Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v31i1.5166

IN-SILICO ANALYSIS OF DELETERIOUS SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS) OF LEUKEMIA INHIBITORY FACTOR (LIF), AND THEIR CONFORMATIONAL PREDICTIONS

Maham Afzal[1], Umer Ali[2], Adeel Riaz[3], Fouzia Tanvir[4], Asif Bilal[5], Sibtain Ahmad[6*]

[1,3] Department of Biochemistry, University of Okara, Okara Pakistan
[2]Department of Biological Sciences, Tennessee State University, Nashville, TN 37209, USA
[4,5]Department of Zoology, University of Okara, Okara Pakistan
[6*]Institute of Animal and Dairy Sciences, Faculty of Animal Husbandry, University of Agriculture, Faisalabad, Pakistan
[6*]Center for Advanced Studies Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad, Pakistan

***Corresponding Author:** Sibtain Ahmad
[*]Institute of Animal and Dairy Sciences, Faculty of Animal Husbandry, University of Agriculture, Faisalabad, Pakistan Email: dr_sibtainhmd6@uaf.edu.pk

Received: September 18, 2023 Revised: November 29, 2023 Accepted: December 12, 2023
Published: January 02, 2024

Abstract
Leukemia inhibitory factor (LIF) is a multifunctional gene belonging to the interleukin-6 cytokine family. It plays crucial roles in various biological processes such as neuron development, wound healing, maintenance of adrenocorticotropic hormonal secretions in the pituitary glands, reproductive system, and alveolus development. Previous studies have associated LIF polymorphisms with female infertility, schizophrenia (SCZ), and osteoporosis. However, comprehensive computational analyses examining the functional and structural impacts of damaging non-synonymous single-nucleotide polymorphisms (nsSNPs) in LIF have not been conducted. The main objective of this study was to identify and classify nsSNPs that have the most detrimental effects on the LIF gene. A total of nine deleterious mutations (C156F, C153G, L147P, Y111C, Q70H, Y66C, Y66H, T120N, and V164M) were detected, which resulted in altered protein structure. Subsequently, these deleterious mutations were assessed for potential post-translational modification sites using molecular docking and molecular dynamic simulation techniques. The results revealed that the C156F mutant displayed greater conservation and structural dissimilarity compared to the other mutants. Docking analysis demonstrated that EC330 inhibits LIF/LIF-R signaling, thereby impeding LIF's tumor-promoting effects. This finding suggests that EC330 could be a potential candidate for targeted cancer therapy in cases where LIF is overexpressed in malignancies.

Keywords: LIF, nsSNPs, Molecular docking, Molecular dynamic simulation

1. INTRODUCTION

Leukemia inhibitory factor (LIF) is a monomeric glycoprotein that is frequently subject to glycosylation (1). The glycosylated LIF protein has a molecular weight of 20 - 25 kDa bilal(1, 2). The LIF

protein, consisting of 202 amino acids, undergoes post-translational modifications that result in the removal of 22 amino acids from the N-terminus, leading to its conversion into a 20kDa form(3). These modifications are crucial as LIF possesses multiple potential N-glycosylation sites, allowing for significant alterations after translation (4). Studies utilizing nuclear magnetic resonance and x-ray crystallography have revealed the structural characteristics of LIF, demonstrating its similarity to other cytokines within the IL-6 family (3). Specifically, LIF adopts a four-helix bundle conformation, with Helix A initiating at Leu44 (residue 22 in the mature chain) and forming covalent connections with the N-terminal region of helix 3 through two disulfide bonds (Cys34-156Cys and Cys40-153Cys). The N-terminus is crucial for binding of receptor (5).The Helix D is linked between helices A and B by a third disulfide bond. (6).

LIF overexpression in tumor tissue has been linked to oral squamous cell carcinoma (7), chordomas (8), pancreatic adenocarcinoma (9), nasopharyngeal carcinoma (10) renal (11), cervical (11, 12), Breast (13) and skin cancer (14). It is also shown in body fluids (15) as well as other tissues including cardiac muscle (16), thymus (17), hypophysis (18), lungs (17), kidney (19), neuronal tissue, involved in inflammation (20), autoimmune diseases (21) and blastocyst implantation (22). LIF exerts its biological effects by activating various signaling pathways, including the Janus kinase/signal transducer and activator of transcription 3 (JAK/STAT3) pathway, Janus tyrosine kinase pathway, phosphoinositide 3-kinase (PI3K) signaling pathway, and p44/42 mitogen-activated protein kinase (ERK1/2) pathway. These pathways play critical roles in mediating the cellular responses and downstream effects of LIF signaling.(23). Melanomas (24), stimulated monocytes and T-lymphocytes (25), immune cells (24) and fibroblasts associated with cancer (26) cause production of LIF (27).

A variant in the LIF gene has been linked to schizophrenia (SCZ), osteoporosis and infertility in women (28-31),. To date, computational methods have not been used to identify pathogenic nonsynonymous single-nucleotide polymorphisms (nsSNPs) in the human LIF gene. (32, 33). To comprehensively examine the different types of variations, a thorough analysis of single nucleotide polymorphisms (SNPs) linked to a specific disease-related gene, along with extensive associative investigations, is essential in the present time. Unlike time-intensive molecular techniques, computational approaches now exist that elucidate the impacts of substituted amino acids, as well as alterations in protein structure and sequence information. In recent years, numerous in silico analyses have been developed to assess the physiological effects of deleterious nsSNPson genes (34-36).

Although there have been many studies on LIF gene, the pathogenic nsSNPs and genetic changes, as well as their impact on protein phenotypes, are not being fully investigated. This study aimed to determine, via molecular docking and molecular dynamic simulation of wild and variant proteins, how the polymorphism effect the function of LIF and causes symptoms in patients.

2. MATERIALS AND METHODS

The Human LIF gene sequence (Accession number: NC_000022.11) and LIF protein sequence (NP_002300.1) were obtained from NCBI. Relevant information from dbSNP and Protein ID (P15018) from UniProt were also retrieved. OMIM provided additional gene and protein data for LIF. Drug Bank databases were used for virtual screening of LIF-related compounds. These datasets were collected for subsequent computational analysis.

2.1. Identification of disease associated nsSNPs

Seven computational tools, including SNPNEXUS from https://www.snp-nexus.org/, were used to predict the deleterious effects of nsSNPs. The input file contained LIF SNPs IDs obtained from dbSNP. SNPNEXUS includes two tools, SIFT and PolyPhen, to sort out intolerant nsSNPs and categorize variants as probably damaging or benign. CADD provides detailed analysis and a C-score to assess the deleterious effect of variants on the protein. PolyPhen2 predicts harmful effects based on structural features. Condel integrates multiple algorithms to assess deleteriousness, while PROVEAN, SNAP2, and PMut predict variations as deleterious or neutral. SNP&GO and PHD-SNP tools were used to

predict disease-associated nsSNPs based on database analysis and reliability index. MetaSNP was used to filter out disease-related mutations.

