Article

Elucidating the Epigenetic and Protein Interaction Landscapes in Amyotrophic Lateral Sclerosis: An Integrated Bioinformatics Analysis

Katerina Kadena and Panagiotis Vlamos *

Department of Informatics, Ionian University, 49100 Corfu, Greece; c18kade@ionio.gr
* Correspondence: vlamos@ionio.gr

Abstract: Background: Amyotrophic Lateral Sclerosis (ALS) is a debilitating neurodegenerative disorder characterized by the progressive degeneration of motor neurons, leading to muscle weakness and paralysis. Understanding the molecular basis of ALS is crucial for the development of effective therapies. Objective: This study aims to explore the genetic and epigenetic underpinnings of ALS, focusing on the interplay between gene mutations, protein interactions, and epigenetic factors. Methods: We conducted an extensive analysis of key ALS-associated genes including TARDBP, SOD1, ANG, VAPB, and CHMP2B. We used computational tools to assess the functional consequences of identified mutations on neuronal health and explored DNA methylation patterns in gene promoters to investigate epigenetic regulation. Results: Our findings reveal that mutations in ALS-associated genes disrupt critical processes such as amyloid fibril formation and autophagy. We also identified altered DNA methylation patterns, suggesting a mechanism for changes in gene expression linked to ALS. Molecular docking studies highlighted Humulene and Buddledin C as compounds with high binding affinities to the SOD1 enzyme, suggesting their potential to mitigate hallmark features of ALS pathology such as SOD1 aggregation and oxidative stress. Conclusions: Our comprehensive analysis underscores the complexity of ALS pathogenesis, combining genetic, epigenetic, and proteomic approaches. The insights gained not only enhance our understanding of ALS but also pave the way for novel therapeutic strategies, highlighting the importance of integrated approaches in tackling this challenging neurodegenerative disease.

Keywords: Amyotrophic Lateral Sclerosis (ALS); protein–protein interaction (PPI) network; gene ontology; epigenetics; molecular docking simulations

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a debilitating neurodegenerative disorder that progressively impairs the motor neurons in the spinal cord and brainstem. The disease predominantly affects men and is characterized by a high degree of clinical heterogeneity, reflecting its complex pathogenic mechanisms. Annually, ALS affects approximately 2–3 individuals per 100,000 in the European population, underscoring a significant health concern [1]. The severity of the disease ranges from complete loss of bulbular and limb function to patient death after three years of symptom onset caused by failure of the respiratory tract. The diagnostic and therapeutic challenges posed by ALS are compounded by the lack of specific biomarkers, making clinical criteria the current standard for diagnosis [2].

Approximately 5–10% of ALS cases are familial, often linked to mutations in genes such as SOD1, TARDBP, FUS, ANG, and OPTN, which contribute to its clinical phenotype [3]. Moreover, the overlap of ALS with frontotemporal dementia in about 15% of patients complicates the disease management, as these patients exhibit significant cognitive and behavioral impairments [4]. Patients have trouble in making judgements and doing routine tasks and are more impulsive than others. This is followed by verbal impairment.
and cognitive disability, which makes the patient more difficult to handle [5,6]. The leading cause of ALS in most patients remains unclear, as the symptoms differ in different population origins. Neurotoxins like β-methyl-amino-L-alanine and glutamate-induced excitotoxicity are linked to ALS [7].

The current study is specifically designed to address the complex genetic and epigenetic landscape of ALS by focusing on the critical interplay among key genetic mutations and their epigenetic modulations. Specifically, mutations in TARDBP are implicated in approximately 3% of familial ALS cases and about 1.5% of sporadic ALS cases, leading to TDP-43 proteinopathy, a core pathological feature of ALS [8]. Additionally, mutations in FUS are associated with neuronal dysfunction in ALS, distinguishing affected patients from healthy controls [9]. Our research aims to critically analyze the interactions between TARDBP, FUS, SOD1, and other significant genes by leveraging advanced bioinformatics tools and integrative analytical approaches.

Our study represents a comprehensive effort to elucidate the complex genetic and epigenetic landscape of ALS, aiming to bridge the gaps in our current understanding of its pathogenesis. By integrating genetic mutation analysis, protein–protein interaction studies, and epigenetic profiling, we seek to unravel the multifaceted mechanisms underlying ALS. The identification of potential therapeutic compounds through molecular docking studies marks a significant stride towards novel treatment strategies. This research not only contributes to the fundamental knowledge of ALS but also holds promise for the development of targeted therapies, offering hope in the fight against this devastating neurodegenerative disorder.

2. Methodology

2.1. Identification of Genetic Variations in ALS

A systematic search strategy was implemented to identify mutations associated with Amyotrophic Lateral Sclerosis (ALS), utilizing the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/; accessed on 30 January 2024) [10]. The search term “Amyotrophic Lateral Sclerosis” was applied to explore relevant databases and genetic repositories comprehensively. The objective was to compile an exhaustive list of mutations that have been previously documented in ALS patients and are recognized for their high pathogenic potential. The details of these mutations are presented in Supplementary File S1: Table S1.

