

Differential Gene Analysis To Identify Risk Genes for Amyotrophic Lateral Sclerosis

Razin M.Rasheed

ABSTRACT

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease that is projected to affect more than 375,000 people by the year 2040. Despite the extensive amount of research, it remains unclear what specific mechanisms are directly contributing to ALS, limiting the number of viable treatments that are available for people with this condition. It is the most common motor neuron disease, with its progression being very fast, with conditions of muscle weakness and slurred speech occurring only a few years before complete paralysis, respiratory failure, and even death. There are many causes that have been proposed, but none of them have led to a cure, only treatments to temporarily extend the patient's life by targeting specific symptoms.

Currently, the genetic drivers of sporadic ALS are poorly understood, which is why new approaches such as those that integrate data from multiple tissues may highlight specific signals that are not derived from a single tissue sample. This study uses differential analysis that has been applied to publicly available datasets to identify key genes that may be associated with ALS. By using sequencing data derived from both spinal cord tissue and the blood, this study aims to maximize the statistical confidence in the genes that are identified as contributing to ALS to allow for future research into the pathology of ALS. Genes demonstrating a statistically significant expression (exceeding a threshold of the FDR or $p < 0.05$) when comparing the datasets are designated as high risk genes. Those top 100 genes were then compared and analyzed based on 3 factors to create a list of 24: (1) how significantly up or down regulated the individual genes were, (2) whether or not the genes were associated with one another or contributed to the same biological process, (3) if they have or haven't been mentioned in prior research. The findings of this study aim to expand on and/or corroborate the current understanding of the genetic role of genes in ALS to allow for further research into the complex neurodegenerative disease and the genes that lead to it. The finding of this study was that ALS is associated with an upregulation of the inflammatory response that results from certain genes. These inflammatory responses lead to damage and degeneration of motor neurons, with the permanent damage to these motor neurons inhibiting normal bodily activity, contributing to the symptoms of muscle weakness, paralysis, and even death. Twenty four of the top 100 genes were found to be part of the same neuron uptake process while also being significantly dysregulated in both the spinal cord and the blood. These genes were part of an inflammatory response pathway and with interferon gamma regulation in the neurons. These genes may serve as future biomarkers and therapeutic targets with this research's validation.

INTRODUCTION

Overview

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disorder that is distinguished by the loss and degeneration of motor neurons that are responsible for muscle movements. The consequences of losing these neurons include muscle weakness, atrophy, stiffness, twitching, paralysis, and ultimately result in respiratory failure (“NIH”). Although cognition tends to be preserved in most cases, the rapid physical decline the patients experience causes an immense amount of psychological and emotional turmoil, with patients experiencing their paralysis while being completely aware. Following the diagnosis, patients are estimated to have a life expectancy of only 2-5 years (“Centers for Disease Control and Prevention”). In rare cases, some patients can live longer depending on the variability of the disease’s progression or the treatments administered.

ALS progresses by first attacking the neurons in the upper motor cortex and the lower brainstem, as well as the spinal cord. As these neurons deteriorate, so does the rate of which communication between the muscles and the nervous system, leading to trouble swallowing, impaired speech, impaired movement, trouble breathing, and paralysis (NINDS, n.d.). This disease often appears in early to middle adulthood but varies widely, with numerous genes having been implicated in causing ALS, with approximately 10% of the cases being directly inherited and the remaining 90% arising from a complex combination of genetic and environmental factors (ALS Association, D. Hartzfeld).

Impact

Despite ALS being considered a rare disease, its impact on families and individuals is substantial. The disease has a lifetime risk of 1 in 350 for men and 1 in 450 for women (Ryan et al.). Roughly 30,000 people in the United States have ALS, with a higher prevalence in men and veterans. It is projected that by the year 2040, 375,000 people around the world will be afflicted with ALS (Vesta et al., 2025). The projected increase is driven by global population aging, implying that ALS diagnoses will continue to grow even if age specific instances stay at their current rate.

