolism, proteostasis, mitochondrial function, and axonal transport pathways.

The *SOD1* variants V15G and I19M showed high deleteriousness predictions and understanding among independent computational tools, consistent with evidence that early exon *SOD1* variants often carry strong pathogenic signatures and correlate with ALS phenotypes (Ruffo et al., 2022). Stability prediction showed consistent thermodynamic destabilization of the *SOD1* variants V15G and I19M, consistent with previous reports classifying V15G as likely pathogenic due to recurrent destabilizing signatures, low population frequency, and localization within a structural hotspot (Ruffo et al., 2022). Such destabilization aligns with established mechanisms in which mutation-induced instability increases misfolding, impaired metal coordination, and aggregation of *SOD1* (Taylor et al., 2016; Trist et al., 2021). In contrast, *TARDBP* variants showed stronger deleterious agreement among sequence-based functional predictors than among clinical classification tools. Although G335D and I222T were predicted to exert biochemical impact, only I222T showed consistent destabilization across stability models. Such variance reflects recognized challenges in *TARDBP* variant interpretation, where biochemical disruption does not always translate into clear clinical classification (Balendra & Isaacs, 2018). Structural assessment suggested preserved global architecture but increased local conformational strain, particularly relevant given the intrinsically dynamic nature of *TARDBP*. Mutations within the low-complexity domain have been shown to alter phase behaviour and aggregation propensity without inducing major structural collapse (Conicella et al., 2016; Balendra & Isaacs, 2018). The predicted local instability observed here is therefore compatible with mechanisms involving changed liquid–liquid phase separation and aberrant aggregation.

Post-translational modification predictions indicated conservation of major phosphorylation motifs throughout variants, suggesting that pathogenic effects are unlikely to arise from direct disruption of phosphorylation sites. This observation aligns with earlier findings that *SOD1* mutations typically impair folding stability and metal binding rather than PTM motifs (Ruffo et al., 2022), and that *TARDBP* aggregation is more closely linked to conformational instability and dysregulated phase behaviour than to phosphorylation loss (Balendra & Isaacs, 2018). Secondary-structure analysis using GOR4 and SOPMA indicated localized increases in α-helical content for *SOD1* V15G and I19M, whereas *TARDBP* variants G335D and I222T exhibited minimal changes. These modest shifts are consistent with ALS-associated mutations that alter local folding dynamics without disrupting global structure (Trist et al., 2021).Variants within the *TARDBP* low-complexity domain are known to influence aggregation propensity rather than induce major structural rearrangements (Conicella et al., 2016).Network-level analyses provided additional context, where STRING and Cytoscape topology analyses identified *TARDBP* and *SOD1* as highly connected hubs within ALS-related pathways. Secondary clusters included genes such as *FUS, C9orf72, OPTN, TBK1, VCP, SETX, CHCHD10*, and *NEK1*, which participate in RNA metabolism, autophagy, mitochondrial function, and vesicular trafficking. KEGG enrichment confirmed ALS as the most significantly represented pathway and highlighted convergence with calcium signaling, autophagy, mitophagy, and RNA processing. These findings support the view that ALS arises from coordinated dysfunction across interconnected cellular systems rather than a single molecular defect (Goutman et al., 2022; Taylor et al., 2016). Thus, even modestly destabilizing variants in *SOD1* and *TARDBP* can propagate dysfunction across multiple cellular systems.

Several limitations should also be considered. All results are based on computational predictions and homology modeling, which cannot fully replace for experimental validation. Stability predictions represent thermodynamic approximations and may not capture intracellular dynamics. The modeling of *TARDBP* is constrained by intrinsically disordered regions, reducing confidence in predicted conformational predictions. Also network analyses reflect predicted or literature-derived associations rather than neuronal interactions, and population-genetic integration was not performed. Future studies incorporating molecular dynamics simulations, biochemical aggregation assays, liquid–liquid phase separation experiments, and iPSC-derived motor neuron systems will be required to validate the functional consequences of these variants. Genome-editing approaches such as CRISPR knock-in models could further clarify physiological relevance under endogenous expression conditions.