2.2. Effect of nsSNPs on protein stability and conservation of amino acids

To predict the impact of SNPs on protein stability, three tools were utilized for reliable results. MuPro uses a support vector machine (SVM) method to predict changes in protein stability (Cheng, Randall, & Baldi, 2006). The output provides a score ranging from 0 to 1, where a score < 0 indicates protein destabilization. I-Mutant 2.0 is an SVM-based server that calculates the free energy change (DDG) to predict protein destabilization upon mutations (E. Capriotti, Fariselli, & Casadio, 2005). The reliability index, ranging from 0 to 10, is also provided, with a higher index indicating greater reliability. The I-Stable tool integrates eleven structure and sequence-based prediction methods to forecast protein destabilization caused by single amino acid mutations (Chen, Lin, Liao, Chang, & Chu, 2020).
Conservation analysis of amino acid residues in human LIF was performed using Consurf (Ashkenazy et al., 2010). Enzymatic sites in proteins tend to have conserved and deleterious amino acids compared to variable regions (Williamson et al., 2013). Consurf calculates a score of 7-9 for highly conservative amino acids using a Bayesian method, while lower scores indicate less conserved residues (Ashkenazy et al., 2010).

2.3. Identification of post translational modification (PTM) site & Molecular network interactions

Post-translational modifications (PTMs) play a crucial role in expanding proteomic diversity, regulating protein function, and contributing to disease processes. These modifications involve adding functional groups to specific amino acid residues within a protein.
In this study, the Musitedeep server was used to predict potential PTM sites in the Human LIF protein. Musitedeep utilizes a deep-learning framework and provides visualization tools for better understanding of the results. The amino acid sequence of LIF protein was used as input for the prediction analysis. For visualizing protein-protein interactions involving LIF, Cytoscape, a Java-based program, was employed. The input for Cytoscape can be the protein name, data in GML format, or simple interaction format (SIF).

2.4. 3D Modeling of LIF protein

Comparative homology modeling was done using Modeller 10.1 (37) which runs on Python scripts. The protein sequence was used for the selection of template using Blast, the model was built based on the alignment of protein sequence with the template structure and selection of final template. In case of mutant, the built model was individually mutated to their respective position through Pymol.

2.5. Refinement and structural validation of native and mutant LIF protein

The wild-type and mutant protein structures were refined using ModRefiner, which improves structure quality through a minimization process. The quality of the refined structures was assessed using ERRAT and RAMPAGE. RAMPAGE utilizes a Ramachandran plot analysis to evaluate structural stability and amino acid distribution. The TM-align algorithm was used to compare wild-type and mutant structures, providing a measure of structural similarity through TM-score and RMSD values. A higher RMSD value indicates greater variations between the native and mutant structures, while a TM-score of 1 signifies a perfect match between the superimposed structures.

2.6. Molecular docking and molecular dynamic simulation of wild and mutant LIF Protein

PyRx virtual screening tool (38) was used to docked ligands with protein in order to find ligand protein interactions along with their affinities. Based on binding affinities of Protein-ligand complex the selected complexes were then visualized by Discovery studio (39). Molecular dynamic simulation (MDS) checked the atoms and molecules movement of protein over a given time. Desmond was used for molecular dynamic simulation of wild and mutant protein structures.

3. RESULTS

3.1. Evaluation of SNPs of LIF gene

The human LIF gene contained a total 5103 single nucleotide polymorphisms (dbSNP-NCBI) (Figure S1). Out of which there were 114 synonymous SNPs, 233 were non-synonymous SNPs, 1456 in 5' upstream region, 1503 in 3' downstream region, 1291 intronic, 384 exonic, 785 lay in the 3'UTR region, 67 in the 5'UTR region (Figure 1, 2). For further analysis, we specifically focused on non-synonymous single nucleotide polymorphisms (nsSNPs). These nsSNPs involve changes in the codons, leading to the incorporation of different amino acids. Such alterations have the potential to exert structural and functional impacts on the protein. By narrowing our analysis to nsSNPs, we aimed to prioritize variations that could have significant implications for the protein's structure and function.

3.2. Identification of pathogenic and disease associated nsSNPs in LIF

Total of 233 nsSNPs were analyzed using SNPNEXUS (Figure 2). SIFT predicted 45 nsSNPs as deleterious with score ≤ 0.05,105 as tolerated with a score ≥ 0.05 and the rest were not reported (Figure S2). Further, based on structural information and Multiple Sequence Alignment (MSA), the PolyPhen indexing categorized 75 nsSNPs as probably damaging with score ranges 0.912-1.00 ,23 as possibly damaging score 0.452-0.868 and 107 as benign with score 0.423-0.001(Figure S3). To ensure robust and reliable outcomes, a stringent filtering process was employed, integrating data from SIFT and PolyPhen algorithms. Through this approach, a total of 28 non-synonymous single nucleotide polymorphisms (nsSNPs) were identified and retained for further analysis. By combining the predictive power of these computational tools, we aimed to enhance the confidence level of our results, focusing on nsSNPs that exhibited a higher likelihood of functional impact on the protein (Table S1). The corresponding nsSNPs were further validated using CADD, Polyphen-2 and SNAP2.In them, all nsSNPs were damaging and were found to have an "effect" on protein function. While Condel and Provean predicted 26 and 23 nsSNPs as deleterious respectively (Table 1)

Further, all 28 nsSNPs were submitted for associated disease (Table 2). The PHD-SNP and SNP&GO analysis calculated 22 and 27 nsSNPs as disease associated respectively. In addition, P-Mut predicted only one nsSNPs (G71E) was disease causing while Meta-SNP suggested 18 nsSNPs as disease related nsSNPs.

3.3. Effect of nsSNPs on protein stability and conservation of amino acids

Previous studies showed that disease associated nsSNPs change the protein stability (Aftab et al., 2021),(Hossain, Roy, & Islam, 2020; Jia et al., 2014). Protein stability of nsSNPs was determined by using Mu-Pro, I-mutant and I-stable online servers. The results showed that most of nsSNPs has decreased protein stability in all tested servers (Table 3). Mu-Pro analysis indicated that only one G71E had increased stability. While I-mutant and I-stable analysis revealed 25 and 21 SNPs decrease protein stability.