Beyond the devastating medical outcome, ALS presents int patients with severe social and economic burdens. Patients, for the rest of their lives, need assisted living technologies and 24.7 caregiving, which are all extremely significant financial burdens on both the patients and their families (Londra et al.,2022). These impacts show the importance of advancing current research aimed at understanding ALS to allow for intervention in the disease in families before the inevitable outcome.

History

ALS symptoms were first identified by Charles Bell in 1924, but ALS was first formally identified and classified as its own condition by French neurologist Jean-Martin Charcot, the father of modern neurology, in 1869 (Foundation Ispen). Early research into ALS consisted of neuropathological

examination of patients. For much of the twentieth century, the disease remained elusive, with it being classified as an idiopathic disease.

Progress into the disease continued through the 20th and 21st centuries with the discovery of the SOD1 Gene, a genetic mutation that is associated with familial ALS (Foundation Ispen). Its discovery in the 1990s is considered a major milestone into the understanding of ALS, with it being direct evidence for a genetic component to ALS. Research cited the discoveries of other genes that have been associated with ALS, including C9orf72, TRADBP, and FUS. The C9orf72 gene in particular has been recognized as the most common genetic cause of ALS. These genes helped to further expand the understanding of ALS as well as revealed common molecular pathways, including protein aggregation and RNA processing dysfunction as causes for the disease. Despite the multitude of findings, mutations in these identified genes have only accounted for a minimal amount of the many ALS cases.

Contributions

Roughly 10% of ALS cases are classified as familial and occur with a high penetrance. The other 90% of cases are sporadic and occur with no clear familial history (Muscular Dystrophy Association, 2015). Sporadic ALS is thought to arise from a complex interplay of multiple genetic and environmental factors such as toxins, lifestyle patterns, and hazardous conditions to all culminate in the disease. Identifying the genetic components of sporadic ALS has proven particularly challenging because of the multifaceted, complex nature of the disease. Large genome wide studies may offer an alternative way of identifying shared signatures across patients, which is a motivation for the approach taken here in this study.

Current treatments

There are currently a very limited number of treatment options for ALS, and the ones available only seek to slow the progression, rather than cure the disease by targeting symptoms experienced by the patients rather than the disease itself. The two primary approved treatments for sporadic ALS are Riluzole and Edaravone, and the main treatment for genetic ALS is Tofersen. These three treatments only provide modest benefits as disease modifying drugs. Riluzole works by dampening the neurotransmitter glutamate and reducing excitotoxicity, reducing the amount of signaling to slow down damage caused to the neurons (Saitoh Y. & Takashi Y., 2020). Edaravone is a newer treatment for ALS that is thought to work by reducing oxidative stress, which is linked to ALS, and helps reduce damage to the neural cells by absorbing free radicals harming the neurons (Cho & Shukla, 2020). For genetic ALS, tofersen works by blocking the mRNA from the mutated SOD1 gene linked with genetic ALS. This protects patients from the toxic buildup of the SOD1 protein (U.S. Food and Drug Administration, 2023). Tofersen is only approved for patients with the SOD1 mutation, which currently is a small fraction of all ALS cases. There currently is no comparable target therapy that exists.

There are also additional treatments for ALS that target the physical symptoms of the disease, such as physical therapy to improve mobility, speech therapy to improve communication, and dietary counseling to help manage swallowing difficulties. However, these therapies offer support to slow the effects of the disease and help aid in the lifestyle challenges with ALS, rather than directly treating the root cause.

Thesis

The overall aim of this study was to apply differential gene expression analysis to publicly available bulk RNA-sequencing datasets to identify candidate risk genes associated with ALS. This was achieved by first setting minimum and maximum r values for the dataset, then graphing the data into volcano plots to find genes that were significantly up or down regulated, pointing towards a significant difference between the ALS and the control populations. Genes exhibiting a statistically significant differential expression across both tissue types were hypothesized to represent as risk genes with a high relevance to ALS pathology. I hypothesized that gene dysregulation in both the central nervous system tissue and the peripheral blood sample would reflect the core progression of the disease, rather than individual tissue effects that would be found by only comparing a single tissue type. This approach also validates the utility of using bioinformatic methods in complex disease gene discovery and will also provide a foundation for future functional validation studies.