Despite these limitations, the findings have significant clinical implications. The destabilizing *SOD1* variants identified here may represent candidates for targeted therapeutic strategies, including antisense approaches aimed at reducing toxic SOD1 protein levels. Likewise, *TARDBP* variants predicted to influence conformational stability and phase behaviour may inform development of small molecules that modulate *TDP-43* aggregation and restore RNA metabolism. Additionally network analysis highlighted hub genes such as *OPTN, TBK1, VCP*, and *C9orf72*, which may serve as biomarkers or therapeutic targets within pathway-oriented intervention strategies. Together, these integrative results provide a framework for prioritizing variants and pathways for experimental validation and therapeutic targets in ALS.

Conclusion

This in-silico analysis systematically evaluated nonsynonymous variants in *SOD1* and *TARDBP* to assess their potential relevance in amyotrophic lateral sclerosis. Two *SOD1* variants, V15G and I19M, demonstrated strong consensus for pathogenicity, consistent thermodynamic destabilization, and structural features aligned with established mechanisms of *SOD1* misfolding and aggregation. In contrast, *TARDBP* variants exhibited variable pathogenicity predictions but showed structural and network-level characteristics consistent with altered conformational stability and RNA-related dysfunction. The Functional enrichment and interaction analyses further underscored *SOD1* and *TARDBP* as central nodes within ALS-related pathways involving proteostasis, autophagy, mitochondrial integrity, and RNA metabolism. Although the findings are based on computational modessling and require experimental validation, this integrative approach prioritizes structurally and network-relevant variants for future investigation. Overall, the study refines the molecular understanding of ALS-associated variation in *SOD1* and *TARDBP* and highlights candidate variants and pathways that may inform mechanistic studies and therapeutic intervention.

Declarations

Competing interests

The authors declare that they have no financial interests or personal conflicts that could have influenced the findings reported in this paper.

Funding

This work was supported by Indian Council of Medical Research (Sanction number. 54/8/GER/2019-NCD-II), DBT-BUILDER (BT/INF/22/SP43065/2021), Govt. of India, Manipal Research Board (MRB) Grant, and MAHE Seed Money Grant.

Author Contribution

MBM, SM, performed the analysis, drafted the manuscript and prepared the figures. SP conceptualized, edited, and revised the manuscript.

Data Availability

All Data are included in the manuscript and supplementary information.

References

1. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protocols Hum. Genet.* **76** (1). https://doi.org/10.1002/0471142905.hg0720s76 (2013).
2. Alam, S. S. M., Samanta, A., Ali, S. & Hoque, M. Structural insights into the impacts of non-synonymous single nucleotide polymorphisms in CD274 gene on the PD-1/PD-L1 interaction: An in silico approach. *Biochem. Biophys. Res. Commun.* **784**, 152679. https://doi.org/10.1016/j.bbrc.2025.152679 (2025).
3. Al-Ayari, E. A., Shehata, M. G., EL-Hadidi, M. & Shaalan, M. G. silico SNP prediction of selected protein orthologues in insect models for Alzheimer's, Parkinson's, and Huntington's diseases. *Sci. Rep.* **13** (1), 18986. https://doi.org/10.1038/s41598-023-46250-5 (2023).
4. Ali, E. W. et al. Exploring the Structural and Functional Consequences of Deleterious Missense Nonsynonymous SNPs in the EPOR Gene: A Computational Approach. *J. Personalized Med.* **14** (11), 1111. https://doi.org/10.3390/jpm14111111 (2024).
5. Al-nakhle, H. H. & Khateb, A. M. Comprehensive In Silico Characterization of the Coding and Non-Coding SNPs in Human Dectin-1 Gene with the Potential of High-Risk Pathogenicity Associated with Fungal Infections. *Diagnostics* **13** (10), 1785. https://doi.org/10.3390/diagnostics13101785 (2023).
6. Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Reviews Neurol.* **14** (9), 544–558. https://doi.org/10.1038/s41582-018-0047-2 (2018).
7. Balendra, R. et al. Amyotrophic lateral sclerosis caused by TARDBP mutations: from genetics to TDP-43 proteinopathy. *Lancet Neurol.* **24** (5), 456–470. https://doi.org/10.1016/S1474-4422(25)00109-7 (2025).
8. Berdyński, M. et al. SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. *Sci. Rep.* **12** (1), 103. https://doi.org/10.1038/s41598-021-03891-8 (2022).
9. Berdyński, M., Safranow, K., Andersen, P. M. & Żekanowski, C. Phenotypic Characterization of ALS-Causing SOD1 Mutations Affecting Polypeptide Length. Human Mutation, 2025(1). (2025). https://doi.org/10.1155/humu/9792233
10. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24** (9), 1537–1549. https://doi.org/10.1016/j.str.2016.07.007 (2016).
11. Crosara, K. T. B., Moffa, E. B., Xiao, Y. & Siqueira, W. L. Merging in-silico and in vitro salivary protein complex partners using the STRING database: A tutorial. *J. Proteom.* **171**, 87–94. https://doi.org/10.1016/j.jprot.2017.08.002 (2018).
12. Dash, B. P., Freischmidt, A., Weishaupt, J. H. & Hermann, A. Downstream Effects of Mutations in SOD1 and TARDBP Converge on Gene Expression Impairment in Patient-Derived Motor Neurons. *Int. J. Mol. Sci.* **23** (17), 9652. https://doi.org/10.3390/ijms23179652 (2022).
13. de Boer, E. M. J. et al. TDP-43 proteinopathies: a new wave of neurodegenerative diseases. *J. Neurol. Neurosurg. Psychiatry*. **92** (1), 86–95. https://doi.org/10.1136/jnnp-2020-322983 (2021).
14. Feneberg, E. et al. TDP-43 pathology is associated with divergent protein profiles in ALS brain and spinal cord. *Acta Neuropathol. Commun.* **13** (1), 175. https://doi.org/10.1186/s40478-025-02084-y (2025).
15. Ganssauge, J. et al. Rapid and inducible mislocalization of endogenous TDP43 in a novel human model of amyotrophic lateral sclerosis. *ELife* 13. https://doi.org/10.7554/eLife.95062 (2025).
16. Glashutter, M., Wijesinghe, P. & Matsubara, J. A. TDP-43 as a potential retinal biomarker for neurodegenerative diseases. *Front. NeuroSci.* **19** https://doi.org/10.3389/fnins.2025.1533045 (2025).