2.2. Functional Prediction of Identified Mutations

The functional implications of amino acid substitutions were predicted using the SIFT tool (http://sift-dna.org/; accessed on 30 January 2024) [11]. Specifically, SIFT scores an amino acid substitution from 0 to 1, with a score below 0.05 indicating a deleterious or intolerant substitution likely to affect protein function adversely. This tool assesses the degree of conservation among amino acid residues across a wide range of species, utilizing multiple sequence alignments to determine which amino acids are critical for function and are thus conserved. Furthermore, the PolyPhen-2 tool (http://genetics.bwh.harvard.edu/pph2/; accessed on 30 January 2024) [12] was also utilized. PolyPhen-2 integrates multiple sources of information including sequence alignment, phylogenetic profiles, and structural attributes to calculate a score that predicts whether a substitution is probably damaging, possibly damaging, or benign. This score is derived from machine learning algorithms based on a training set of human missense mutations along with their known effects. Thus, this dual approach ensures a comprehensive assessment of the mutational impact, aiding in the identification of variants with significant functional consequences.

2.3. Comprehensive Network Analysis of ALS-Associated Genes

After the elucidation of the functional roles of the selected genes in ALS, a protein–protein interaction (PPI) network was constructed to assess their connectivity. The 23 genes were submitted to the STRING database (https://string-db.org/; accessed on 30 January
2.4. Identification and Characterization of Methylation Sites in ALS-Associated Genes

To further investigate the epigenetic regulation of ALS-associated genes, a detailed analysis of methylation and gene expression microarray data was conducted using R software version 3.1.2 [14]. The microarray data, normalized and scrutinized using the Bioconductor minfi and limma packages, identified aberrant methylation patterns and differential gene expression between cases and controls. The Benjamini–Hochberg false discovery rate (FDR) method was applied to adjust P values for each CpG site and gene. An FDR-adjusted P value below 0.05 was established as the threshold to determine differentially methylated CpG sites (DMCs) and differentially expressed genes (DEGs). Additionally, delta β values greater than 0.2 or less than −0.2 were used to define hypermethylated or hypomethylated genes, respectively. For the expression data, a delta expression value beyond 1 was classified as upregulated, whereas a value below −1 indicated downregulation. All identified DMCs were annotated to their corresponding differentially methylated genes (DMGs) based on the microarray platform’s annotation file. To corroborate these findings, methylation and expression microarray data were also obtained from the TCGA database [15], utilizing the Illumina HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA, USA) and IlluminaHiSeq platforms (Illumina, Inc., San Diego, CA, USA), respectively. This validation process involved over 450,000 CpG sites annotated in the human genome, including multiple sites within the body of hub genes that are implicated in the regulation of gene expression.

2.5. Molecular Docking Analysis

In this study, proteins with a higher number of methylation sites were selected for further analysis. A comprehensive library of 7241 phytochemicals was docked against these target proteins. For the docking simulations, a specific inhibitor binding site was targeted. The PyRx tool [16], functioning as a front-end for AutoDock Vina, facilitated these simulations, employing both rigid and flexible docking parameters. The docking grid was defined with dimensions of 20 units in each axis \((x, y, z)\) and was positioned to encompass the entire binding site, with coordinates set at \((x = 112.015, y = 106.402, z = 131.8428)\). AutoDock Vina’s empirical scoring function, aggregating contributions from various individual terms, determined the affinity of protein–compound binding. The docked complex with the lowest root mean square deviation (RMSD) was considered optimal. Binding energies between ligands and target proteins were evaluated based on their affinity, with values below −5.00 kcal/mol indicating good binding strength, and values below −7.00 kcal/mol suggesting very good affinity. The top five phytochemicals with the highest binding affinity were selected, exhibiting structural diversity and potential as potent inhibitors of the target protein, as indicated by their docking scores and interaction patterns.

2.6. Molecular Dynamics (MD) Simulation

Molecular dynamics (MD) simulations offer a comprehensive, atom-level representation of molecular dynamics, crucial for understanding complex biological interactions. This approach, employing GROMACS 2018 software (version 18) [17] and the OPLS-AA/L force field, was utilized to simulate the dynamics of docked complexes. Initial structures, derived from the 3D structure of the protein and optimized using DockPrep, served as
the basis for simulations. MD simulations commenced using complexes with the highest binding affinity, positioned initially as per their docked states. The ligand molecules were parameterized using the SwissParam webserver (http://swissparam.ch/; accessed on 30 January 2024). These simulations, conducted over a 20 ns period, were in line with established protocols from previous studies. Essential parameters, including the radius of gyration, RMSD, SSE, and root mean square fluctuation (RMSF), were evaluated to assess the stability and interaction dynamics of each complex.

3. Results

3.1. Identification and Functional Prediction of Pathogenic Mutations in ALS

The study began with an extensive search for genetic variations associated with Amyotrophic Lateral Sclerosis (ALS), leading to the identification of 354 mutations deemed highly pathogenic. These mutations are meticulously detailed in Supplementary File S1: Table S1 and Figure 1. To evaluate the functional impact of these mutations, two computational prediction tools, SIFT (Sorting Intolerant From Tolerant, accessed on 30 January 2024) and PolyPhen-2 (Polymorphism Phenotyping v2, accessed on 30 January 2024), were employed. These tools are specifically designed to assess the deleterious effects of amino acid substitutions on protein function, thereby providing insights into the potential pathogenicity of these mutations.

Figure 1. Comprehensive analysis of SNP variations in human genes: This figure illustrates various aspects of Single-Nucleotide Polymorphisms (SNPs) in human genes. (A) depicts the count of SNPs identified per gene, highlighting genes with the highest variability. (B) shows the distribution of SNP coordinates, providing insights into their genomic spread. (C) illustrates a scatter plot of SNPs across different chromosomes, color-coded by gene, to visualize chromosomal distribution. Finally, (D) presents the distribution of SNP coordinates for each gene, offering a detailed view of genomic location variability within each gene. These plots collectively provide a multifaceted view of SNP characteristics, crucial for understanding genetic variations and potential implications in human health and disease.

