Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2017

Research Article

# Missense single nucleotide polymorphisms analysis using bioinformatics tools in *SURF1* and *NDUFV1* genes associated with leigh syndrome

**Aleyna ÖZTÜRK[1], Ömer Faruk KARASAKAL[2,*], Muhsin KONUK[3]**

*[1]Department of Biotechnology, Graduate School of Science, Üsküdar University, İstanbul, Türkiye*
*[2]Medical Laboratory Techniques, Vocational School of Health Services, Üsküdar University, İstanbul, Türkiye*
*[3]Department of Molecular Biology and Genetics, Faculty of Engineering and Natural Sciences, Üsküdar University, İstanbul, Türkiye*

## ARTICLE INFO

*Article history*
Received: 16 August 2024
Revised: 17 October 2024
Accepted: 13 February 2025

**Keywords:**
*In silico*; Leigh Syndrome, NDUFV1; Single Nucleotide Polymorphisms; SURF1

## ABSTRACT

Leigh syndrome is a mitochondrial condition that causes severe neurological damage beginning early in life. By determining the possible harmful consequences of missense single nucleotide polymorphisms in the *SURF1* and *NDUFV1* genes linked to Leigh syndrome, this work seeks to clarify the molecular process of the illness. In the study, deleterious single nucleotide polymorphisms were identified using bioinformatics tools such as SIFT, PolyPhen-2, PredictProtein, PHD-SNP, SNP&GO, PANTHER and Meta-SNP, then their changes on protein integrity were analyzed using I-Mutant 2.0 and MUpro. The findings indicate that certain single nucleotide polymorphisms may significantly disrupt protein structure and stability, increasing the risk of Leigh syndrome. Using STRING and GeneMANIA databases, the interaction networks of these proteins with other proteins were examined, providing a more comprehensive understanding of the biological basis of Leigh syndrome. In our study, 6 single nucleotide polymorphisms in the *SURF1* gene and 11 single nucleotide polymorphisms in the *NDUFV1* gene were identified as potentially deleterious. Our study extends the previous literature on the genetic basis of Leigh syndrome and provides new information on the role of *SURF1* and *NDUFV1* gene mutations in disease pathogenesis. These findings may contribute to the creation of diagnostic and treatment strategies for Leigh syndrome in the future.

**Cite this article as:** Öztürk A, Karasakal ÖF, Konuk M. Missense single nucleotide polymorphisms analysis using bioinformatics tools in *SURF1* and *NDUFV1* genes associated with leigh syndrome. Sigma J Eng Nat Sci 2026;44(2):908–918.

## INTRODUCTION

Leigh syndrome (LS) is a rapidly progressive neurological condition most often occurring in the first years of life. The condition presents with symmetric necrotic lesions in specific regions of the brain, including the brainstem, basal ganglia, and thalamus [1]. Mitochondrial dysfunction is the underlying cause of this disease. Symptoms include growth

***Corresponding author.**
*E-mail address: omerfaruk.karasakal@uskudar.edu.tr

*This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic*

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

retardation, muscle weakness (myopathy), respiratory distress (dyspnea), lactic acidosis and progressive brain damage (encephalopathy) [2]. Single nucleotide polymorphism (SNP) analyses can be used to investigate the genetic foundations of mitochondrial diseases such as LS. These analyses are crucial for determining the pathogenesis, clinical features and genetic characteristics of the disease. The SURF1 cytochrome c oxidase assembly factor *(SURF1)* gene is located on chromosome 9 and encodes a cytochrome c oxidase (COX) assembly factor. This gene is required for the assembly of the cytochrome c oxidase (COX; complex IV) enzyme, which is the terminal enzyme in mitochondrial respiratory chain [3]. Mutations in the SURF1 gene not only disrupt the assembly of COX and its enzymatic activity, resulting in a functional deficiency but also involving molecular mechanisms of Leigh syndrome (LS) [4]. The Human NADH: ubiquinone oxidoreductase core subunit V1 (NDUFV1) is encoded by the NDUFV1 gene on chromosome 11, and is a 51kDa component of nicotinamide adenine dinucleotide hydride-ubiquinone reductase that is essential for complex-I of mitochondrial respiratory chain, and plays an important role in NADH oxidation and ROS production [5]. Mutations in NDUFV1 result in phenotypes including myoclonic epilepsy, leukodystrophy and developmental arrest [6]. In addition to these, mutations in NDUFV1 were reported to be associated with a spectrum of diseases ranging from late-onset LS, ataxia and macrocephaly [7]. The *SURF1* and *NDUFV1* genes were associated with LS by genetic studies in the late 1990s and early 2000s. In particular, mutations in the *SURF1* gene have been directly linked to disease progression [8]. Mutations in the *NDUFV1* gene were first identified at least in 2000 association with LS and it has been observed that this gene is crucial for energy production and can increase the severity of symptoms [9].

This study used bioinformatics tools to examine the impact of SNPs leading to changes in amino acid residues in the *SURF1* and *NDUFV1* genes associated with LS on protein structure, function and stability. The SNPs are the most common genetic polymorphisms, which there is one nucleotide change in a specific position along genome. They tend to segregate into two alleles, which are the most numerous kind of genetic variants in the human genome [10].