METHODS

Dataset Selection

To identify candidate risk genes for ALS, publicly available bulk RNA sequencing datasets were surveyed and selected for analysis. An initial literature search was conducted to identify the studies that performed a bulk RNA sequencing on samples derived from the ALS patients. Publications were screened based on the availability of corresponding sequencing data in public research and the inclusion of either the raw read counts or normalized gene expression values were suitable for analysis.

From this screening process, two studies were selected. These datasets represented RNA sequencing of the peripheral blood mononuclear cells (PBMCs) (Zucca et al., 2019) and the motor neurons from the spinal cord (Masrori et al., 2025) from sporadic and genetically defined ALS patients and control groups. Together, these datasets encompassed multiple ALS relevant cell types and tissue types, enabling a cross dataset comparisons of the gene expression patterns. Sample sizes and group assignments (ALS versus control) were obtained from the original studies' metadata and used as provided.

Differential Gene Expression

Differential expression analysis was performed on the selected publicly available bulk mRNA-sequencing datasets to identify potential risk genes associated with ALS. For each of the selected studies, normalized RNA sequencing counts were used. Before analysis, the low abundance genes were first filtered out by removing those with counts below 5 across all samples to ensure maximum biological relevance and reduce noise from genes that had minimally expressed transcripts.

For each dataset, differential gene expression was determined by calculating the log2 fold change between the ALS samples and controls. Variability within each group was then assessed by computing the standard deviation. Multiple testing corrections were performed using the False Discovery Rate (FDR) method to control for any possible error. Genes were then adjusted, with a p-value less than 0.05 being marked as statistically significant. Significant genes were then ranked based on the magnitude of the fold change to prioritize candidate genes that may have a large effect on the disease.

Following differential expression analysis, the resulting gene lists from each dataset were compared using applications such as STRING and KEGG pathways to analyze the impact of the present genes and their possible contributions to the disease, and then further corroborated with other research papers.

RESULTS

Key biological processes such as defense responses, regulation of interferon gamma production, and inflammatory responses were significantly upregulated in patients who had ALS, with key genes acting as nodes for these symptoms, with upregulation in certain genes seeming to directly be a factor in the progression of ALS.

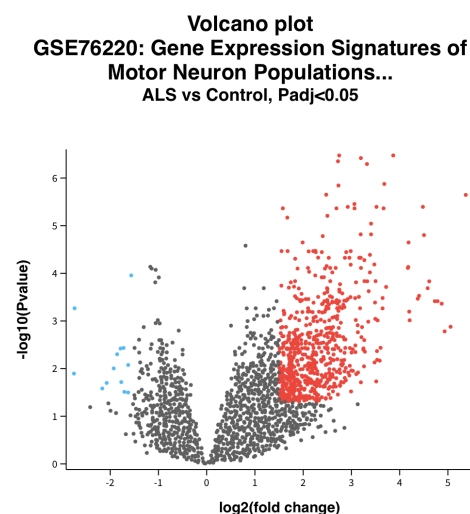


Figure 1. Differentially Expressed Genes in Spine Samples. The diagram above is a volcano plot created from dataset GSE76220, highlighting upregulated and downregulated genes (NCBI) (GEO2R)

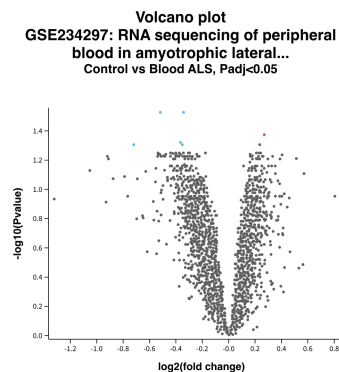


Figure 2. Differentially Expressed Genes from Blood Sample. The diagram above is a volcano plot created from dataset GSE234397, highlighting upregulated and downregulated genes (NCBI) (GEO2R)

The volcano plots Figure 1. and Figure 2 show genes that were statistically up and down regulated in the spinal samples (Fig. 1) and the peripheral blood samples (Fig. 2) compared to the controls. Genes on the graph that appear red are upregulated, while genes that appear blue are downregulated. The X-axis is the Log2 fold change, which is the magnitude of the gene (up or down regulation). The Y-axis is the p-value, which is the likelihood that the notoriety is not random.