To identify mutations that may impact human health, it is necessary to consider evolutionary information. By utilizing Consurf, calculations were conducted to determine the evolutionary conservation of amino acid residues in the LIF protein. (Figure 3). Based on Consurf results C156F, C153G, L137F, V164M, C153Y, P90L, Q70H, N56S were highly conserved. Among these C156F, C153G, L137F, V164M, C153Y were structural and buried, P90L, Q70H, N56S were exposed and functional. G71E, N127K, N127D, P73L, L62V were conserved. And L147P, Y111C, Y66C, Y66H, L116V, T120N were slightly conserved while other nsSNPs were predicted as least conserved (Table 3).

Among the 28 identified non-synonymous single nucleotide polymorphisms (nsSNPs), a subset of nine nsSNPs (C156F, C153G, L147P, Y111C, Q70H, Y66C, Y66H, T120N, V164M) was determined to be highly significant in the context of the human LIF gene (Table S2). These findings were supported by the collective predictions of 15 different computational tools, including SIFT, PolyPhen, PROVEAN, PolyPhen2, SNAP2, CADD, Condel, PHD-SNP, SNP&GO, MuPro, I-Mutant, iStable, PMut, MetaSNP, and Consurf. The summarized results of these tools, illustrating the prediction of

common deleterious nsSNPs, are depicted in Figure 4.Among the 28 identified non-synonymous single nucleotide polymorphisms (nsSNPs), a subset of nine nsSNPs (C156F, C153G, L147P, Y111C, Q70H, Y66C, Y66H, T120N, V164M) was determined to be highly significant in the context of the human LIF gene (refer to Table S2). These findings were supported by the collective predictions of 15 different computational tools, including SIFT, PolyPhen, PROVEAN, PolyPhen2, SNAP2, CADD, Condel, PHD-SNP, SNP&GO, MuPro, I-Mutant, iStable, PMut, MetaSNP, and Consurf. The summarized results of these tools, illustrating the prediction of common deleterious nsSNPs, are depicted in Figure 4.

3.4. Prediction of post translational modification (PTM) sites and molecular network interactions

Post translational modification sites (PTM) plays prominent role in folding and degradation of proteins as well as gene expressions regulation. An extensive study between PTMs, SNPs and diseases are vital as SNPs induces PTMs and predicting the harmful nsSNPs associated with PTM sites may be helpful in the analyzing and interpretation of diseases (Kim, Kang, Min, & Yi, 2015). Musitedeep predicted four mutations comprising Post-translational modification sites (Table S3). Among them, N56S was found to regulate Glycosylation, A32T were associated with phosphorylation and C156F, and C153G were associated with palmitoylation (Figure 5A).

The interaction network of the LIF protein with other proteins was visualized using Cytoscape (Figure 5B). LIF protein was found to be associated with various interacting partners, including Interleukin-6 Cytokine (IL6), Vascular endothelial growth factor A (VEGFA), Signal transducer and activator of transcription 3 (STAT3), Interleukin-1 alpha (IL1A), Granulocyte colony stimulating factor (CSF3), Interleukin-1 beta (IL1B), Cardiotrophin-1 (CTF1), Interleukin-6 receptor subunit beta (IL6ST), Leukemia inhibitory factor receptor (LIFR), and Fibroblast growth factor 2 (FGF2). These interactions suggest potential functional connections and signaling pathways involving the LIF protein and its associated partners.

3.5. 3D Modeling of LIF protein

The Protein Data Bank (PDB) does not contain the complete structure of LIF protein. The complete protein structure was necessary to further analyze the effect of above shortlisted nine nsSNPs into protein structure. The protein templates were generated though protein data bank protein (pdb) selected as a search database in BLASTp along with the psiBlast as algorithm and protein sequence. The templates are selected based on E-value and % identity. The PDB ID 2Q7N, 1PVH have 100%, 1EMR found to have 98.11%, and 1A7M have 85% identical with the query sequence (Table S4). These four 2Q7N, 1PVH, 1EMR, 1A7M PDB ID structures were further used as a template for comparative Homology modeling by Modeller10.1. Modeller 10.1 generated five similar models based on 1A7M template (Table S5). The query sequence (NP_002300.1) was aligned against the selected template structures to build the model. The best model was selected based on the low DOPE (Discrete Optimized protein energy) score and high GA341.

Among the five model "qseq1.B99990002.pdb" had low -20176.61523 dope score and highest 1.00000 GA341 score (Table S5). The selected model"qseq1.B99990002.pdb" (Figure 5C) undergo point mutations in Pymol. The native model and point mutated structures were refined by ModRefiner.

3.6. Model validation by ERRAT and Ramachandran Plot

The refined native and mutant structures was further validated by ERRAT and Ramachandran (Table 4). The quality factor of native model was 86.2857%. In ProCheck, Ramachandran plot was used for further assessment of wild and mutant structure. In Ramachandran plot the generated native protein model was found to have 90.8% residues in the favored region, 6.9% in allowed region, 1.7% in generously disallowed regions and 0.6% disallowed region (Table 4; Figure 5D). For a good and reliable protein structures, there should be more than 90% residues in the favored region of native protein While among mutants C156F found to have 90.2% residues in favored region, 8.7% in allowed

region, 0.6% generously disallowed and 0.6% disallowed, Q70H have 91.9% in favored region, 6.4% in allowed,1.2% disallowed and 0.6 disallowed while remaining mutant structures showed less residues in favored region.

3.7. RMSD and TM Score calculations through TM-align

TM align were used to investigate the structural similarities between wild and mutant (Table 4). Among the 9 mutants, C156F found to have lowest. TM score 0.96626 and highest RMSD value 1.17Å, followed by T120N, V164M, Q70H having 0.97184, 0.96824, 0.97257 TM score and 1.11 Å, 1.10 Å, 1.02 Å RMSD value respectively. The remaining mutants were found to have less than 1 Å RMSD values. Based on Higher RMSD value and low TM score C156F have shown greater structural dissimilarity and selected for superimposition over the wild protein structures and further docking analysis (Figure 6).

3.8. Molecular docking analysis

Total 20 ligands associated with LIF protein were retrieved from PubChem Drug bank. PyRx were used to dock all 20 ligands with LIF protein along with the selected mutant C156F (Table S6). The grid box was set with axes X=6.3363, Y=-12.8628, Z=-1.0061. Eight ligands Lonaparisan(-8), EC330(-7.8), Coumestrol(-7.5), Estradiol(-7.3), desmethylmifepristone (-7.1), GH1(-7.1), Toripristone(-7.1), Mifepristone(-7) showed stronger binding affinities while in case of mutant C156F these eight ligands showed -7.5, -7.1, -6.4, -7.2, -6.8, -6.6, -7, -6.9 binding affinities and were selected for further analysis (Table 5). Less the value stronger the binding affinity of ligand and protein. The docked compound was visualized by Discovery studio (Figure 7A, B).