Variants within genes can negatively or positively affect a gene's function, protein structure and predisposition to diseases [11]. Public databases record more than 10 million SNPs that are located in the genome about 100–300 base pairs apart. These polymorphisms add to genetic diversity, influencing characteristics, disease risk and response to drugs [12]. *In silico* simulation is the artificial replication of biological functions used to predict the organization and stability of living systems. These tests are often used for the stability of proteins, enzymes, genes, or drugs. *In silico* analysis is a realistic method to study protein, enzyme and gene stability, molecular basis of diseases and also it can be applied in drug development for understanding the molecular interaction. These analyses yield a wealth of useful information for researchers and concomitantly decrease the expense and time involved in laboratory studies [13]. Studies on variants of these genes in the literature are limited. This study is based solely on bioinformatic analyses and does not include experimental validation. Furthermore, our analyses do not include a population-based genetic diversity assessment. It may be useful to validate *in silico* analyses with laboratory experiments to better understand protein stability and biological effects. The distribution of these variants in different populations and their disease associations should be examined. Additionally, developing 3-D models of the SURF1 protein may contribute to comprehensive biological impact analyses.

This study provides an innovative contribution to the literature by comprehensively addressing the effects of SNPs on protein stability, functions and structural properties through multidimensional analyses. In addition, gene-gene and protein-protein interactions were analyzed to reveal new information about the molecular mechanisms of the disease. It was emphasized what role missense mutations in the *NDUFV1* and *SURF1* genes would play in the pathogenesis of Leigh syndrome and how the effects of these mutations could be characterized. These findings may enable the development of new approaches for early diagnosis and treatment strategies of LS in the future.

## MATERIALS AND METHODS

### Data Collection

Genes for analysis were selected based on a comprehensive literature review of LS. SNP identities, locations, nucleotide changes and their associated amino acid changes were obtained using resources from National Center for Biotechnology Information (NCBI). The UniProt platform was consulted to collect amino acid sequences in FASTA format and obtain UniProtKB protein identifiers.

### *In Silico* Software Tools and Evaluation Criteria

Missense SNPs may contribute to changes in protein structure, stability and function, thus the impact of selected genes on proteins was analyzed *in silico* [14]. A number of software utilities were used to examine the potential for missense SNPs in SURF1 and NDUFV1 genes to disrupt protein structure. Only the SNPs which were potentially deleterious were chosen for further analysis. SIFT, PolyPhen-2 (HumVar and HumDiv), PredictProtein, PHD-SNP and SNPs&GO were the primary tools employed in the study followed by PANTHER and Meta-SNP. The software packages use a range of bioinformatics methods to find SNPs that are predicted to have an impact on the conformation or function of the encoded protein.

SIFT predicts whether an amino acid substitution in a protein will have a phenotypic effect and works by

calculating sequence homology, the properties of amino acids and evolutionary information to determine how changes from one state to another impact function [15]. PolyPhen-2, estimates the impact of amino acid substitutions on protein structure and activity using basic physical and evolutionary information. HumVar Model: Used for diagnosing Mendelian diseases and evaluating rare alleles, HumDiv Model: Effective for distinguishing significant mutations within human variations [16]. PredictProtein, enables the prediction of structural and functional characteristics of proteins. This is a system that brings together many programs for interactive protein structure and function studies [17]. Web-based PHD-SNP is designed to predict the possible impact of SNPs in protein coding regions. It is intended to distinguish the disease-related polymorphism from the neutral one [18]. SNPs&GO is a bioinformatics tool that traverses protein functional annotation (Gene Ontology terms) to infer the pathogenicity of single-point variations. It examines protein sequences to determine if a mutation is associated with disease [19]. PANTHER is a publicly available resource that classifies genes and gene products under ontologies by tracing their evolutionary relationship. This resource classifies proteins based on function [20]. Meta-SNP is a computational tool used to predict the pathogenicity of SNPs, by combining several models. It refines the prediction for disease-associated non-synonymous SNVs by integrating the predictions of SNAP, SIFT, PANTHER and PhD-SNP [21].

**Bioinformatics Analysis Steps**

The analyses conducted using these tools and algorithms were divided into the following steps:

**SNP identification**

SNPs in *SURF1* and *NDUFV1* genes were identified using NCBI and UniProtKB databases and their possible impacts on protein structure were examined.

Protein Function and Structural Analysis: SIFT and PolyPhen-2 were employed to predict the impact of amino acid changes on protein function. The predictions from these tools were analysed independently with HumVar and HumDiv to compare the outcome.

**Pathogenicity assessment**

Web-based software PHD-SNP and SNPs&GO were utilized to predict whether the single nucleotide polymorphisms (SNPs) in the present study had potential associations with disease. These analyses included estimating SNP effects with the contribution of Gene Ontology terms.

**Integrated predictions**

Meta-SNP integrated the results of SNAP, SIFT, PANTHER and PhD-SNP, allowing for a more accurate estimation of the disease-causing potential of SNPs. The data obtained as a result of the analyses were filtered and common deleterious SNPs were identified.

**Gene-Gene and Protein-Protein Interactions**

Gene-gene interactions were analyzed utilizing GeneMANIA, an online resource designed to evaluate gene functions and relationships. GeneMANIA is a software tool that provides information about the function of each gene, including pathways, co-localization, genetic interactions, physical relationships and common protein domains [22]. This tool facilitates the study of possible gene interactions.

Protein-protein interactions were evaluated utilizing another platform STRING. STRING improves the accuracy of predicted protein interactions by integrating various data sources, such as experimental datasets and co-expression analyses, thus enabling more reliable bioinformatic analyses [23]. Based on the provided data, possible functional partners of genes were identified.