From the graph of Figure 1, we notice that a significant number of genes are upregulated and downregulated. This is beneficial, as a significant number of genes are meant to be upregulated in the spinal sample, as it is the source of the disease. In Figure 2. There is a minimal number of genes that were dysregulated. This is by design, as this is meant to help narrow down the possible notable genes found in the sample of the spinal cord. A minimal amount of genes is also expected, as the blood is not the root cause of the disease.



7

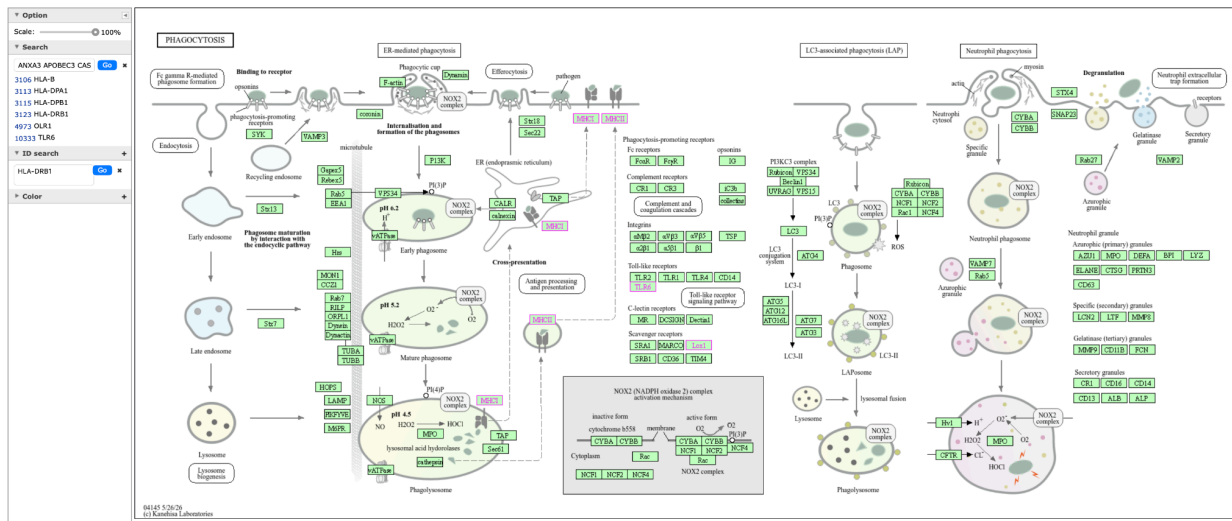


Fig. 4. The chart shown in Figure 4 was created by inputting the 24 highlighted genes from the STRING functional protein network into the KEGG pathway (KEGG Pathway Database)

In Figure 4, the key highlighted genes form the STRING functional protein network that was inputted into the KEGG pathway database to create a chart showing molecular, biological, and chemical interactions taking place in the body from the genes. Not surprisingly, the genes are associated with reuptake of signaling molecules and free radicals between neurons and synapses, which can explain a reason as to the degeneration of neurons in ALS. This can then possibly result in key factors in ALS, such as the inflammatory responses.

The results do not corroborate published findings of the involvement of SOD1 gene mutations in ALS (Foundation Aspen), but other genes that were identified have been previously reported to be involved in the pathophysiology of ALS. From the analysis of 152 patient samples from both the patient's blood and spinal tissue, 24 significant genes were present in both of the samples: ANXA3, APOBEC3, CASP4, CHI3L1, EIF2A, K2 F2RL1, F3, GBP2, HLA-B, HLA-DP, A1 HLA-DPB1, HLA-DRB1, IL27RA, IRF2, LAPTM5, NFATC4, OLR1, PLSCR1, RIPK2, SERPING1, SRPK1, TLR6, TNFRSF1B, VNN1 (Figure 3). Using a network, all of these highlighted genes are involved in inflammation and inflammatory responses in neurons and can explain the symptoms of ALS as well as allude to a root cause (Figure 4).