The analysis revealed a subset of mutations across 23 genes predicted to be highly deleterious. This significant finding (Table 1) highlights the genetic complexity underlying ALS. Due to their predicted high impact on protein function, the identified mutations may play crucial roles in the disease's pathogenesis. This discovery not only contributes to the understanding of the genetic underpinnings of ALS but also opens avenues for future research into targeted therapies and genetic counseling for this condition.

Table 1. Summary of deleterious Single-Nucleotide Polymorphisms (SNPs) and associated amino acid changes in various genes linked to ALS.
identified per gene, highlighting genes with the highest variability. (B) shows the distribution of SNP coordinates, providing insights into their genomic spread. (C) illustrates a scatter plot of SNPs across different chromosomes, color-coded by gene, to visualize chromosomal distribution. Finally, (D) presents the distribution of SNP coordinates for each gene, offering a detailed view of genomic location variability within each gene. These plots collectively provide a multifaceted view of SNP characteristics, crucial for understanding genetic variations and potential implications in human health and disease.

The analysis revealed a subset of mutations across 23 genes predicted to be highly deleterious. This significant finding (Table 1) highlights the genetic complexity underlying ALS. Due to their predicted high impact on protein function, the identified mutations may play crucial roles in the disease’s pathogenesis. This discovery not only contributes to the understanding of the genetic underpinnings of ALS but also opens avenues for future research into targeted therapies and genetic counseling for this condition.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Amino Acid Change</th>
<th>Gene Name</th>
<th>Predictions</th>
<th>SNP</th>
<th>Amino Acid Change</th>
<th>Gene Name</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs80356717</td>
<td>D53G</td>
<td>TARDBP</td>
<td>Deleterious</td>
<td>rs267606685</td>
<td>F224L</td>
<td>SLC52A3</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs80356723</td>
<td>G179R</td>
<td>TARDBP</td>
<td>Deleterious</td>
<td>rs267606688</td>
<td>P28T</td>
<td>SLC52A3</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs80356733</td>
<td>G232C</td>
<td>TARDBP</td>
<td>Deleterious</td>
<td>rs281875284</td>
<td>T46I</td>
<td>VAPB</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs72466487</td>
<td>Q74P</td>
<td>DCTN1</td>
<td>Deleterious</td>
<td>rs74315431</td>
<td>P56S</td>
<td>VAPB</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs67586389</td>
<td>G71A</td>
<td>DCTN1</td>
<td>Deleterious</td>
<td>rs121912444</td>
<td>A5T</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs121909342</td>
<td>G59S</td>
<td>DCTN1</td>
<td>Deleterious</td>
<td>rs121912442</td>
<td>A5V</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs12190138</td>
<td>C157Y</td>
<td>ALS2</td>
<td>Deleterious</td>
<td>rs121912448</td>
<td>C7F</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs368743618</td>
<td>A383T</td>
<td>TUBA4A</td>
<td>Deleterious</td>
<td>rs121912453</td>
<td>G17S</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs63750653</td>
<td>D118Y</td>
<td>CHMP2B</td>
<td>Deleterious</td>
<td>rs121912431</td>
<td>G38R</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs63751126</td>
<td>Q206H</td>
<td>CHMP2B</td>
<td>Deleterious</td>
<td>rs121912433</td>
<td>G42S</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs207482230</td>
<td>P281L</td>
<td>TFG</td>
<td>Deleterious</td>
<td>rs121912434</td>
<td>G42D</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs371575563</td>
<td>R550G</td>
<td>NEK1</td>
<td>Deleterious</td>
<td>rs121912457</td>
<td>P27C</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs121434591</td>
<td>S85C</td>
<td>MATR3</td>
<td>Deleterious</td>
<td>rs121912443</td>
<td>H47R</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs143511494</td>
<td>G327S</td>
<td>SQSTM1</td>
<td>Deleterious</td>
<td>rs121912458</td>
<td>H81R</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs387906829</td>
<td>E102Q</td>
<td>SIGMAR1</td>
<td>Deleterious</td>
<td>rs121912452</td>
<td>L85V</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs140376902</td>
<td>L65Q</td>
<td>SIGMAR1</td>
<td>Deleterious</td>
<td>rs121912436</td>
<td>G86R</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs267607087</td>
<td>S331F</td>
<td>SPTLC1</td>
<td>Deleterious</td>
<td>rs11556620</td>
<td>N87S</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs29001584</td>
<td>L389S</td>
<td>SETX</td>
<td>Deleterious</td>
<td>rs121912437</td>
<td>G94R</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs267606929</td>
<td>E472G</td>
<td>OPTN</td>
<td>Deleterious</td>
<td>rs121912438</td>
<td>G75A</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs397507527</td>
<td>G268C</td>
<td>PTPN11</td>
<td>Deleterious</td>
<td>rs121912440</td>
<td>L88V</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs121909535</td>
<td>Q36L</td>
<td>ANG</td>
<td>Deleterious</td>
<td>rs74315452</td>
<td>I113T</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs121909539</td>
<td>C63W</td>
<td>ANG</td>
<td>Deleterious</td>
<td>rs121912441</td>
<td>I114T</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs121909540</td>
<td>K64I</td>
<td>ANG</td>
<td>Deleterious</td>
<td>rs121912451</td>
<td>S135N</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs147713329</td>
<td>R2031G</td>
<td>SPG11</td>
<td>Deleterious</td>
<td>rs121912449</td>
<td>I133T</td>
<td>SOD1</td>
<td>Deleterious</td>
</tr>
<tr>
<td>rs267606831</td>
<td>G507D</td>
<td>FUS</td>
<td>Deleterious</td>
<td>rs387906709</td>
<td>P497H</td>
<td>UBQLN2</td>
<td>Deleterious</td>
</tr>
</tbody>
</table>