**Effects of Predicted Deleterıous Single Nucleotide Polymorphisms on Protein Stabilization**

I-Mutant 2.0 and MUpro were employed to assess the effects of potentially damaging SNPs on protein stability. Support vector machine (SVM) technology is the basis of the I-Mutant 2.0, which predicts changes in protein stability induced by mutations. It calculates the change in free energy ($\Delta\Delta G$), which reflects the stabilizing or destabilizing effect of the mutation, based on the protein sequence and structure [24]. MUpro employs SVM to predict changes in protein stability and offers important information about protein behavior under mutational stress by predicting changes in protein stability induced by mutations [25]. These tools are specifically designed to predict the effects of variations at different positions on protein stability.

**Construction of Three-Dimensional Models of Predicted Deleterious Variations**

Three-dimensional models of the *NDUFV1* gene were created using the project HOPE platform, which provides tools to visualize and analyze protein structures [26]. The Project HOPE tool is a platform designed to analyze the structural and functional effects of SNPs or point mutations in proteins. Project HOPE combines data from various web services and databases, including UniProt and the Protein Data Bank (PDB), to generate detailed reports on how a mutation can affect the three-dimensional structure of a protein and its interactions. Project HOPE provided results on amino acid size, charge, hydrophobicity and conservation [27]. However, modeling information for the *SURF1* gene could not be obtained by Project HOPE because the three-dimensional structure of the *SURF1* gene is not available.

## RESULTS AND DISCUSSION

**Prediction Results of Deleterious Single Nucleotide Polymorphisms Using *In-Silico* Methods**

5321 SNPs were detected in the *SURF1* gene, of which 734 were identified as missense variations. Among these

missense SNPs, 602 amino acid substitutions were evaluated. 3417 SNPs were detected in the *NDUFV1* gene, of which 431 were missense SNPs. Within this subset, 547 amino acid changes were analyzed.

Using all software tools, 6 deleterious SNPs (rs28933402, rs121918658, rs145615218, rs201492662, rs373551988, rs369707095) were identified in the *SURF1* gene (Table 1). These 6 missense SNPs include a total of 6 variants (G124E, Y274D, Y117S, R118W, R137W and G159W) predicted to be deleterious. In the *NDUFV1* gene (Table 1), 11 deleterious SNPs (rs121913661, rs138526825, rs138583785, rs150966634, rs201787156, rs11540007, rs139966879, rs143866203, rs201382784, rs375322599, rs375897089) were identified, with a total of 12 variants (E214K, R88G, R88C, R224C, R386C, E408K, G87S, G146V, R88H, G306C, G321S and E246K) predicted to be deleterious.

**Results of Protein Stabilization**

The effect of SNPs classified as deleterious by all the software tools on the stability of proteins was evaluated utilizing I-Mutant 2.0 and MuPro platforms. An overview of these results is provided in Table 1 and Table 2.

**Results of Gene-Gene and Protein-Protein Interactions**

Gene-gene and protein-protein interactions involving the *SURF1* and *NDUFV1* genes, as well as their connections with other genes, were examined using GeneMANIA. Figure 1 depicts these gene interactions, highlighting aspects such as co-expression, localization, genetic and physical interactions, predicted pathways and common protein domains. Additionally, the STRING database was used to map out the connections between *SURF1* and *NDUFV1* and ten related proteins, as illustrated in Figure 1.

**Results of Construction of Three-Dimensional Models of Predicted Deleterious Variations**

Each amino acid has its own characteristics, including size, charge, and hydrophobicity. Each of these characteristics can be affected by mutations. Hydrophobicity mutations can influence hydrogen bonding, while charge mutations can affect ionic bonds. Size variation between

**Table 1.** Possible prediction results of single nucleotide polymorphisms in *SURF1* gene using software tools

| Analysis | rs28933402 | rs121918658 | rs145615218 | rs201492662 | rs373551988 | rs369707095 |
|---|---|---|---|---|---|---|
| Nucleotide Change | C>T | A>C | T>A / T>G | T>A / T>C | G>A | C>A |
| Amino Acid Change | G124E | Y274D | Y117S | R118W | R137W | G159W |
| PredictProtein Result | E | E | E | E | E | E |
| PredictProtein Score | 90 | 95 | 81 | 68 | 91 | 83 |
| SIFT Result | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt |
| SIFT Score | 0.000 | 0.000 | 0.001 | 0.008 | 0.043 | 0.001 |
| PolyPhen-2 HumDiv Result | PH | PH | PH | PH | PH | PH |
| PolyPhen-2 HumDiv Score | 1000 | 1000 | 1000 | 1000 | 0.999 | 1000 |
| PolyPhen-2 HumVar Result | PH | PH | PH | PH | PH | PH |
| PolyPhen-2 HumVar Score | 1000 | 1000 | 0.995 | 0.998 | 0.948 | 1000 |
| SNP&GO Result | Ds | Ds | Ds | Ds | Ds | Ds |
| SNP&GO RI | 9 | 9 | 9 | 7 | 8 | 9 |
| PHD-SNP Result | Ds | Ds | Ds | Ds | Ds | Ds |
| PHD-SNP RI | 8 | 10 | 9 | 9 | 8 | 8 |
| PANTHER Result | PH | PH | PH | PH | PH | PH |
| META-SNP Result | Ds | Ds | Ds | Ds | Ds | Ds |
| META-SNP Score | 0.854 | 0.875 | 0.807 | 0.824 | 0.803 | 0.834 |
| META-SNP RI | 7 | 8 | 6 | 6 | 6 | 7 |
| I-Mutant Result | Dec | Dec | Dec | Dec | Dec | Dec |
| I-Mutant Score | 1 | 4 | 8 | 8 | 7 | 7 |
| MUpro Result | Dec | Dec | Dec | Dec | Dec | Dec |
| MUpro DDG Value | -0,86483608 | -1,3281959 | -1,5154971 | -0,98251135 | -0,39536249 | -0,065944524 |

E: Effective; Dlt: Deleterious; Ds: Disease; Dec: Decrease; PH: Potentially harmful.