Discussion

The results corroborate the current understanding that ALS is caused by degeneration in the motor neurons. This degeneration is more specifically caused by upregulation of genes that result in an inflammatory process that inhibits the process of neurons. The inflammatory process is then what causes the damage to them over time, which is seen in the slow process of motor degeneration seen in ALS patients as their motor neurons deteriorate.

Another factor that could be damaging the neurons is the upregulation of genes involved with phagocytosis. Phagocytosis is similar to other inflammatory responses, as it is meant to clear debris and other radicals that can cause harm (Foundation Ipsen). With an increase in phagocytosis, neurons may be engulfed and destroyed, further leading to motor neuron degeneration. This can specifically be attributed to genes such as the HLR subgroup of genes OLR1 and TLR6, which are genes specifically associated with the phagocytosis immune response, causing vascular inflammation and immune signaling(Yao, H. et al.)(Gene Cards).

These results demonstrate that combining different datasets across various tissue types can effectively identify genetic markers associated with ALS. The identification of both known and new risk genes supports the viability of approaches such as these to further understand the genetic causes of ALS. The newly highlighted genes represent unrecognized contributors or effects of ALS, but still need validation.

Conclusion

This study demonstrates the potential of genetic analyses to identify risk genes associated with ALS by using publicly available datasets. By comparing the sequences of both patients with and without ALS, we can identify key genes that are associated with the disease. This work is aimed at highlighting the genetic mechanisms and factors that contribute to ALS that need to be validated further. This work is also aimed at showing the ability of models to be used in understanding and curing genetic diseases and possible implementations that could be made to help improve the process.

The primary value of these findings lies in their clinical and diagnostic utility. By identifying 24 priority candidate genes that show consistent dysregulation across both local neurological and systemic biological compartments, this research provides a concrete foundation for the development of minimally invasive, blood-based diagnostic biomarkers. Currently, diagnosing ALS is a prolonged process of elimination, often taking over a year from the onset of symptoms and causing a devastating delay for a disease with a post-diagnosis life expectancy of just 2 to 5 years. Utilizing these 24 shared genetic targets could help pave the way for early stage screening tools for ALS, enabling clinicians to intercept the disease and help them initiate protective, symptomatic therapies far earlier in the disease's progression.

Furthermore, prevalent familial mutations like *SOD1* open up entirely new avenues for targeted drug discovery in sporadic cases of ALS, which do make up roughly 90% of all ALS diagnoses. These highlighted genes represent active cellular stress signals and disruptions that can be directly targeted by



newer therapeutic drugs. Ultimately, this work is meant to help bridge the gap between what is and isn't currently actionable medicine for these patients. By advancing the overarching goal of identifying therapeutics, it could improve patient outcomes and accelerate the global effort to slow, halt, or ultimately completely cure the complex neurodegenerative conditions such as ALS.