3.2. Gene Ontology Analysis of Identified Genes in ALS

The gene ontology analysis revealed that the genes associated with ALS were significantly enriched in key biological processes (Supplementary File S2: Table S1). These
processes include amyloid fibril formation, autophagy, and neuromuscular junction development, highlighting their potential roles in the pathogenesis of ALS. Other notable processes include endosome organization, protein homooligomerization, and motor behavior, indicating a broad impact on cellular functioning. Further, genes were also involved in viral release from the host cell, regulation of superoxide anion generation, and positive regulation of microtubule polymerization. Processes like neuron cellular homeostasis, regulation of I-kappaB kinase/NF-kappaB signaling, and cell death underscore the complex interplay of cellular mechanisms in ALS. Additionally, genes were linked to developmental processes like placenta and ovarian follicle development, as well as responses to external stimuli such as heat and reactive oxygen species. The findings also encompassed broader aspects like fibroblast growth factor receptor signaling pathway, macroautophagy, nervous system development, and endosomal transport. The modulation of synaptic transmission and locomotory behavior, positive regulation of autophagy and memory, and regulation of GTPase activity were among other key processes identified.

Further, the cellular component analysis indicated a significant enrichment of genes in areas such as the cytoplasm, growth cone, and cytosol, reflecting their widespread distribution within the cell. Other components like cytoplasmic vesicles, neuronal cell bodies, and the nucleus were also prominent (Supplementary File S3: Table S1). The genes were found in specialized structures such as autophagosomes, axons, microtubule cytoskeleton, intracellular non-membrane-bounded organelles, postsynaptic density, lysosomes, and microtubules. Additional components included the axon cytoplasm, dendrites, and the nuclear inner membrane, highlighting the diverse cellular localizations of these genes.

Lastly, in terms of molecular functions, the genes were primarily enriched in functions like identical protein binding and protein binding, indicating their critical roles in protein-protein interactions (Supplementary File S4: Table S1). Protein kinase binding, small GTPase binding, and cadherin binding were also identified as significant functions. The analysis showed enrichment in protein binding, bridging, RNA binding, and K63-linked polyubiquitin binding, suggesting a role in complex molecular interactions. Other functions included microtubule binding, polyubiquitin binding, phosphotyrosine binding, actin binding, and DNA binding. Additionally, the genes were associated with copper ion binding, enzyme binding, and receptor tyrosine kinase binding, emphasizing their diverse functional roles in cellular processes.

### 3.3. Network Analysis

To further elucidate the interplay between the identified genes in ALS, the genes were submitted to the STRING database, a comprehensive resource for the analysis of protein-protein interactions. This analysis aimed to construct a PPI network, providing insights into how these proteins might interact with each other and contribute to the disease pathology. The PPI analysis, as depicted in Figure 2D, revealed a highly interconnected network among the genes. This intricate network underscores the complex molecular interactions and pathways potentially involved in ALS. The high degree of connectivity associated with these genes suggests a stronger association with ALS pathogenesis. This finding is particularly noteworthy, as proteins with higher connectivity within PPI networks are often crucial in disease processes, potentially serving as key regulators or hubs in disease-related pathways. The network analysis using the STRING database has provided valuable insights into the protein interactions of ALS-associated genes. It highlights the importance of considering the interconnected nature of biological pathways and molecular interactions in understanding the mechanisms underlying complex diseases like ALS.
Abstract: Background: Amyotrophic Lateral Sclerosis (ALS) is a debilitating neurodegenerative disorder characterized by the progressive degeneration of motor neurons, leading to muscle weakness and paralysis. Understanding the molecular basis of ALS is crucial for the development of effective therapies. Objective: This study aims to explore the genetic and epigenetic underpinnings of ALS, focusing on the interplay between gene mutations, protein interactions, and epigenetic factors.

Methods: We conducted an extensive analysis of key ALS-associated genes including TARDBP, SOD1, ANG, VAPB, and CHMP2B. We used computational tools to assess the functional consequences of identified mutations on neuronal health and explored DNA methylation patterns in gene promoters to investigate epigenetic regulation.

Results: Our findings reveal that mutations in ALS-associated genes disrupt critical processes such as amyloid fibril formation and autophagy. We also identified methylation sites within target genes.

Figure 2. Gene ontology and protein–protein interaction network analysis of 23 ALS-associated genes. (A) Bar graph showing the distribution of 23 ALS-associated genes across different biological processes. The length of each bar represents the count of genes involved in each process, with amyloid fibril formation, autophagy, and neuromuscular junction development being prominently featured. (B) Volcano plot of molecular functions highlighting the $-\log(p\text{-value})$ against the count of genes associated with each function. Larger and more colored dots represent higher significance and greater gene counts, respectively, with protein binding, RNA binding, and enzyme binding being the most significant functions. (C) Pie chart detailing the cellular components with which the 23 ALS-associated genes are associated. Each segment’s size reflects the percentage of genes associated with each cellular component, with the cytoplasm and growth cone being the most represented. (D) Protein–protein interaction (PPI) network of the 23 ALS-associated genes. Nodes represent individual genes, with their size proportional to the number of connections, indicating their centrality in the network. Lines represent the interactions between genes, illustrating a complex web of connectivity with SOD1, FUS, and TFG among the most connected nodes.