**Table 2.** Possible prediction results of single nucleotide polymorphisms in *NDUFV1* gene using software tools

| Analysis | s121913661 | s138526825 | s138526825 | s138583785 | s150966634 | S201787156 | s11540007 | s139966879 | s143866203 | s201382784 | s375322599 | s375897089 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nucleotide change | G>A | C>G / C>T | C>G / C>T | C>A / C>T | C>T | G>A | G>A | G>A / G>T | G>A | G>C / G>T | G>A | G>A |
| Amino Acid change | 214K | 88G | 88C | 224C | 386C | 408K | 87S | 146V | 88H | 306C | 321S | 246K |
| Predict protein result | E | E | E | E | E | E | E | E | E | E | E | E |
| PredictProtein score | 70 | 93 | 80 | 73 | 87 | 57 | 73 | 71 | 73 | 57 | 79 | 88 |
| SIFT result | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt | Dlt |
| SIFT score | 0.013 | 0.001 | 0.000 | 0.003 | 0.000 | 0.002 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 |
| PolyPhen-2 HumDiv result | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH |
| PolyPhen-2 HumDiv score | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 0.977 | 1000 | 1000 | 1000 | 0.999 |
| PolyPhen-2 HumVar result | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH |
| PolyPhen-2 HumVar score | 0.986 | 1000 | 1000 | 1000 | 0.999 | 0.999 | 1000 | 0.931 | 1000 | 0.997 | 1000 | 0.999 |
| SNP&GO result | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds |
| SNP&GO RI | 0.93 | 0.89 | 0.92 | 0.95 | 0.95 | 0.54 | 0.72 | 0.93 | 0.86 | 0.72 | 0.89 | 0.93 |
| PHD-SNP result | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds |
| PHD-SNP RI | 9 | 8 | 8 | 9 | 9 | 1 | 4 | 9 | 7 | 4 | 8 | 9 |
| PANTHER result | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH | PH |
| META-SNP result | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds | Ds |
| META-SNP score | 0.715 | 0.875 | 0.912 | 0.885 | 0.889 | 0.725 | 0.884 | 0.771 | 0.875 | 0.932 | 0.853 | 0.888 |
| META-SNP RI | 4 | 8 | 8 | 8 | 8 | 5 | 8 | 5 | 8 | 9 | 7 | 8 |
| I-Mutant result | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec |
| I-Mutant score | 5 | 6 | 5 | 6 | 4 | 8 | 1 | 4 | 5 | 6 | 8 | 7 |
| MUpro result | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec | Dec |
| MUpro DDG value | -1.0879523 | -1.5640386 | -0.72560616 | -0.2860499 | -0.47840161 | -1.8670333 | -0.97483618 | -0.60336513 | -1.1489642 | -0.83887817 | -0.61125374 | -0.99532374 |

E: Effective; Dlt: Deleterious; Ds: Disease; Dec: Decrease; PH: Potentially Harmful

normal and mutant-type amino acids can impair the formation of a hydrogen bond. These characteristics can impair protein function by causing the loss of ligand binding. Mutations can also influence local structure and binding sites. Moreover, mutations can impair interactions between crucial domains of proteins, thereby influencing protein function. The differences in hydrophobicity, electric charge, and molecular size of normal and mutated amino acids in the *SURF1* and *NDUFV1* genes were evaluated using Project HOPE. Details of these outcomes are provided in Table 3. Moreover, models and annotations for deleterious mutations caused by changes in amino acids in the *SURF1* and *NDUFV1* genes and three-dimensional structures for the *NDUFV1* gene were created using Project HOPE (Table 4).

**Biological Implications of the Findings**

The findings indicate that certain SNPs identified in the *SURF1* and *NDUFV1* genes may negatively affect protein stability, which may lead to impaired mitochondrial functions and contribute to the progression of LS. Specifically, variations in functional protein interactions can be triggered by some of the identified deleterious SNPs. These can lead to problems in energy production in mitochondria and the emergence of symptoms associated with cellular energy deficiency [28]. The results can be evaluated to develop more specific treatment and follow-up strategies for LS and to provide important clues regarding the potential impact of genetic differences on the underlying molecular mechanisms of the condition.

Studying SNPs is vital for elucidating the genetic foundations of diseases, assessing disease risk and predicting individual responses to treatment. Since there are so many known SNPs in the literature, conducting wet-lab experiments aimed at understanding the biological effects of each SNP is quite labor-intensive. Therefore, bioinformatics analyses are used as an effective pre-screening strategy to prioritize potentially pathogenic SNPs before proceeding to wet lab-based experimental validation studies [29].

It is predicted that the E214K and G146V variants identified in the NDUFV1 gene may negatively affect the NADH oxidation process by disrupting the structural stability and enzymatic activity of mitochondrial Complex I. Such structural and functional disruptions can lead to a decrease in oxidative phosphorylation efficiency and a reduction in cellular energy production. This disruption may lead to disruptions in electron transport system and accumulation of oxidative species [5,7]. Increased oxidative imbalance and metabolic deficiency may contribute to the pathogenesis of the disease by causing cellular damage and neurodegeneration.