References

- Amyotrophic Lateral Sclerosis (ALS)* | National Institute of Neurological Disorders and Stroke, www.ninds.nih.gov/health-information/disorders/amyotrophic-lateral-sclerosis-als. Accessed 4 Oct. 2026.
- “About Amyotrophic Lateral Sclerosis (ALS).” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, www.cdc.gov/als/abouttheregistrymain/about-amyotrophic-lateral-sclerosis-als.html. Accessed 4 Oct. 2026.
- “History of Amyotrophic Lateral Sclerosis.” *Fondation Ipsen*, 3 Apr. 2025, www.fondation-ipsen.org/podcast/history-of-amyotrophic-lateral-sclerosis/.
- Londral, Ana. “Assistive Technologies for communication empower patients with ALS to generate and self-report health data.” *Frontiers in Neurology*, vol. 13, 26 Apr. 2022, <https://doi.org/10.3389/fneur.2022.867567>.
- “Causes/Inheritance - Amyotrophic Lateral Sclerosis (ALS) - Diseases.” Muscular Dystrophy Association, 24 Oct. 2025, www.mda.org/disease/amyotrophic-lateral-sclerosis/causes-inheritance.
- Amyotrophic Lateral Sclerosis (ALS)* | National Institute of Neurological Disorders and Stroke, www.ninds.nih.gov/health-information/disorders/amyotrophic-lateral-sclerosis-als. Accessed 4 Oct. 2026.
- Ryan, Marie, et al. “Lifetime risk and heritability of amyotrophic lateral sclerosis.” *JAMA Neurology*, vol. 76, no. 11, 1 Nov. 2019, p. 1367, <https://doi.org/10.1001/jamaneurol.2019.2044>.
- Vasta, Rosario, et al. “Amyotrophic lateral sclerosis prevalence projection in 2040: A less rare disease.” *Annals of Clinical and Translational Neurology*, vol. 13, no. 2, 14 Oct. 2025, pp. 379–386, <https://doi.org/10.1002/acn3.70226>.
- Saitoh, Yuji, and Yuji Takahashi. “Riluzole for the treatment of amyotrophic lateral sclerosis.” *Neurodegenerative Disease Management*, vol. 10, no. 6, 27 Aug. 2020, pp. 343–355, <https://doi.org/10.2217/nmt-2020-0033>.
- Center for Drug Evaluation and Research. “FDA Approves Treatment of ALS Associated with a Mutation in the Supero.” *U.S. Food and Drug Administration*, FDA, www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-amyotrophic-lateral-sclerosis-associated-mutation-sod1-gene. Accessed 4 Oct. 2026.
- Williams, Kelly. “RNA Sequencing of Peripheral Blood in Amyotrophic Lateral Sclerosis Reveals Distinct Molecular Subtypes: Considerations for Biomarker Discovery.” *National Center for*

Biotechnology Information, U.S. National Library of Medicine, ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76220. Accessed 4 Oct. 2026.

Williams, Kelly. “RNA Sequencing of Peripheral Blood in Amyotrophic Lateral Sclerosis Reveals Distinct Molecular Subtypes: Considerations for Biomarker Discovery.” *National Center for Biotechnology Information*, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234297. Accessed 4 Oct. 2026.

Cho, HaEun, and Surabhi Shukla. “Role of edaravone as a treatment option for patients with amyotrophic lateral sclerosis.” *Pharmaceuticals*, vol. 14, no. 1, 31 Dec. 2020, p. 29, <https://doi.org/10.3390/ph14010029>.

“Database Commons a Catalog of Worldwide Biological Databases.” *STRING - Database Commons*, ngdc.cncb.ac.cn/databasecommons/database/id/62. Accessed 4 Oct. 2026.

Gene Ontology Amigo, amigo.geneontology.org/amigo/landing. Accessed 4 Oct. 2026.

Kegg Pathway Database, www.genome.jp/kegg/pathway.html. Accessed 4 Oct. 2026.

Hartzfeld, Deborah. “FYI: Familial Amyotrophic Lateral Sclerosis (FALS) and Genetic Testing.” *The ALS Association*, 10 June 2020, www.als.org/navigating-als/resources/familial-amyotrophic-lateral-sclerosis-fals-and-genetic.

Zucca, Susanna, et al. “RNA-seq profiling in peripheral blood mononuclear cells of amyotrophic lateral sclerosis patients and controls.” *Scientific Data*, vol. 6, no. 1, 5 Feb. 2019, <https://doi.org/10.1038/sdata.2019.6>.

Masrori, Pegah, et al. “C9orf72 hexanucleotide repeat expansions impair microglial response in ALS.” *Nature Neuroscience*, vol. 28, no. 11, 14 Oct. 2025, pp. 2217–2230, <https://doi.org/10.1038/s41593-025-02075-1>.

Yao, H., et al. “OLR1 Knockdown Mitigates Traumatic-Hemorrhagic Shock-Induced Liver Injury by Inhibiting the TLR4/Myd88 Pathway and NLRP3 Inflammasome Activation - Molecular Biology.” *SpringerLink*, Pleiades Publishing, 21 July 2025, link.springer.com/article/10.1134/S0026893325600539.

OLR1 Gene Oxidized Low Density Lipoprotein Receptor 1, www.genecards.org/card/OLR1. Accessed 14 Oct. 2026.