17. Goutman, S. A. et al. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol.* **21** (5), 465–479. https://doi.org/10.1016/S1474-4422(21)00414-2 (2022).
18. Huang, M., Liu, Y. U., Yao, X., Qin, D. & Su, H. Variability in SOD1-associated amyotrophic lateral sclerosis: geographic patterns, clinical heterogeneity, molecular alterations, and therapeutic implications. *Translational Neurodegeneration*. **13** (1), 28. https://doi.org/10.1186/s40035-024-00416-x (2024).
19. Irfan, M., Iqbal, T., Hashmi, S., Ghani, U. & Bhatti, A. Insilico prediction and functional analysis of nonsynonymous SNPs in human CTLA4 gene. *Sci. Rep.* **12** (1), 20441. https://doi.org/10.1038/s41598-022-24699-0 (2022).
20. Kamal, M. M. et al. In silico functional, structural and pathogenicity analysis of missense single nucleotide polymorphisms in human MCM6 gene. *Sci. Rep.* **14** (1), 11607. https://doi.org/10.1038/s41598-024-62299-2 (2024).
21. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51** (D1), D587–D592. https://doi.org/10.1093/nar/gkac963 (2023).
22. Karunakara, S. H. et al. Analysis of miR-497/195 cluster identifies new therapeutic targets in cervical cancer. *BMC Res. Notes*. **17** (1), 217. https://doi.org/10.1186/s13104-024-06876-8 (2024).
23. Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S.,Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla,Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., … Maglott, D. R.(2018). ClinVar: improving access to variant interpretations and supporting evidence.Nucleic Acids Research, 46(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153.
24. Lépine, S. et al. Homozygous ALS-linked mutations in TARDBP/TDP-43 lead to hypoactivity and synaptic abnormalities in human iPSC-derived motor neurons. *IScience* **27** (3), 109166. https://doi.org/10.1016/j.isci.2024.109166 (2024).
25. Majeed, A. & Mukhtar, S. Protein–Protein Interaction Network Exploration Using Cytoscape (pp. 419–427). (2023). https://doi.org/10.1007/978-1-0716-3327-4_32
26. Meng, L. et al. Mini-review: Recent advances in post-translational modification site prediction based on deep learning. *Comput. Struct. Biotechnol. J.* **20**, 3522–3532. https://doi.org/10.1016/j.csbj.2022.06.045 (2022).
27. Van Es, M. A. Amyotrophic lateral sclerosis; clinical features, differential diagnosis and pathology (pp. 1–47). (2024). https://doi.org/10.1016/bs.irn.2024.04.011
28. Mohamed, N. M., Mohamed, R. H., Kennedy, J. F., Elhefnawi, M. M. & Hamdy, N. M. A comprehensive review and in silico analysis of the role of survivin (BIRC5) in hepatocellular carcinoma hallmarks: A step toward precision. *Int. J. Biol. Macromol.* **311**, 143616. https://doi.org/10.1016/j.ijbiomac.2025.143616 (2025).
29. Mukherjee, S., Das, S., Sriram, N., Chakraborty, S. & Sah, M. K. In silico investigation of the role of vitamins in cancer therapy through inhibition of MCM7 oncoprotein. *RSC Adv.* **12** (48), 31004–31015. https://doi.org/10.1039/D2RA03703C (2022).
30. Opie-Martin, S., Iacoangeli, A., Topp, S. D., Abel, O., Mayl, K., Mehta, P. R., Shatunov,A., Fogh, I., Bowles, H., Limbachiya, N., Spargo, T. P., Al-Khleifat, A., Williams,K. L., Jockel-Balsarotti, J., Bali, T., Self, W., Henden, L., Nicholson, G. A., Ticozzi,N., … Shaw, C. E. (2022). The SOD1-mediated ALS phenotype shows a decoupling between age of symptom onset and disease duration. Nature Communications, 13(1), 6901. https://doi.org/10.1038/s41467-022-34620-y.
31. Pavithran, H. & Kumavath, R. In silico analysis of nsSNPs in CYP19A1 gene affecting breast cancer associated aromatase enzyme. *J. Genet.* **100** (2), 23. https://doi.org/10.1007/s12041-021-01274-6 (2021).
32. Ruffo, P., Perrone, B. & Conforti, F. L. SOD-1 Variants in Amyotrophic Lateral Sclerosis: Systematic Re-Evaluation According to ACMG-AMP Guidelines. *Genes* **13** (3), 537. https://doi.org/10.3390/genes13030537 (2022).