It has been reported in the literature that P119L mutation in *SURF1* gene causes defects in COX mechanism and is associated with LS [4]. This mutation affects the structural stability of the COX complex, similar to the G124E and G159W mutations detected in our study. All three mutations lead to a decrease in COX activity and cause disruptions in energy metabolism. R88C and R386C mutations previously reported in *NDUFV1* gene are also related

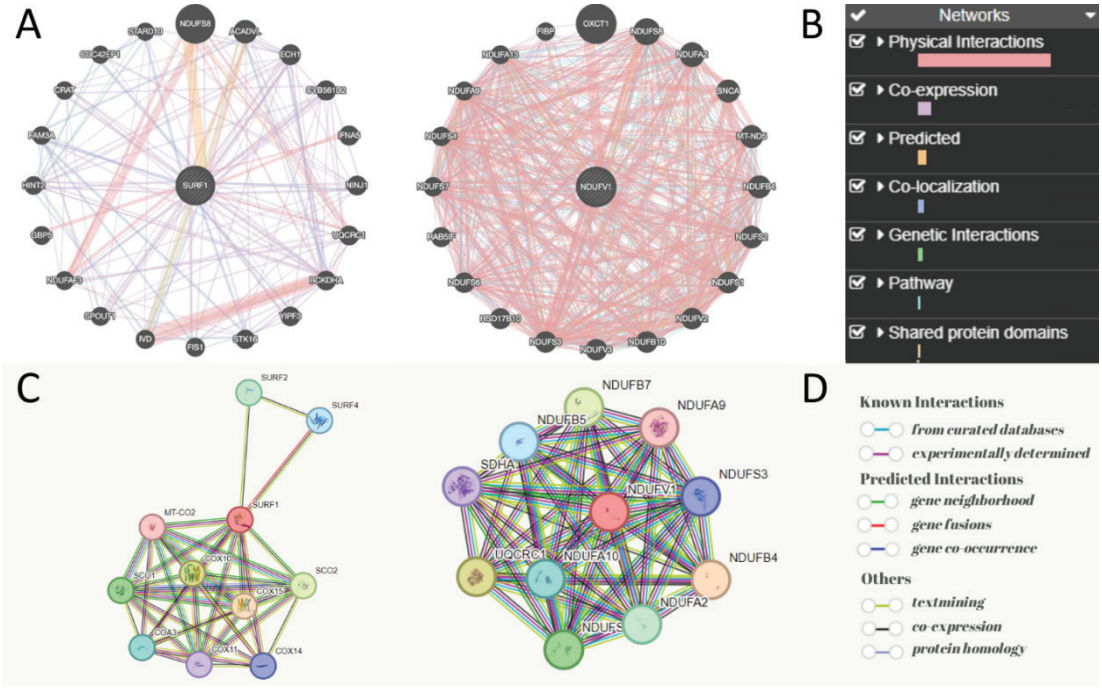



**Figure 1A.** *SURF1*, **B.** *NDUFV1* gene interaction map (GeneMANIA), **C.** *SURF1*, **D.** *NDUFV1* protein interaction map (STRING).

to complex I deficiency and LS [6,30]. These mutations may disrupt protein function by mechanisms similar to the E214K and G146V mutations we detected. In both cases, the decrease in complex I activity may affect cellular energy production and contribute to the clinical symptoms of the disease.

The study by Özkan et al., examining the effects of genetic variations on the disease, reveals that there is no significant relationship between prostate cancer and the rs7501939 polymorphism. This finding shows that SNP analyses are important for understanding both disease risk and disease characteristics [31].

The literature reported that Y204C and E214K variants in the *NDUFV1* gene have deleterious effects [32]. Another study mentioned the disease-associated effects of R88G, R386C and R199P variants in the *NDUFV1* gene; however, our study did not identify R199P as a common deleterious variant [30]. For the *SURF1* gene, the P119L variant has been documented in the literature, but was not found in our data in the NCBI database. In this context, the results of this study are consistent with the data in the literature and reveal the deleterious effects of SNPs in more detail.

When examining other mitochondrial diseases, MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes) disease is associated with the A3243G mutation in the MT-TL1 gene. This mutation disrupts the structure of mitochondrial transfer ribonucleic acid (tRNA) and negatively affects energy production. Mutations in LS are generally of nuclear deoxyribonucleic acid (DNA) origin and affect complexes I and IV of the mitochondrial respiratory chain, leading to disruptions in adenosine triphosphate (ATP) production. While *in silico* analyses predict the structural effects of this mutation by examining the tRNA structure in MELAS, mutations that disrupt protein function in LS are examined with protein modeling [33].

LHON (Leber's Hereditary Optic Neuropathy) is associated with mutations in the *MT-ND1*, *MT-ND4* and *MT-ND6* genes in mitochondrial DNA. One of the most common mutations in this disease, G11778A, disrupts the function of mitochondrial complex I and prevents energy production in optic nerve cells, leading to vision loss. *NDUFV1* mutations in LS similarly disrupt complex I function, but unlike LHON, neurological symptoms are more common and usually appear in infancy [34].

The difference in the size of mutant amino acids and wild-type amino acids can greatly influence protein structure and function. Variations in the size of amino acids can influence the overall structure and function of proteins [35].