3.4. Identification of Methylation Site within Target Genes

Methylation data from GSE242475 and GSE12525 were downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/, accessed on 30 January...
There were 28 ALS and six control specimens in GSE33202 and 10 ALS and 10 control specimens in GSE12525. The DMCs of these two methylation datasets were analyzed separately. Similarly, a total of 784 DEGS were obtained from GSE233881. After identification of multiple DMCs and DEGs from these microarray datasets, a matching function was performed to identify hypermethylated overlapping genes in ALS. A total of nine genes named TARDBP, DCTN1, ALS2, CHMP2B, SPTLC1, SETX, ANG, VAPB, and SOD1 were found by comparing differentially expressed miRNAs (DEMs) and DEGs data with the genes identified to be highly mutated.

For the TARDBP gene, methylation within the promoter region’s CpG island might modulate gene expression, which is significant considering that the TARDBP gene’s product, TDP-43, is a known factor in ALS pathology when mutated. This epigenetic regulation could potentially contribute to the disease’s progression by altering TDP-43 levels. Similarly, the CHMP2B and SOD1 genes, both implicated in ALS, also exhibit CpG islands within their promoter regions, suggesting a possible regulatory role for methylation. In particular, SOD1 is a well-established gene in familial ALS cases, where its expression level, potentially controlled by promoter methylation, is crucial to disease onset and progression. Though no methylation sites are specified for DCTN1, and no CpG islands are identified within the gene domain, this does not rule out a regulatory role for methylation elsewhere that could impact ALS. On the other hand, the ALS2 gene shows an upstream methylation site outside a CpG island but also contains a promoter CpG island, indicating a complex regulatory landscape that may influence juvenile ALS forms. Notably, several genes, including SPTLC1, SETX, ANG, and VAPB, exhibit intronic methylation, with SPTLC1’s site situated within a CpG island. Such intronic methylation could influence alternative splicing or gene expression patterns, which might be particularly relevant to ALS, where protein function and cellular homeostasis are compromised. Overall, the presence of methylation sites, especially within CpG islands of promoter regions, points to epigenetic mechanisms that could downregulate gene expression and contribute to the pathology of ALS. Understanding these patterns could offer novel insights into ALS’s etiology, possibly opening avenues for epigenetic therapies alongside genetic approaches to manage or slow the disease’s progression.

3.5. Molecular Docking Analysis

The quest to uncover novel inhibitors for the SOD1 enzyme, a significant target in ALS research, led to an extensive molecular docking study. Initiating with the retrieval of the SOD1 enzyme’s 3D structure from the Protein Data Bank (PDB ID: 5YTU) (https://www.rcsb.org/structure/5ytu; accessed on 30 January 2024), a focused approach was adopted. A large-scale virtual screening was performed, involving a library of 7241 phytochemicals, to discern compounds that exhibit the highest binding affinity to the active site of SOD1—particularly within the same binding pockets known to harbor ALS-causing mutations. The docking process was designed to simulate the interaction between the SOD1 enzyme and each phytochemical compound in the library, calculating the binding affinity to predict the strength and stability of the potential inhibitor–enzyme complexes. This computational
prediction is critical, as it informs the likelihood of a compound’s efficacy in modulating the enzyme’s activity, which is pivotal in the progression of ALS.

From the extensive screening, five compounds emerged as potential candidates based on their binding affinity scores, which are inversely related to the strength of the interaction—the more negative the score, the stronger the predicted binding affinity. These compounds, along with their calculated root mean square deviation (RMSD) values, provide insights into the stability of the docked poses. RMSD values closer to zero suggest a more accurate and reliable docked conformation, reflecting how closely the simulation matches the predicted native binding mode.

The compound Humulene showed the most promising result with the highest binding affinity of $-10.0984$ kcal/mol, coupled with a RMSD of 1.86057. This suggests not only a strong interaction with the SOD1 enzyme but also a potentially stable binding conformation. Following closely was Buddledin C, with a binding affinity of $-9.9865$ kcal/mol and a low RMSD of 0.73937, indicating a highly favorable interaction. Deoxynupharidine, though having a moderate binding affinity of $-5.8311$ kcal/mol, presented a RMSD of 2.40470, implying a less optimal fit. Azadiradione’s binding affinity of $-7.4170$ kcal/mol, along with the highest RMSD of 3.58678, suggests a weaker, yet potentially modulatory, interaction with SOD1. Lastly, [3-(Benzyloxy)isoxazol-5-yl]methanol demonstrated a robust binding affinity at $-9.4057$ kcal/mol and a RMSD of 1.59431, suggesting it too could be a stable and effective SOD1 inhibitor. The results of this molecular docking study, particularly the high binding affinity scores of Humulene and Buddledin C, alongside their favorable RMSD values, indicate strong potential for these compounds as SOD1 inhibitors (Figure 3). Their predicted interactions with the SOD1 enzyme offer a valuable foundation for further experimental studies, which are imperative to validate their efficacy and potential as therapeutic agents against ALS.