33. Schubach, M., Maass, T., Nazaretyan, L., Röner, S. & Kircher, M. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Res.* **52** (D1), D1143–D1154. https://doi.org/10.1093/nar/gkad989 (2024).
34. Sim, N. L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40** (W1), W452–W457. https://doi.org/10.1093/nar/gks539 (2012).
35. Suk, T. R. & Rousseaux, M. W. C. The role of TDP-43 mislocalization in amyotrophic lateral sclerosis. *Mol. Neurodegeneration*. **15** (1), 45. https://doi.org/10.1186/s13024-020-00397-1 (2020).
36. Szklarczyk, D. et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51** (D1), D638–D646. https://doi.org/10.1093/nar/gkac1000 (2023).
37. Taylor, J. P., Brown, R. H. & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539** (7628), 197–206. https://doi.org/10.1038/nature20413 (2016).
38. Trist, B. G., Genoud, S., Roudeau, S., Rookyard, A., Abdeen, A., Cottam, V., Hare,D. J., White, M., Altvater, J., Fifita, J. A., Hogan, A., Grima, N., Blair, I. P.,Kysenius, K., Crouch, P. J., Carmona, A., Rufin, Y., Claverol, S., Van Malderen, S.,… Double, K. L. (2022). Altered SOD1 maturation and post-translational modification in amyotrophic lateral sclerosis spinal cord. Brain, 145(9), 3108–3130. https://doi.org/10.1093/brain/awac165.
39. Trist, B. G., Hilton, J. B., Hare, D. J., Crouch, P. J. & Double, K. L. Superoxide Dismutase 1 in Health and Disease: How a Frontline Antioxidant Becomes Neurotoxic. *Angew. Chem. Int. Ed.* **60** (17), 9215–9246. https://doi.org/10.1002/anie.202000451 (2021).
40. Vintilescu, C. R., Afreen, S., Rubino, A. E. & Ferreira, A. The Neurotoxic Tau45-230 Fragment Accumulates in Upper and Lower Motor Neurons in Amyotrophic Lateral Sclerosis Subjects. *Mol. Med.* **22** (1), 477–486. https://doi.org/10.2119/molmed.2016.00095 (2016).
41. Wang, X., Hu, Y. & Xu, R. The pathogenic mechanism of TAR DNA-binding protein 43 (TDP-43) in amyotrophic lateral sclerosis. *Neural Regeneration Res.* **19** (4), 800–806. https://doi.org/10.4103/1673-5374.382233 (2024).
42. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46** (W1), W296–W303. https://doi.org/10.1093/nar/gky427 (2018).
43. Xia, F., Dou, Y., Lei, G. & Tan, Y. FPGA accelerator for protein secondary structure prediction based on the GOR algorithm. *BMC Bioinform.* **12** (S1). S5. https://doi.org/10.1186/1471-2105-12-S1-S5 (2011).
44. Yuan, D., Jiang, S. & Xu, R. Clinical features and progress in diagnosis and treatment of amyotrophic lateral sclerosis. *Ann. Med.* **56** (1). https://doi.org/10.1080/07853890.2024.2399962 (2024).
45. Zeng, J. et al. Decoding TDP-43: the molecular chameleon of neurodegenerative diseases. *Acta Neuropathol. Commun.* **12** (1), 205. https://doi.org/10.1186/s40478-024-01914-9 (2024).