Glycine, the smallest amino acid with a side chain chain made up of only one hydrogen atom, enhances the flexibility

**Table 3.** Amino acid changes of specific single nucleotide polymorphisms in *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool, with effects of these changes on attributes like such as size, charge and hydrophobicity for wild-type and mutant-type

| Gene name | SNP ID | Amino acid substitution | Wild-type | | | Mutant-type | | |
|---|---|---|---|---|---|---|---|---|
| | | | Size | Charge | Hydrophobicity | Size | Charge | Hydrophobicity |
| *SURF1* | **rs28933402** | **G124E** | < | Neutral | > | > | Negative | < |
| | **rs121918658** | **Y274D** | > | Neutral | > | < | Negative | < |
| | **rs145615218** | **Y117S** | > | - | - | < | - | - |
| | **rs201492662** | **R118W** | < | Positive | < | > | Neutral | > |
| | **rs373551988** | **R137W** | < | Positive | < | > | Neutral | > |
| | **rs369707095** | **G159W** | < | - | < | > | - | > |
| *NDUFV1* | **rs121913661** | **E214K** | < | Negative | - | > | Positive | - |
| | **rs138526825** | **R88G** | > | Positive | < | < | Neutral | > |
| | **rs138526825** | **R88C** | > | Positive | < | < | Neutral | > |
| | **rs138583785** | **R224C** | > | Positive | < | < | Neutral | > |
| | **rs150966634** | **R386C** | > | Positive | < | < | Neutral | > |
| | **rs201787156** | **E408K** | < | Negative | - | > | Positive | - |
| | **rs139966879** | **G146V** | < | - | < | > | - | > |
| | **rs143866203** | **R88H** | < | Positive | - | > | Neutral | - |
| | **rs201382784** | **G306C** | < | - | < | > | - | > |
| | **rs375322599** | **G321S** | < | - | - | > | - | - |
| | **rs375897089** | **E246K** | < | Negative | - | > | Positive | - |

of protein structures. The flexibility of protein structures is crucial for the function of the proteins encoded by the SURF1 and *NDUFV1* genes, and the substitution of glycine with another amino acid can impair this flexibility [26]. For example, G124E, G159W in *SURF1* and G87S, G146V, G306C and G321S in *NDUFV1* represent substitutions of glycine with other amino acids that may affect protein function. Similarly, changes to or from glycine may affect the required rigidity at certain positions, thus affecting both function and structure. Table 4 shows the variations in electrical charge among wild and mutant-type residues. Proteins may form ionic interactions between oppositely charged residues, known as salt bridges, due to electrostatic forces. These salt bridges contribute to the protein's structure and interactions with other biomolecules. In an aqueous environment, certain ionic interactions occur on the protein surface, while others happen on the inside [26]. For example, in the E214K variant of the *NDUFV1* gene, the hydrogen bond between the lysine at position 219 and the arginine at position 224 may not reoccur due to the smaller size and different electrical charge of the new residue. Variations in hydrophobicity among the residues are also presented in Table 4.

As a result of HOPE analyses, Y274D variant in the *SURF1* gene, numbered rs121918658, identified in our study, was found to cause the substitution of aspartic acid instead of tyrosine and COX deficiency. These findings are consistent with those reported by [36] study, which is consistent with the destabilizing effects of the mutation on the COX complex.

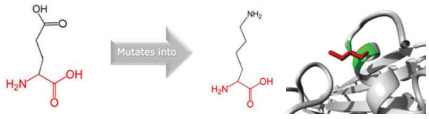
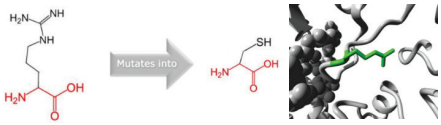
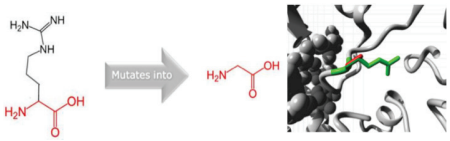
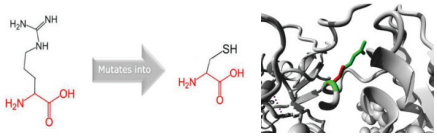
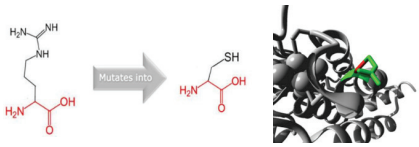
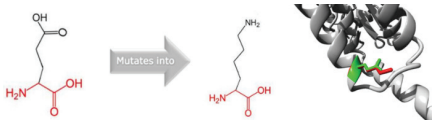
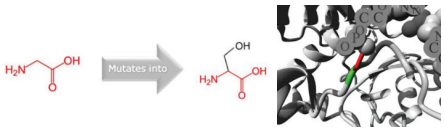
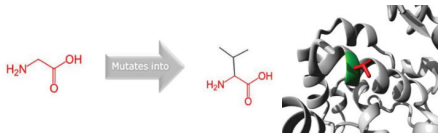
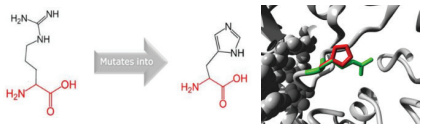
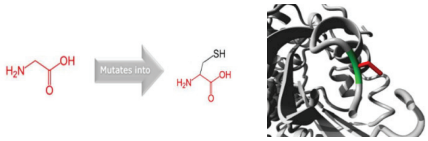
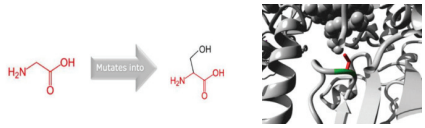
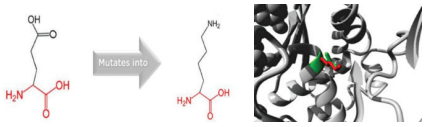
E214K variant numbered rs121913661 in the *NDUFV1* gene, identified in our study, was determined to be deleterious as a result of our analysis. Additionally, this variant was described by [37] study, it was detected in Leigh syndrome patients with complex 1 deficiency. The data in this experimental study are parallel to the data we obtained in our study.