![Figure 3. Structural representations of SOD1 complexes with bioactive compounds. (A) Depicts the Humulene–SOD1 complex, with Humulene shown in red, docked within the SOD1 protein depicted in blue. (B) Illustrates the Buddledin C–SOD1 complex, with Buddledin C in red, interacting with the SOD1 protein shown in light blue. SOD1, or superoxide dismutase 1, is a protein that plays a critical role in cellular antioxidant defense, and the complexes shown highlight potential interactions with these natural compounds.](image-url)
3.6. Molecular Dynamics Simulation

Initially, the MD simulations were conducted between the SOD1 protein and the compound Buddledin C over a 20-nanosecond time frame, providing insights into the stability and behavior of the ligand–protein complex. Figure 4A presents a dual-axis graph where the root mean square deviation (RMSD) of the protein and the ligand fitted on the protein are plotted against simulation time. The protein RMSD fluctuates moderately, suggesting relative stability of the protein’s backbone throughout the simulation. The ligand RMSD shows higher variability, which could indicate conformational changes or flexibility in the ligand when bound to the protein. These fluctuations are crucial, as they can signify the binding stability and the dynamic nature of the interaction between Buddledin C and SOD1.

Figure 4B illustrates the root mean square fluctuation (RMSF) of the protein’s alpha carbons across each residue index. Peaks in the RMSF graph highlight residues that experience greater flexibility during the simulation, providing clues about the potential impact of Buddledin C binding on the protein’s conformational dynamics. Figure 4C
displays the secondary structure elements’ stability over the course of the simulation. The consistent maintenance of secondary structure, as indicated by the minimal variation in the percentage of secondary structure elements (% SSE), suggests that the protein structure is not significantly perturbed by the binding of Buddledin C.

Lastly, Figure 4D offers a detailed interaction map of Buddledin C with active site residues within SOD1. The interaction percentages denote the proportion of simulation time during which specific contacts were maintained. A key interaction is observed between Buddledin C and the residue TRP32, with a high interaction frequency, suggesting a strong and persistent binding. Other interactions, such as those with LYS30 and GLU21, are also noted, albeit with lower interaction frequencies, suggesting that though these interactions contribute to the binding, they are not as strong or consistent as those with TRP32. Together, these MD simulation results indicate that Buddledin C forms a stable complex with the SOD1 protein, engaging predominantly with the residue TRP32, which could be critical for its inhibitory action. The observed interactions and stability over the simulation period support the potential of Buddledin C as a candidate for further experimental validation in ALS therapeutic development.

Further, the MD simulation results of Humulene with SOD1 indicated a suitable interaction between the compounds and ligand proteins over a 20-nanosecond period. These results are essential for understanding the dynamic nature of the ligand–protein complex and for assessing the stability of Humulene as a potential therapeutic agent for ALS. Figure 5A illustrates the root mean square deviation (RMSD) for both the SOD1 protein backbone and Humulene, plotted over the simulation time. The RMSD values for the protein show moderate fluctuations, maintaining overall stability with no drastic conformational changes. Meanwhile, the RMSD for Humulene attached to the protein also displays fluctuations, which could indicate some level of flexibility or adaptability of the ligand within the active site. The dual-axis format allows for the direct comparison of the fluctuations in protein and ligand stability simultaneously.

In Figure 5B, the root mean square fluctuation (RMSF) of the alpha carbons in SOD1 is depicted, revealing the flexibility of individual residues when the protein interacts with Humulene. Peaks in the RMSF graph correspond to residues that experience greater mobility, which could be critical for understanding the binding dynamics and the potential impact on SOD1’s functionality upon ligand binding. The third panel, Figure 5C, shows the percentage of secondary structure elements (% SSE) throughout the duration of the simulation. The consistency in % SSE indicates that the overall secondary structure of SOD1 remains largely unaffected by the presence of Humulene, suggesting that the binding of the ligand does not disrupt the protein’s structural integrity. Finally, Figure 5D provides an interaction map between Humulene and specific residues of SOD1. It highlights the interaction frequencies, showing that certain residues, such as LYS30, maintain contact with the ligand for a significant portion of the simulation time. These persistent interactions are critical, as they indicate potential key binding sites and the strength of the ligand’s association with the protein. Overall, the MD simulation results suggest that Humulene binds to SOD1 with sufficient stability and interacts predominantly with specific residues that may be crucial for its inhibitory activity. These findings support further investigation of Humulene as a potential therapeutic candidate for ALS, with particular attention to the residues that exhibit strong and sustained interactions with the compound.
Figure 5. Molecular dynamics simulation of humulene bound to SOD1. (A) Root mean square deviation (RMSD) of the SOD1 protein backbone (blue) and Humulene (red) over the course of a 20 ns MD simulation, displaying the stability of the protein and the dynamic binding of the ligand. (B) Root mean square fluctuation (RMSF) for the alpha carbons of SOD1, highlighting the flexibility of individual amino acid residues upon Humulene interaction. (C) Percentage of secondary structure elements (% SSE) in SOD1, indicating the preservation of protein structure during ligand binding. (D) Interaction frequency map showing the binding interactions of Humulene with key residues of SOD1, notably with a significant interaction with LYS30, suggesting a strong and potentially influential binding affinity.

4. Discussion

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disorder known as Lou Gehrig’s disease, characterized by the degeneration and death of motor neurons in the brain and spinal cord. This results in muscle wasting and paralysis, with the disease presenting either sporadically or familially, adding complexity to its pathogenesis. Despite significant research, a complete understanding of ALS’s molecular mechanisms remains elusive, and treatments are largely focused on symptom management and slowing disease progression.