Figures

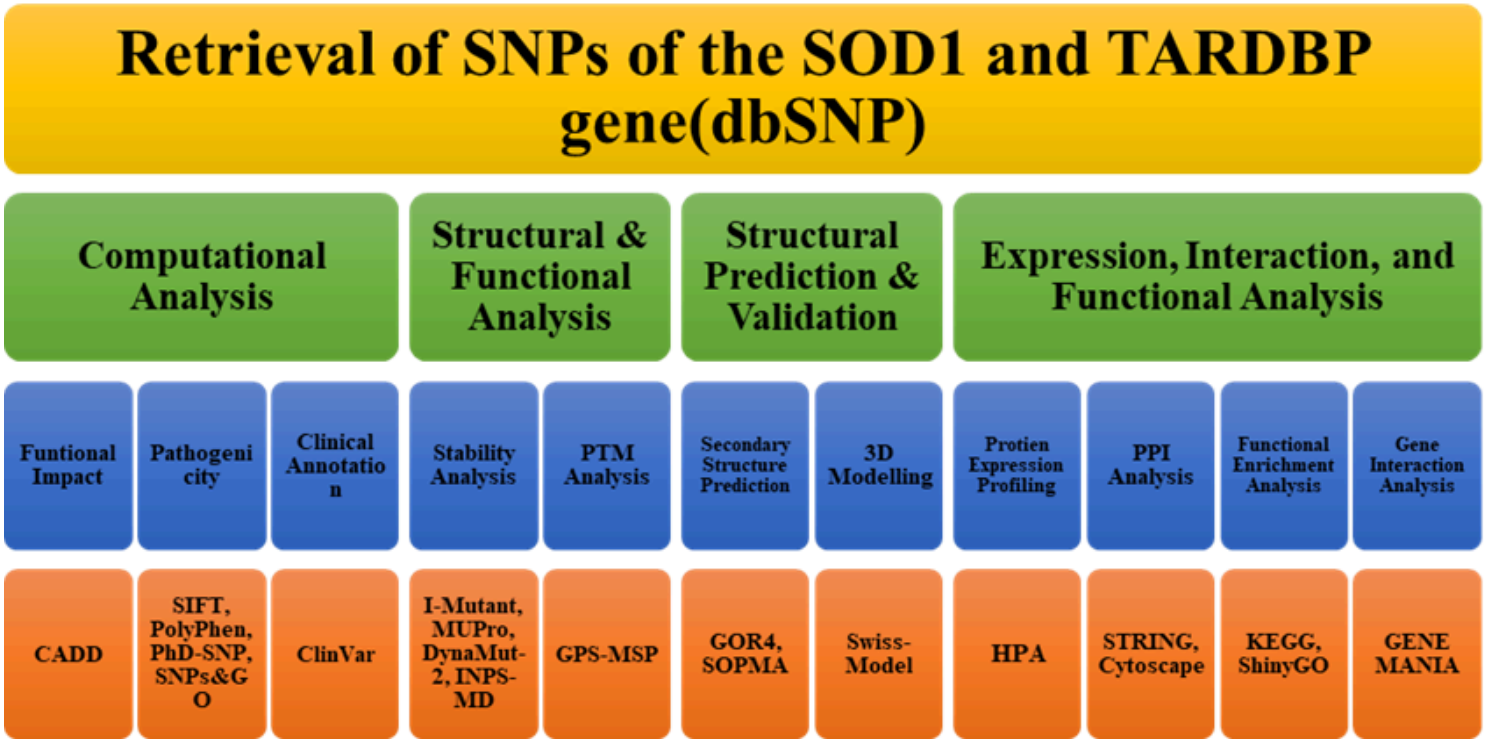


Figure 1

Workflow for SNP retrieval, inclusion, and analysis using *in silico* approaches.