The findings show that these SNPs may disrupt protein stability and lead to disruptions in mitochondrial functions, which may induce the neurological symptoms of the disease. The study can be an important data source for determining genetic risk factors and personalized treatment approaches. The information obtained can shed light on new research areas for the understanding of LS pathogenesis and mitochondrial diseases.

**Table 4.** Models and explanations of the *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool

| Genes | SNP ID | Modelling | Explanation |
|---|---|---|---|
| ***SURF1*** | **rs28933402** |  | Glycine → Glutamic acid at position 124 |
| | **rs121918658** |  | Tyrosine → Aspartic acid at position 274 |
| | **rs145615218** |  | Tyrosine → Serine at position 117 |
| | **rs201492662** |  | Arginine → Tryptophan at position118 |
| | **rs373551988** |  | Arginine → Tryptophan at position 137 |
| | **rs369707095** |  | Glycine → Tryptophan at position 159 |

**Table 4.** Models and explanations of the *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool

| Genes | SNP ID | Modelling | Explanation |
|---|---|---|---|
| *NDUFV1* | **rs121913661** |  | Glutamic acid → Lysine at position 214 |
| | **rs138526825** |  | Arginine → Glycine at position 88 |
| | **rs138526825** |  | Arginine → Cysteine at position 88 |
| | **rs138583785** |  | Arginine → Cysteine at position 224 |
| | **rs150966634** |  | Arginine → Cysteine at position 386 |
| | **rs201787156** |  | Glutamic acid → Lysine at position 408 |
| | **rs11540007** |  | Glycine → Serine at position 87 |
| | **rs139966879** |  | Glycine → Valine at position 146 |
| | **rs143866203** |  | Arginine → Histidine at position 88 |
| | **rs201382784** |  | Glycine → Cysteine at position 306 |
| | **rs375322599** |  | Glycine → Serine at position 321 |
| | **rs375897089** |  | Glutamic acid → Lysine at position 246 |

Green colour: Wild-type residue; Red colour: Mutant-type residue.

## CONCLUSION

In this study, we evaluated the effects of SNPs causing amino acid substitutions in the SURF1 and NDUFV1 genes on the protein structure, function, and stability in LS, using bioinformatics tools for analysis. The findings obtained from our study are expected to serve as a valuable data source for both experimental and bioinformatics research. Future studies should include experimental studies in cellular and animal models to confirm the effects of these single nucleotide polymorphisms on protein function. In particular, their effects on protein-protein interactions, cellular energy metabolism and mitochondrial functions may be detailed with biochemical experiments. In addition, genotype-phenotype analyses may be performed in large patient cohorts to investigate the clinical biomarker potential of these deleterious single nucleotide polymorphisms. Another crucial step is the identification of targeted treatment strategies for these variants. For instance, the identification of small molecules or gene therapy strategies that could enhance the stability of mutant proteins could help reduce the symptoms of the patients. In this case, the clinical interpretation of the results and the identification of targeted treatment approaches could offer major breakthroughs in the treatment of mitochondrial disorders like LS.

## ACKNOWLEDGEMENTS

This study was produced from the master's thesis of the first author, which was accepted at Üsküdar University, Department of Biotechnology.

## AUTHORSHIP CONTRIBUTIONS

AÖ: Data collection, *in silico* analysis, writing—original draft preparation. KÖF: Organizing the research, designing the research and methodology, writing (review and editing). MK: Organizing the research, designing the research and methodology, writing (review and editing).

## DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

## CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ETHICS

There are no ethical issues with the publication of this manuscript.

## STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

## REFERENCES

[1] Chae JH, Lee JS, Kim KJ, Hwang YS, Hirano M. Biochemical and genetic analysis of Leigh syndrome patients in Korea. Brain Dev 2008;30:387–390. [CrossRef]

[2] Johnson SC, Yanos ME, Kayser EB, Quintana A, Sangesland M, Castanza A, et al. MTOR inhibition alleviates mitochondrial disease in a mouse model of Leigh syndrome. Science 2013;342:1524–1528. [CrossRef]

[3] Péquignot MO, Desguerre I, Dey R, Tartari M, Zeviani M, Agostino A, et al. New splicing-site mutations in the SURF1 gene in Leigh syndrome patients. J Biol Chem 2001;276:15326–15329. [CrossRef]

[4] Pecina P, Čapková M, Chowdhury SKR, Drahota Z, Dubot A, Vojtíšková A, et al. Functional alteration of cytochrome c oxidase by SURF1 mutations in Leigh syndrome. Biochim Biophys Acta Mol Basis Dis 2003;1639:53–63. [CrossRef]

[5] Zanette V, Valle D, Telles BA, Robinson AJ, Monteiro V, Santos MLSF, et al. Ndufv1 mutations in complex I deficiency: Case reports and review of symptoms. Genet Mol Biol 2021. [CrossRef]

[6] Tsang WY, Sayles LC, Grad LI, Pilgrim DB, Lemire BD. Mitochondrial respiratory chain deficiency in Caenorhabditis elegans results in developmental arrest and increased life span. J Biol Chem 2001;276:32240–32246. [CrossRef]