Using the discovery studio, the effect of mutations on hydrogen bond and other interactions were observed. When Lonaparisan interact with LIF protein, it showed C-H bond at GLY83, Alkyl and Pi-alkyl association at LYS175 and halogens interactions at GLY83 residues. Whereas when Lonaparisan interact with the C156F mutant showed H-bond at three different points (VAL110, THR114 and ARG145), C-H bond at HIS163, alkyl and Pi-alkyl association at ARG145 and halogens at GLY113. EC330 interact with LIF showed H-bond at GLN186, C-H bond at THR172, alkyl and Pi-alkyl interactions at LYS175 and LYS182 and halogens at ASP88. Whereas when interact with C156F mutant it showed H-bonding at ARG154 and CYS153, C-H bond at ARG154, alkyl and Pi-alkyl association at ARG154. Coumestrol interact with LIF protein, it only showed alkyl and Pi-alkyl interactions at LYS182 and LYS175. When interact with C156F mutant, it showed H-bonding at ASN150, CYS40 and GLU98, Pi-sigma at ARG154, Pi-Pi shaped at TYR159, alkyl and Pi-alkyl at ARG154 and Pi-sulphur at CYS40. Estradiol interacts with LIF it formed alkyl and Pi-alkyl association at LYS80 and LYS175. Whereas, when interact with mutant C156F, it showed H-bond at LEU44 and alkyl, pi-alkyl interactions at ARG154, PHE156, ALA35, PRO29 and ILE27. Desmethylmifepristone interact with LIF it showed C-H bond at LYS182, GLN186, alkyl and pi-alkyl interactions at LYS175. Whereas when interact with mutant C156F, it showed H-bond at SER149, Pi charged at ARG37, alkyl and pi-alkyl interactions at ARG145, VAL110, TYR168 and unfavorable donor-donor interactions at ARG37.GHI interact with LIF, it showed H-bond at different sites like GLY174, SER173 and LEU81, C-H bond at GLN186, LYS182, Pi-sigma at LYS175, alkyl and pi-alkyl interactions at LYS182 and unfavorable donor-donor interactions at SER173. Whereas when it interacts with mutant C156F it showed H- bonding at VAL166, C-H bond at THR114, Pi-charged at ARG145, Pi-sigma at VAL110 and unfavorable donor-donor interactions at ARG145. Toripristone interact with LIF protein, it showed C-H bond at ASP171 and when it interacts with C156F mutant it also showed C-H bond at ASP171. Mifepristone interacts with LIF protein it showed H-bond at LEU81 and alkyl and pi-alkyl interactions at LYS182. When it combines with C156F mutant, it showed H-bond at LEU81 and alkyl and pi-alkyl interactions at LYS175 and LYS182.

RMSD (roof mean square deviation and RMSF (root mean square fluctuations) were analyzed between native and mutant C156F complex in Desmond. Simulation was carried out at 100ns for each complex independently. RMSD value of protein and mutants were analyzed. In wild LIF protein, the

position of the backbone atoms of protein doesn't change much during simulation. RMSD is higher than standard value that means, after simulation time deviation was occurred. The protein is not stable and would result in abnormal function while mutant C156F has different spikes than the wild type as mutational protein has more hyper action than wild type. The RMSF value of wild and mutant is different that showed that mutation alter the protein flexibility (Figure 8).

4. DISCUSSION

Computational biology has firmly established its position in the field of genomic research. (40) It is also a common practice to employ computational biology techniques for the identification of deleterious mutations in target genes, which contribute to the underlying causes of diverse diseases. (41-43). The human SNP database (dbSNP) has documented over 4 million human SNPs, with approximately 2% of these SNPs found within coding regions. These coding region SNPs have been associated with various genetic diseases. Presently, conducting a comprehensive investigation regarding the effects of non-synonymous single nucleotide polymorphisms (nsSNPs) poses a significant challenge. However, computational tools offer a valuable means to gather information concerning the impact of nsSNPs on protein structure and function. (44, 45).

The LIF gene has accumulated over 5000 reported SNPs in the dbSNP-NCBI database. These SNPs can be found in coding, non-coding, or regulatory regions. Another valuable resource for human gene mutation data is the Human Gene Mutation Database (HGMD) (46). R. Giess et al. (1999) reported three mutations in the LIF gene. Among these mutations, two are missense mutations, the third mutation is located at the start codon of exon 1, which is part of the regulatory region of the gene (29), The other two mutations were found in the third exon of the LIF gene. These regions are crucial for the interaction between LIF and its receptor. Polymorphisms in these regions can lead to a decrease in the biological activity of the LIF protein, which in turn may contribute to female infertility (29), In a separate investigation, a gene mutation involving the substitution of valine with methionine at codon 64 (V64M) was examined (47), in addition to the V64M mutation, two polymorphisms were identified in the upstream region of the LIF gene and are associated with infertility, (32). In the Japanese population, the LIF gene was found to harbor four single-nucleotide polymorphisms (SNPs) located within the third exons at positions 3951 C/T, 4442 A/C, 4376 C/G, and 5961 G/A(48). However, it is worth noting that these specific SNPs, namely rs929271 in the 3' untranslated region (UTR) of LIF, as well as rs929273 and rs737812, have been linked to schizophrenia (SCZ), neural memory degradation, and pregnancy loss (30).

Currently, there is a lack of reported studies specifically focusing on the utilization of computational tools for predicting deleterious single nucleotide polymorphisms (SNPs). In our study, we employed a range of computational tools such as SNPnexus, CADD, Condel, Polyphen2, Provean, SNP&GO, PHD-SNP, P-Mut, I-Mutant 2.0, Istable, and MuPro to filter and analyze nsSNPs. These tools were utilized to assess the potential impact of these nsSNPs on protein structure and function. These tools were employed to identify and prioritize nsSNPs that are likely to have the greatest impact and association with diseases. By employing these computational approaches, we aimed to enhance the efficiency and accuracy of identifying deleterious nsSNPs with potential disease relevance. It is important to distinguish between deleterious and neutral nsSNPs, as the former has the potential to impact enzymatic activity (49). The computational tools employed in the study predicted nine highly deleterious nsSNPs (C156F, C153G, L147P, Y111C, Q70H, Y66C, Y66H, T120N, and V164M). These nsSNPs were found to be located within the binding regions of proteins, which are critical for protein-protein interactions. These binding regions are typically conserved, comprising charged residues and hydrophobic residues that form a hydrophobic core involved in the binding process. The presence of harmful genetic variability, as indicated by these deleterious nsSNPs, can have adverse effects on the protein structure and function. Such effects include protein destabilization, alteration in protein conformation and dynamics, and modifications in the selectivity and affinity of binding partners. These detrimental consequences may disrupt proper protein-protein interactions, leading to impaired cellular processes and potential disease manifestations. (50). Understanding the correlation between SNPs and