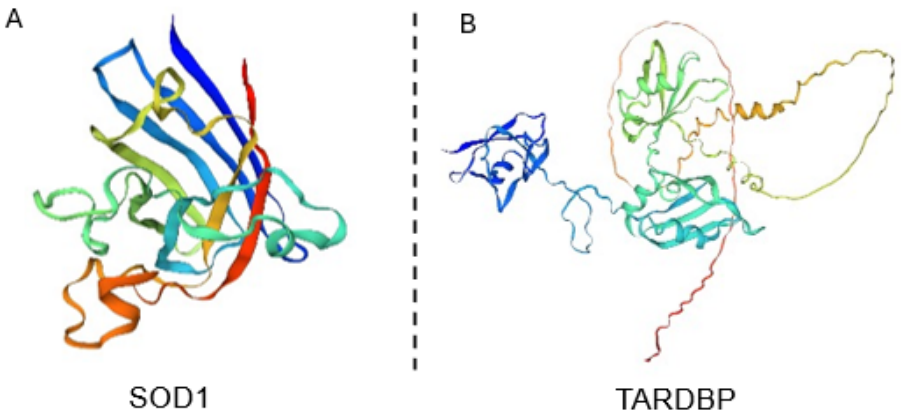


Figure 2

Homology-based three-dimensional structural models of pathogenic variants in *SOD1* and *TARDBP* proteins. The protein structures are displayed as ribbon diagrams in a rainbow gradient (blue to red) from the N-terminal to the C-terminal region. (A) *SOD1* variant model illustrates the characteristic β-barrel fold composed of β-sheets and connecting loops, with localized alterations near the mutation site that may influence metal-binding stability. (B) *TARDBP* variant model showing the multi-domain *TDP-43* architecture, including RNA recognition motifs and flexible coil regions, with the mutation positioned in a structurally dynamic domain that potentially affects RNA binding and protein aggregation behaviour.

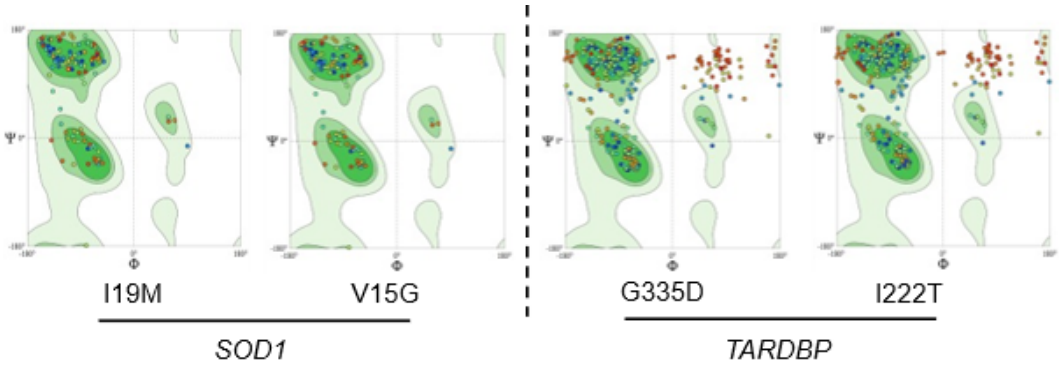


Figure 3

Ramachandran plot validation of three-dimensional structural models for *SOD1*and *TARDBP* variants. The plots illustrate the distribution of backbone dihedral angles phi (φ) and psi (ψ) for all residues, with shaded regions representing sterically allowed conformations. (A) *SOD1* variants rs1182088847 (I19M) and (B) rs1202989817 (V15G) of displayed residues predominantly within favored and allowed regions, indicating well-defined geometries and strongerstereochemical quality. (C) *TARDBP* variants rs80356729 (G335D) and (D) rs1570722030 (I222T) showed the most residue distribution within allowed regions but with a slightly higher proportion of outliers, suggesting localized conformational strain. All models demonstrate acceptable stereochemical reliability (MolProbity ≤ 1.76), confirming their suitability for further structural and functional interpretation.