[7] Incecik F, Herguner O, Besen S, Bozdoğan S, Mungan N. Late-onset Leigh syndrome due to NDUFV1 mutation in a 10-year-old boy initially presenting with ataxia. J Pediatr Neurosci 2018;13:205. [CrossRef]

[8] Péquignot MO, Dey R, Zeviani M, Tiranti V, Godinot C, Poyau A, et al. Mutations in the SURF1 gene associated with Leigh syndrome and cytochrome c oxidase deficiency. Hum Mutat 2001;17:374–381. [CrossRef]

[9] Lake NJ, Compton AG, Rahman S, Thorburn DR. Leigh syndrome: One disorder, more than 75 monogenic causes. Ann Neurol 2016;79:190–203. [CrossRef]

[10] Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 2004;14:414–425. [CrossRef]

[11] Wang Z, Huang C, Lv H, Zhang M, Li X. In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. PLoS One 2020;15:e0227859. [CrossRef]

[12] Zhu J, Qiu C, Palla M, Nguyen T, Russo JJ, Ju J, et al. A microfluidic device for multiplex single-nucleotide polymorphism genotyping. RSC Adv 2014;4:4269–4277. [CrossRef]
[13] Merlino A, Graziano G, Mazzarella L. Structural and dynamic effects of α-helix deletion in Sso7d: Implications for protein thermal stability. Proteins 2004;57:692–701. [CrossRef]
[14] Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–3814. [CrossRef]
[15] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–1082. [CrossRef]
[16] Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009;19:1553–1561. [CrossRef]
[17] Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein – An open resource for online prediction of protein structural and functional features. Nucleic Acids Res 2014;42:W337–W343. [CrossRef]
[18] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009;30:1237–1244. [CrossRef]
[19] Bhardwaj A, Dhar YV, Asif MH, Bag SK. In silico identification of SNP diversity in cultivated and wild tomato species: Insight from molecular simulations. Sci Rep 2016;6:38715. [CrossRef]
[20] Alizadehmohajer N, Zahedifar S, Sohrabi E, Shaddel Basir S, Nourigheimasi S, Falak R, et al. Using in silico bioinformatics algorithms for the accurate prediction of the impact of spike protein mutations on pathogenicity, stability, and functionality of SARS-CoV-2. Biochem Genet 2023;61:778–808. [CrossRef]
[21] Ali AAH, Ibrahim M. In silico analysis of single nucleotide polymorphisms in human CCR5 gene. bioRxiv 2020. Preprint. doi: https://doi.org/10.1101/2020.11.14.382739.
[22] Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. Nucleic Acids Res 2018;46:W60–W64. [CrossRef]
[23] von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, et al. STRING: Known and predicted protein-protein associations integrated and transferred across organisms. Nucleic Acids Res 2005;33:D433–D437. [CrossRef]
[24] Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33. [CrossRef]
[25] Pan Q, Nguyen TB, Ascher DB, Pires DEV. Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures. Brief Bioinform 2022;23. [CrossRef]
[26] Venselaar H, te Beek TAH, Kuipers RKP, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases: An e-Science approach with life scientist friendly interfaces. BMC Bioinform 2010;11:bbac025. [CrossRef]
[27] Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol 2008;65:45–53. [CrossRef]
[28] Lee IC, Chiang KL. Clinical diagnosis and treatment of Leigh syndrome based on SURF1 genotype and phenotype. Antioxidants 2021;10:1950. [CrossRef]
[29] Bénit P, Chretien D, Kadhom N, De Lonlay-Debeney P, Cormier-Daire V, Cabral A, et al. Large-scale deletion and point mutations of the nuclear NDUFV1 and NDUFS1 genes in mitochondrial complex I deficiency. Am J Hum Genet 2001;68:1344–1352. [CrossRef]
[30] Marin SE, Mesterman R, Robinson B, Rodenburg RJ, Smeitink J, Tarnopolsky MA. Leigh syndrome associated with mitochondrial complex I deficiency due to novel mutations in NDUFV1 and NDUFS2. Gene 2013;516:162–167. [CrossRef]
[31] Özkan E, Erdemir A, Törer BD, Taşçı AI, Baskın Y, Ellidokuz H, et al. Genotyping and analysis of rs7501939 polymorphism for prostate cancer. Sigma J Eng Sci 2015;6:101–107.
[32] Kose M, Canda E, Kagnici M, Aykut A, Adebali O, Durmaz A, et al. SURF1 related Leigh syndrome: Clinical and molecular findings of 16 patients from Turkey. Mol Genet Metab Rep 2020;25. [CrossRef]
[33] El-Hattab AW, Adesina AM, Jones J, Scaglia F. MELAS syndrome: Clinical manifestations, pathogenesis, and treatment options. Mol Genet Metab 2015;116:4–12. [CrossRef]
[34] Engvall M, Kawasaki A, Carelli V, Wibom R, Bruhn H, Lesko N, et al. A novel mutation in the mitochondrial MT-ND5 gene associated with Leber hereditary optic neuropathy. Front Neurol 2021;12:652590. [CrossRef]
[35] Biro JC. Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. Theor Biol Med Model 2006;3. [CrossRef]
[36] Munnich A, Rustin P. Cytochrome c oxidase deficiency. Am J Med Genet Semin Med Genet 2001;106:46–52. [CrossRef]
[37] Quaio CRDAC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet 2020;184:955–964. [CrossRef]