In response to this challenge, our study embarked on a comprehensive analysis of the genetic landscape of ALS, identifying numerous mutations with potential pathogenic impacts. We employed computational tools such as SIFT and PolyPhen-2 to predict the functional consequences of these mutations, suggesting significant effects on protein structure and function that may exacerbate disease progression. This genetic complexity highlights the heterogeneous nature of ALS and underscores the necessity of our genetic and epigenetic exploration.
Further deepening our understanding, gene ontology analysis revealed that ALS-associated genes are predominantly involved in crucial biological processes such as amyloid fibril formation, autophagy, and neuromuscular junction development. These genes are enriched in molecular functions like protein and enzyme binding, indicating that disturbances in these functions are central to ALS pathology. Our protein–protein interaction (PPI) network analysis added another layer of insight, illustrating a complex network of interactions among ALS-associated proteins, which could be key targets for therapeutic intervention.

The intricate web of protein interactions mapped in our study not only clarifies the molecular underpinnings of ALS but also highlights potential therapeutic targets that could alter disease progression. Identifying these key interaction nodes within the PPI network opens new avenues for developing treatments that extend beyond current symptom management strategies, suggesting a move towards modifying the disease course itself.

Gene ontology analyses have furthered our understanding by revealing the enrichment of ALS-associated genes in processes and components that are pivotal to neuronal health and disease, such as amyloid fibril formation, autophagy, and neuromuscular junction development. The enriched molecular functions, including protein binding and enzyme binding, align with the notion that perturbations in protein interactions and modifications can be central to ALS pathology. These insights align with the protein–protein interaction analyses, which illustrates the intricate network of interactions among ALS-associated proteins, potentially unveiling targets for therapeutic intervention.

Epigenetic regulation, particularly DNA methylation within promoter regions and CpG islands, emerges as a noteworthy dimension in the regulation of gene expression relevant to ALS. The correlation between promoter methylation patterns and gene expression levels in genes such as TARDBP, SOD1, and CHMP2B may offer a deeper understanding of the regulatory mechanisms that could be altered in ALS, paving the way for the exploration of epigenetic therapies.

The molecular docking studies conducted have yielded promising results, with compounds like Humulene and Buddledin C showing high binding affinities to the SOD1 enzyme, which is a significant target given its role in familial ALS. These findings are not only crucial for understanding the interaction dynamics but also for guiding future experimental validation. Molecular dynamic simulations have provided a detailed view of the stability and behavior of these potential inhibitor complexes with SOD1, revealing specific interactions that could be exploited for therapeutic benefit.

The network and methylation analyses, combined with the insights from molecular docking and dynamics simulations, converge to suggest that a holistic approach, encompassing genetic, epigenetic, and protein interaction perspectives, is essential for a thorough understanding of ALS. The identified compounds, Humulene and Buddledin C, stand out as potential inhibitors of SOD1, warranting further investigation. The interactions observed, particularly those within the active site and involving key residues, offer a strong foundation for the hypothesis that these compounds could modulate SOD1 activity, which could be a strategic avenue for ALS therapy development.

TARDBP encodes the TDP-43 protein, and mutations in this gene can lead to the accumulation of protein aggregates, a common feature in ALS pathology. These aggregates disrupt RNA processing and neuronal function, contributing to disease progression. DCTN1 is involved in the axonal transport system, and its mutations can impair the transport of essential molecules and organelles within the neuron, affecting neuronal health and survival. ALS2 mutations have been associated with juvenile-onset ALS and are thought to affect endosomal trafficking and neuronal development. CHMP2B is part of the endosomal-lysosomal system, and its mutations can lead to defective protein degradation, contributing to neurodegenerative processes. SPTLC1 mutations, although less common in ALS, affect sphingolipid metabolism, which is crucial for maintaining the structure and function of neuronal membranes. SETX is involved in DNA repair and RNA processing, where its dysfunction due to mutations can cause genomic instability and contribute to motor neuron death. ANG is involved in angiogenesis and neuroprotection, and its mutations can lead to
decreased angiogenesis and increased susceptibility to stress in motor neurons. VAPB is implicated in the unfolded protein response and lipid transport, and mutations here can disrupt cellular homeostasis. Lastly, SOD1 is one of the most well-studied genes in relation to familial ALS, with its mutations leading to toxic protein aggregation that contributes to motor neuron death through oxidative stress and other mechanisms.

Several studies have focused on these genes in the context of ALS. A study by Qing Liu et al. included these genes in a screening panel for sequence analysis in Chinese patients with familial and sporadic ALS, underscoring their significance in the disease’s genetic profile [18]. Another study ranked the credibility of these genes in relation to ALS, with SOD1 and TARDBP being among the most credible genes associated with the disease [19]. Moreover, multiple recent studies mentioned that TARDBP, SOD1, ANG, VAPB, and CHMP2B are genes associated with ALS, highlighting their importance in ALS research and potential as therapeutic targets [20].

Compounds like Humulene and Buddledin C, identified through molecular docking studies, show promise as potential therapeutic agents due to their high binding affinity to SOD1. Their interactions with SOD1 suggest they may mitigate the toxic effects of SOD1 aggregation, which is a key pathogenic event in familial ALS cases. By potentially reducing oxidative stress and preventing the misfolding of SOD1, these compounds could help maintain neuronal health and slow the progression of ALS. In conclusion, the collective research efforts on these genes provide a foundation for understanding the genetic underpinnings of ALS and reinforce the potential of targeting these pathways with therapeutic compounds like Humulene and Buddledin C. These approaches not only deepen our understanding of the disease but also open new avenues for treatment strategies that could significantly impact patient outcomes in ALS.