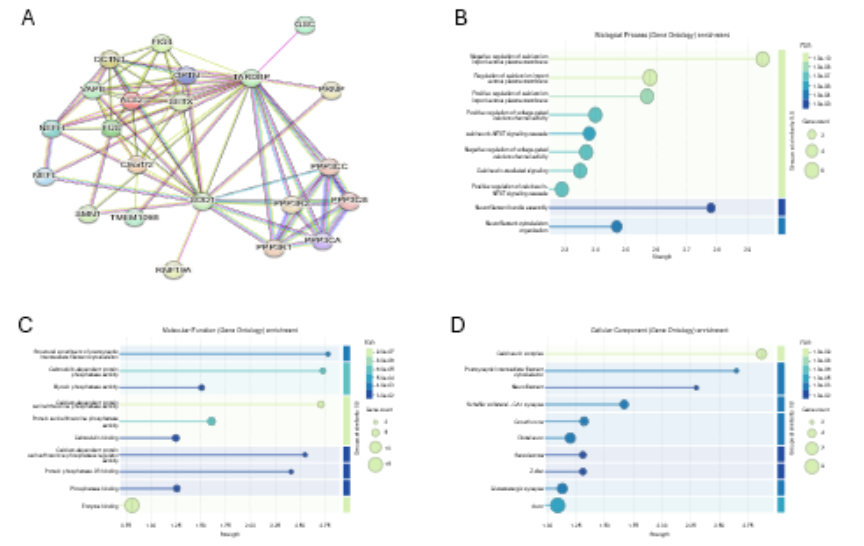


Figure 4

Functional enrichment and interaction landscape of the *SOD1-TARDBP* network. (A) STRING protein-protein interaction (PPI) network showing the functional connectivity between *SOD1, TARDBP,* and ALS-related genes. Edge colors correspond to different evidence types used by STRING, including experimental data, curated pathway information, co-expression patterns, and text-mining associations. (B) Gene Ontology- Biological Process (GO-BP) enrichment of the network. Bubble size represents the number of genes contributing to each term, while color indicates statistical significance (FDR). (C) Gene Ontology Molecular Function (GO-MF) enrichment, showing overrepresented functional activities within the network. (D) Gene Ontology-Cellular Component (GO-CC) enrichment highlighting the predominant subcellular localizations associated with *SOD1-TARDBP*interaction partners.

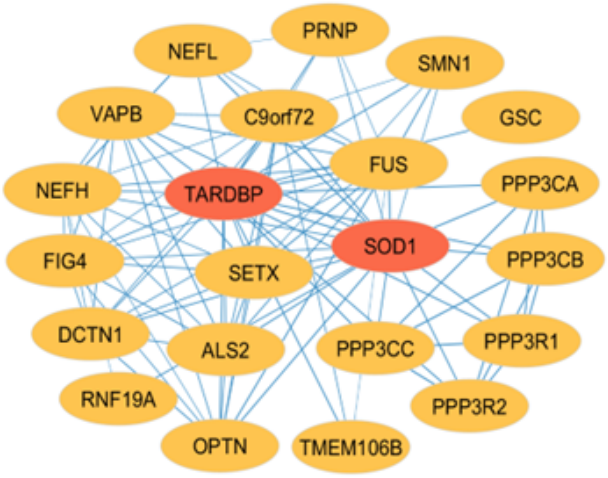


Figure 5

Network topology of *SOD1* and *TARDBP* within the ALS-associated STRING-Cytoscape interaction map. The Cytoscape visualization shows the central positioning of *SOD1* and *TARDBP*(highlighted in red) and their strongest interaction partners (yellow nodes) based on STRING-derived functional connections. Edges represent high-confidence associations including co-expression, physical interaction, pathway co-membership, and shared protein domains. Network topology parameters-including average shortest path length, clustering coefficient, closeness centrality, betweenness centrality, and degree-were computed using the Network-Analyzer plugin in Cytoscape. Genes with higher degree and betweenness centrality serve as major hubs and bottleneck regulators within the ALS network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDataREV2.docx
- TableS1.xlsx