their phenotypic consequences is crucial for elucidating the etiology of various diseases or disorders. By comprehending the impact of SNPs on gene function, protein structure, and biological processes, we can gain valuable insights into the underlying mechanisms of these conditions (51, 52). Such knowledge is essential for identifying disease-associated SNPs, predicting disease risk, and developing personalized therapeutic approaches. (53). Protein stability is a fundamental factor that influences the physiology of biological molecules. Deleterious (nsSNPs) may destabilize and misfold the protein, contributing to significant functional consequences.(54-57). Among the nine non-synonymous single nucleotide polymorphisms (nsSNPs) examined, C156F, C153G, Q70H, and V164M were identified as highly conserved, with a conservation score of 9. Notably, C156F and C153G were found to impact the post-translational modification of the LIF protein.

Further analysis focused on the structural consequences of these deleterious nsSNPs in the three-dimensional (3D) structure of LIF. The Modeller10.1 software was utilized to obtain the 3D structure, followed by refinement and structure validation using Modrefiner, ERRAT, and ProCheck. To assess the structural deviations of the wild-type and mutant proteins, the Tm Align tool was employed to predict the root mean square deviation (RMSD) and TM score. Higher RMSD and lower TM score values indicated a greater deviation of the mutant protein structure from the native counterpart. Based on these evaluation criteria, the C156F nsSNP was selected as the most deleterious, as the remaining eight nsSNPs exhibited lower RMSD and higher TM scores, indicating a lesser deviation of the mutant protein structures from the wild-type structure. Both the wild-type and mutant C156F proteins were found to interact with eight significant ligands. LIF, being a pleiotropic cytokine, has posed challenges in delineating its precise functions, given its multifaceted nature.

In our study, we conducted molecular dynamics simulations (MDS) to investigate the structural impact of nsSNPs on the protein and their potential to alter its biological function. To the best of our knowledge, no previous research has utilized molecular dynamics simulations to explore the effect of nsSNPs on the structural integrity of this gene.

By employing MDS analysis, we aimed to generate comprehensive data on the structural changes induced by pathogenic mutations, including residue variations and conformational alterations in the protein. This approach allows for a deeper understanding of the structural consequences associated with deleterious mutations, complementing experimental techniques.

Interestingly, emerging evidence and substantial scientific investigations have increasingly recognized LIF as a compelling candidate for cancer therapy, particularly in cases where LIF is overexpressed. In line with this, we have investigated a group of small molecule compounds, specifically EC330, as potential LIF inhibitors. The design of EC330 was guided by structure-activity relationship (SAR) studies conducted on human breast cancer MCF7 cells with LIF overexpression. Our research endeavors to provide significant contributions by elucidating the structural consequences of non-synonymous single nucleotide polymorphisms (nsSNPs) on the protein. Furthermore, we aim to provide insight on the potential therapeutic strategies targeting Leukemia inhibitory factor (LIF), especially in cancer cases characterized by LIF overexpression. By understanding the impact of nsSNPs on the protein's structure and function, we can identify potential targets for therapeutic intervention and contribute to the development of effective treatment approaches.(58). To assess the influence of the mutant and wild-type proteins, we generated graphical representations of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF). These parameters offer valuable information regarding the structural alterations and dynamic characteristics of the proteins.

In our computational analysis, we predicted deleterious mutations; however, among all nine it is important to note that V64M was not identified as highly deleterious in our study. As far as our understanding, there is no existing evidence linking the mutations C156F, C153G, L147P, Y111C, Q70H, Y66C, Y66H, T120N, and V164M to any recognized diseases.

It is worth mentioning that one of the limitations of computational tools is that the results obtained from these analyses require further validation through wet lab experiments. While bioinformatics approaches are time-saving and cost-effective, the predictions regarding the deleterious effects of SNPs

on protein physiology necessitate wet lab studies for confirmation. This integration of computational and experimental approaches represents a crucial step forward in drug design and development.

Supplementary Data

Figure S1: SNPNEXUS showed that LIF gene contains 223 non-synonymous SNPs, 114 synonymous SNPs, 785 in the 3'UTR regions, and 67 in the5'UTR. 1456 in the 5'Upstream, 1503 in the 3'Downstream; Figure S2: The SIFT analysis, indicated 44 non synonymous SNPs (nsSNPs) as deleterious and 106 were found to be tolerated; Figure S3: PolyPhen classified 107 nsSNPs to be 'Benign, probably damaging (75) and possibly damaging (23) respectively; Table S1: nsSNPs filtered out by combining the information from SIFT and PolyPhen; Table S2: Significant SNPs in human LIF gene; Table S3: Identification of Post-Translational Modification (PTMs) sites through MusiteDeep; Table S4: Blastp results; Table S5:Five best models predicted by Modeller 10.1; Table S6: Docking of 20 ligands with native and mutant protein along with their binding affinities.

Authors' contributions

M. Afzal perceived the study design and collected data, S. Ahmad, U. Ali and A. Bilal analyzed, interpreted the results and drafted the manuscript. F. Tanvir, A. Riaz and S. Ahmad performed a critical revision of the manuscript and help in writing. All authors approved the version to be published and agreed to be accountable for all aspects of the work.

Acknowledgments

None

Funding

None

Data availability

The data used to support the findings of this research are available from the corresponding author upon request.

Conflict of interest

The authors declare that we have no conflict of interest.

REFERENCES

1. Metcalfe SJG, Immunity. LIF in the regulation of T-cell fate and as a potential therapeutic. 2011;12(3):157-68.
2. SIMPSON RJ, HILTON DJ, NICE EC, RUBIRA MR, METCALF D, GEARING DP, et al. Structural characterization of a murine myeloid leukaemia inhibitory factor. 1988;175(3):541-7.
3. Robinson R, Grey L, Staunton D, Vankelecom H, Vernallis A, Moreau J-F, et al. The crystal structure and biological function of leukemia inhibitory factor: implications for receptor binding. 1994;77(7):1101-16.
4. Baumann H, Wong GJTJoI. Hepatocyte-stimulating factor III shares structural and functional identity with leukemia-inhibitory factor. 1989;143(4):1163-7.
5. Boulanger MJ, Bankovich AJ, Kortemme T, Baker D, Garcia KCJMc. Convergent mechanisms for recognition of divergent cytokines by the shared signaling receptor gp130. 2003;12(3):577-89.
6. Nicola NA, Babon JJJC, reviews gf. Leukemia inhibitory factor (LIF). 2015;26(5):533-44.
7. Lin T-A, Wu T-S, Li Y-J, Yang C-N, Illescas Ralda MM, Chang H-HJJocm. Role and mechanism of LIF in oral squamous cell carcinoma progression. 2020;9(2):295.
8. Gulluoglu S, Sahin M, Tuysuz EC, Yaltirik CK, Kuskucu A, Ozkan F, et al. Leukemia inhibitory factor promotes aggressiveness of chordoma. 2017;25(7):1177.