To sum up, the integration of genetic, epigenetic, computational, and molecular dynamics studies provides a detailed landscape of ALS-related alterations, paving the way for the development of targeted therapies. These findings not only deepen our understanding of the disease mechanism but also underscore the complex interplay of biological processes contributing to ALS pathogenesis. The potential therapeutic compounds identified, especially their interactions with SOD1, underscore the value of using multifaceted approaches that span multiple levels of biological research to uncover novel treatment avenues for complex diseases like ALS. Expanding on the clinical implications, it is crucial to explore how these discoveries might be translated into practical treatments. The feasibility of developing these targeted therapies hinges on their ability to navigate the rigorous clinical trial process, requiring robust preclinical validation to assess efficacy and safety. Additionally, the challenges associated with delivering these therapies to the affected neuronal populations and overcoming potential off-target effects must be meticulously addressed. By proactively tackling these challenges, we can enhance the potential for these innovative therapies to reach clinical application, offering hope for effective management and possibly altering the course of ALS.

Our integrated bioinformatics analysis has not only deepened our understanding of the molecular underpinnings of ALS but also opened up several promising avenues for therapeutic development. The identification of specific genetic mutations and epigenetic modifications critical to ALS pathogenesis presents potential targets for novel treatment approaches. For instance, targeted gene editing strategies, such as those employing CRISPR-Cas9 technology, could be tailored to correct detrimental genetic mutations in early disease stages. Additionally, our findings suggest that epigenetic therapies, which could reverse abnormal DNA methylation patterns, offer a significant opportunity to modulate gene expression dynamically in response to disease progression. The potential therapeutic compounds identified through molecular docking studies, like Humulene and Buddledin C, provide a starting point for the development of drugs aimed at mitigating key pathological features such as protein aggregation. These bioinformatics-driven insights underscore the necessity of a multidisciplinary approach, combining genetics, molecular biology, and clinical strategies to forge pathways toward personalized medicine in ALS treatment. This
convergence of disciplines promises not only to enhance the precision of therapy but also to improve overall clinical outcomes for patients suffering from this devastating disease.

5. Conclusions

In conclusion, this study provides a comprehensive exploration of the genetic and epigenetic factors contributing to ALS pathogenesis. Our analysis highlights critical mutations in genes such as TARDBP, SOD1, ANG, VAPB, and CHMP2B, which impact neuronal function and survival through mechanisms such as protein aggregation, impaired axonal transport, disrupted endosomal trafficking, and aberrant lipid metabolism. These pathways offer potential targets for therapeutic intervention. Additionally, our findings on DNA methylation patterns within gene promoter regions underscore the importance of epigenetic mechanisms in regulating gene expression and influencing disease progression. The identification of hypermethylation in promoter regions of pivotal genes suggests that epigenetic therapies could provide new avenues for ALS treatment, representing a modifiable aspect of gene regulation with the potential to alter disease trajectories. The discovery of compounds like Humulene and Buddledin C, with high binding affinities to the SOD1 enzyme, illustrates the utility of molecular docking studies in identifying novel therapeutic candidates. These compounds show promise in mitigating the effects of SOD1 aggregation and oxidative stress, which are key aspects of familial ALS, thus offering potential to preserve motor neuron health. Although our study has made significant strides in understanding ALS, it is not without limitations. The complexity of gene–environment interactions in ALS and the variability in genetic expression profiles across individuals pose challenges in translating these findings directly into clinical applications. Future research should focus on overcoming these barriers, possibly through longitudinal studies and enhanced modeling of environmental factors, to refine our understanding of ALS and improve the specificity and efficacy of potential therapies. The path forward in ALS research and therapy involves continued exploration of the intricate biological networks that characterize the disease. By integrating genetic, molecular, and epigenetic approaches, we can advance the development of targeted therapies, enhance early detection methods, and tailor patient-specific treatment strategies. This integrated approach holds profound implications for improving patient care and combating the devastating impacts of ALS.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/sclerosis2030010/s1, Supplementary File S1: Table S1: Genetic variants associated with neurodegenerative disorders. It enlists unique identifiers such as VariationID, AlleleID(s), and dbSNP ID for each variant, offering comprehensive genomic information relevant to ALS; Supplementary File S2: Table S1: Comprehensive overview of biological processes associated with 23 key proteins in ALS. This table categorizes and details the various biological processes in which the identified 23 proteins are predominantly involved, highlighting their roles in the complex network of mechanisms contributing to the pathophysiology of ALS; Supplementary File S3: Table S1: Detailed analysis of molecular functions of 23 ALS-related proteins. This table enumerates the specific molecular functions attributed to the 23 proteins implicated in ALS, illustrating the range of activities they are involved in at the molecular level; Supplementary File S4: Table S1: Distribution of 23 ALS-associated proteins across cellular components. This table provides an exhaustive breakdown of the cellular components where the 23 ALS-associated proteins are predominantly enriched. It underscores the localization and interaction of these proteins within specific cellular structures, elucidating their potential impact on cellular function and integrity in the context of ALS.

Author Contributions: Conceptualization, K.K. and P.V.; methodology, K.K.; software, K.K.; validation K.K.; formal analysis, K.K.; investigation, K.K.; resources, K.K.; data curation, K.K.; writing—original draft preparation, K.K.; writing—review and editing, K.K.; visualization, K.K.; supervision, P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: The data are contained within this article and supplementary files.

Conflicts of Interest: The authors declare no conflict of interest.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.