9. Wang D, Liu K, Yang Y, Wang T, Rao Q, Guo W, et al. Prognostic value of leukemia inhibitory factor and its receptor in pancreatic adenocarcinoma. 2020;16(3):4461-73.
10. Liu S-C, Chang Y-SJM, Oncology C. Role of leukemia inhibitory factor in nasopharyngeal carcinogenesis. 2014;1(1):e29900.
11. Atlas HPJE-Ic. Available online: https://www. proteinatlas. org. 2021.
12. Qian L, Xu F, Wang X, Jiang M, Wang J, Song W, et al. LncRNA expression profile of ΔNp63α in cervical squamous cancers and its suppressive effects on LIF expression. 2017;96:114-22.
13. Li X, Yang Q, Yu H, Wu L, Zhao Y, Zhang C, et al. LIF promotes tumorigenesis and metastasis of breast cancer through the AKT-mTOR pathway. 2014;5(3):788.
14. Kuphal S, Wallner S, Bosserhoff AKJE, pathology m. Impact of LIF (leukemia inhibitory factor) expression in malignant melanoma. 2013;95(2):156-65.
15. Mashayekhi F, Salehi ZJJoCN. Expression of leukemia inhibitory factor in the cerebrospinal fluid of patients with multiple sclerosis. 2011;18(7):951-4.
16. Ancey C, Corbi P, Froger J, Delwail A, Wijdenes J, Gascan H, et al. Secretion of IL-6, IL-11 and LIF by human cardiomyocytes in primary culture. 2002;18(4):199-205.
17. Fukada K, Korsching S, Towle MFJGF. Tissue-specific and ontogenetic regulation of LIF protein levels determined by quantitative enzyme immunoassay. 1997;14(4):279-95.
18. Chesnokova V, Melmed SJE. Leukemia inhibitory factor mediates the hypothalamic pituitary adrenal axis response to inflammation. 2000;141(11):4032-40.
19. Morel DS, Taupin J-L, Potier M, Deminière C, Potaux L, Gualde N, et al. Renal synthesis of leukaemia inhibitory factor (LIF), under normal and inflammatory conditions. 2000;12(3):265-71.
20. Gadient RA, Patterson PHJSc. Leukemia inhibitory factor, Interleukin 6, and other cytokines using the GP130 transducing receptor: roles in inflammation and injury. 1999;17(3):127-37.
21. Kellokumpu-Lehtinen P, Talpaz M, Harris D, Van Q, Kurzrock R, Estrov ZJIjoc. Leukemia-inhibitory factor stimulates breast, kidney and prostate cancer cell proliferation by paracrine and autocrine pathways. 1996;66(4):515-9.
22. Paiva P, Menkhorst E, Salamonsen L, Dimitriadis EJC, reviews gf. Leukemia inhibitory factor and interleukin-11: critical regulators in the establishment of pregnancy. 2009;20(4):319-28.
23. Burdon T, Smith A, Savatier PJTicb. Signalling, cell cycle and pluripotency in embryonic stem cells. 2002;12(9):432-8.
24. Mattei S, Colombo MP, Melani C, Silvani A, Parmiani G, Herlyn MJIjoc. Expression of cytokine/growth factors and their receptors in human melanoma and melanocytes. 1994;56(6):853-7.
25. Kamohara H, Sakamoto K, Ishiko T, Mita S, Masuda Y, Abe T, et al. Human carcinoma cell lines produce biologically active leukemia inhibitory factor (LIF). 1994;85(2):131-40.
26. Ohata Y, Tsuchiya M, Hirai H, Yamaguchi S, Akashi T, Sakamoto K, et al. Leukemia inhibitory factor produced by fibroblasts within tumor stroma participates in invasion of oral squamous cell carcinoma. 2018;13(2):e0191865.
27. Bilal A, Naveed N. A brief note on cancer and its treatment. Occup Med Health Aff. 2021;9(7):1-3.
28. Králícková M, Síma R, Martínek P, Vanecek T, Ulcová-Gallová Z, Síma P, et al. [The leukemia inhibitory factor gene mutations in the population of infertile women: the heterozygote transition G to A on the position 3400 does not affect the outcome of the infertility treatment]. Ceska gynekologie. 2007;72(4):293-8.
29. Giess R, Tanasescu I, Steck T, Sendtner MJMhr. Leukaemia inhibitory factor gene mutations in infertile women. 1999;5(6):581-6.
30. Okahisa Y, Ujike H, Kunugi H, Ishihara T, Kodama M, Takaki M, et al. Leukemia inhibitory factor gene is associated with schizophrenia and working memory function. 2010;34(1):172-6.
31. Niaei G, Taghavi BA, Niaei AJIJoPH. V64M Mutation in Leukemia Inhibitory Factor Gene in Women Infertility. 2017;46(7):1003-4.

32. Tajeddin N, Ahadi AM, Javadi G, Ayat H. Investigation of Polymorphisms in the Upstream Sequence of LIF and LIFR Genes in the Infertile Women. 2021;23(1):e95239.
33. Moudi M, Sargazi S, Heidari Nia M, Saravani R, Shirvaliloo M, Shakiba M. Polymorphism in the 3′-UTR of LIF but Not in the ATF6B Gene Associates with Schizophrenia Susceptibility: a Case-Control Study and In Silico Analyses. Journal of Molecular Neuroscience. 2020;70(12):2093-101.
34. George Priya Doss C, Nagasundaram N, Chakraborty C, Chen L, Zhu H. Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: a molecular dynamics approach. Human Genomics. 2013;7(1):10.
35. Desai M, Chauhan JB. Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. Computational biology and chemistry. 2017;70:7-14.
36. Bilal A, Iqbal A, Rauf A, Azam AR. Top outbreaks of 21st century: a review.
37. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Current protocols in bioinformatics. 2016;54:5.6.1-5.6.37.
38. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods in molecular biology (Clifton, NJ). 2015;1263:243-50.
39. Jejurikar BL, Rohane SHJAJoRiC. Drug Designing in Discovery Studio. 2021;14(2):135-8.
40. Ahmed F, Kumar M, Raghava GP. Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. In silico biology. 2009;9(3):135-48.
41. Kamaraj B, Rajendran V, Sethumadhavan R, Kumar CV, Purohit R. Mutational analysis of FUS gene and its structural and functional role in amyotrophic lateral sclerosis 6. Journal of biomolecular structure & dynamics. 2015;33(4):834-44.
42. Arshad M, Bhatti A, John P. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. PloS one. 2018;13(1):e0188143.
43. Singh D, Rahi A, Kumari R, Gupta V, Gautam G, Aggarwal S, et al. Computational and mutational analysis of TatD DNase of Bacillus anthracis. Journal of cellular biochemistry. 2019.
44. Karchin RJBib. Next generation tools for the annotation of human SNPs. 2009;10(1):35-52.
45. Doss CGP, Sudandiradoss C, Rajasekaran R, Choudhury P, Sinha P, Hota P, et al. Applications of computational algorithm tools to identify functional SNPs. 2008;8(4):309-16.
46. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human gene mutation database (HGMD®): 2003 update. 2003;21(6):577-81.
47. Niaei G, Amir Taghavi B, Niaei A. V64M Mutation in Leukemia Inhibitory Factor Gene in Women Infertility. Iran J Public Health. 2017;46(7):1003-4.
48. Ishida R, Ezura Y, Iwasaki H, Nakazawa I, Kajita M, Kodaira M, et al. Linkage disequilibrium and haplotype analysis among four novel single-nucleotide polymorphisms in the human leukemia inhibitory factor (LIF) gene. 2001;46(10):557-9.
49. Johnson MM, Houck J, Chen CJCE, Biomarkers P. Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. 2005;14(5):1326-9.
50. Ma B, Elkayam T, Wolfson H, Nussinov RJPotNAoS. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. 2003;100(10):5772-7.
51. Bilal A, Ahmad S, Tanvir F, Tariq M, Ramzan K, Saleem M, et al. PREDICTIVE MODELING OF N-ACETYL TRANSFERASE 2 SINGLE NUCLEOTIDE POLYMORPHISMS AND BREAST CANCER RISK USING IN-SILCO APPROACHES. THE JOURNAL OF MICROBIOLOGY AND MOLECULAR GENETICS. 2022;3(2):105-21.
52. Bilal A, Ahmad S, Nisa FU, Ali F, Ramzan K, Tariq M, et al. ROLE OF TUMOR NECROSIS FACTOR-? ETA (TNF-?) IN GASTRIC CANCER: SINGLE NUCLEOTIDE POLYMORPHISMS ANALYSIS-AN IN-SILICO STUDY. THE JOURNAL OF MICROBIOLOGY AND MOLECULAR GENETICS. 2022;3(3):191-206.

53. Thusberg J, Olatubosun A, Vihinen MJHm. Performance of mutation pathogenicity prediction methods on missense variants. 2011;32(4):358-68.
54. Ferrer-Costa C, Orozco M, de la Cruz XJJomb. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. 2002;315(4):771-86.
55. Bross P, Corydon TJ, Andresen BS, Jørgensen MM, Bolund L, Gregersen NJHm. Protein misfolding and degradation in genetic diseases. 1999;14(3):186-98.
56. Shaw GJBi. Polymorphism and single nucleotide polymorphisms (SNP s). 2013;112(5):664-5.
57. Khan IA, Mort M, Buckland PR, O'Donovan MC, Cooper DN, Chuzhanova NAJIsb. In silico discrimination of single nucleotide polymorphisms and pathological mutations in human gene promoter regions by means of local DNA sequence context and regularity. 2006;6(1, 2):23-34.
58. Nair H, Santhamma B, Nickisch K. Cytotoxic agents that preferentially target leukemia inhibitory factor (LIF) for the treatment of malignancies and as new contraceptive agents. Google Patents; 2018.

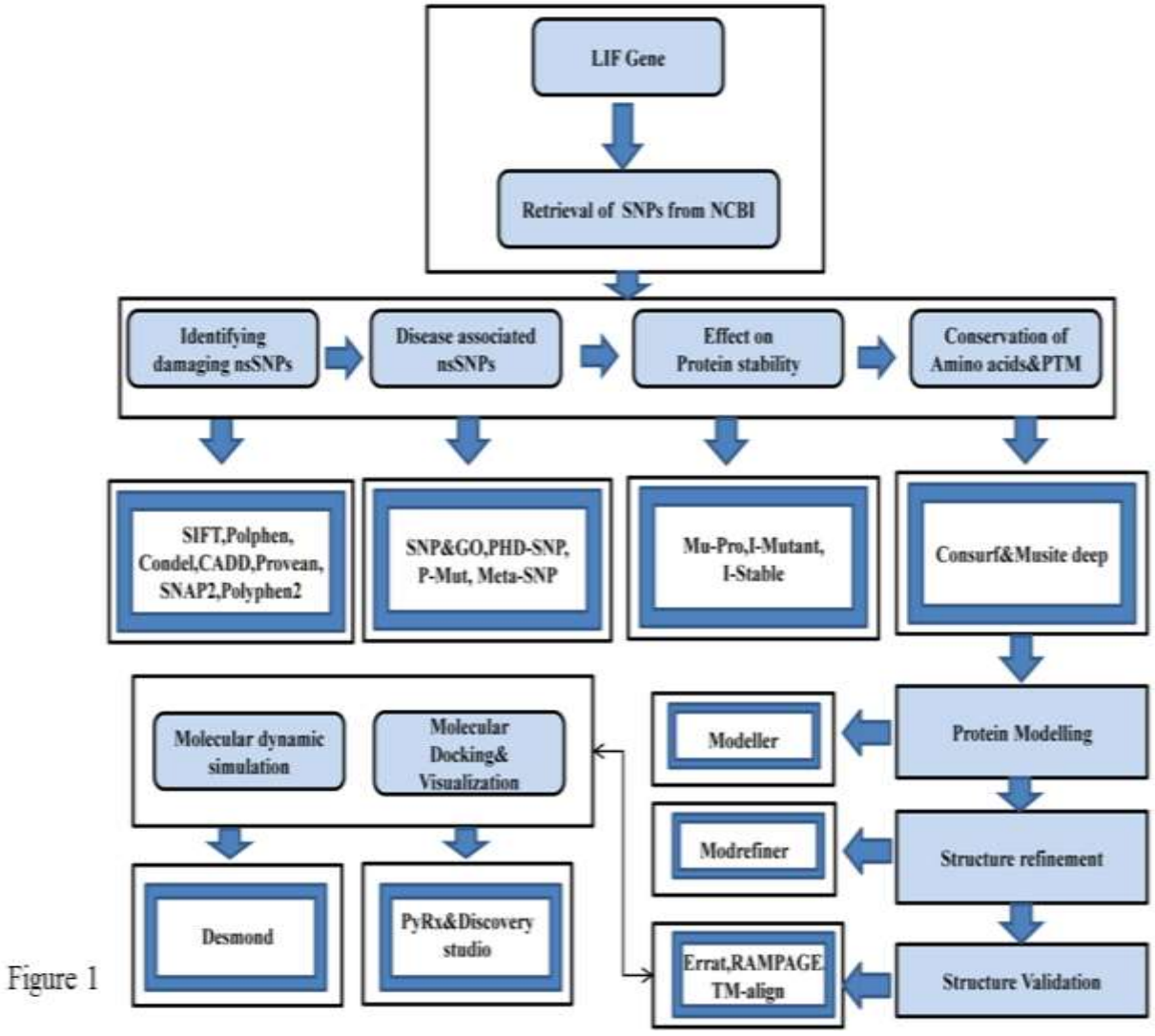

Figure 1. Flow chart of LIF protein analysis.

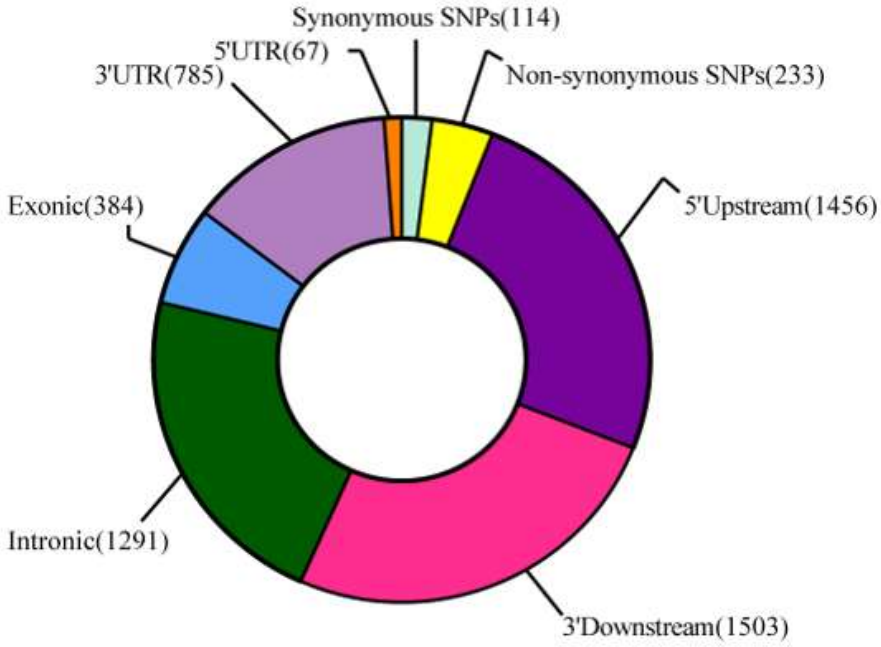

Figure 2. Evaluation of SNPs.

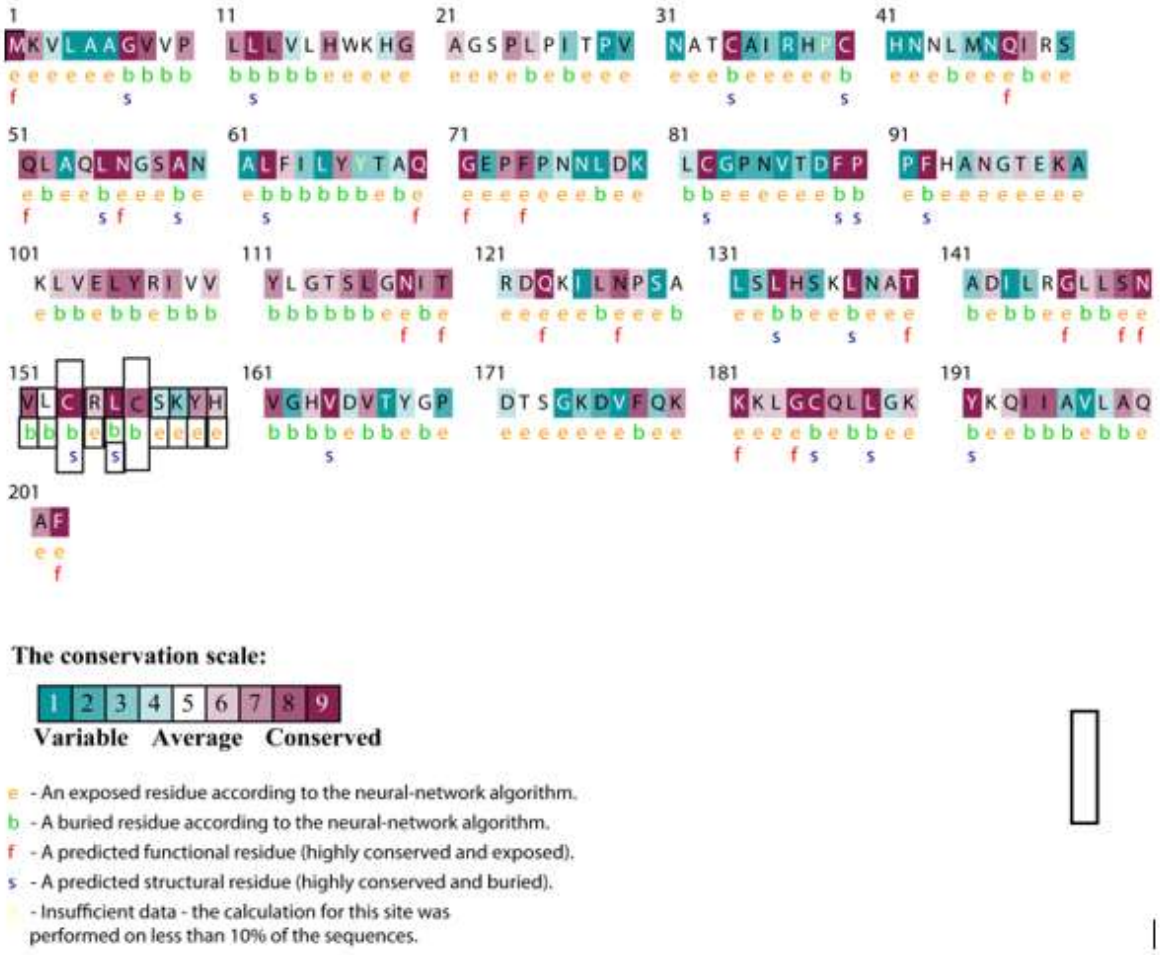

Figure 3. Prediction of evolutionary conservation of amino acids.

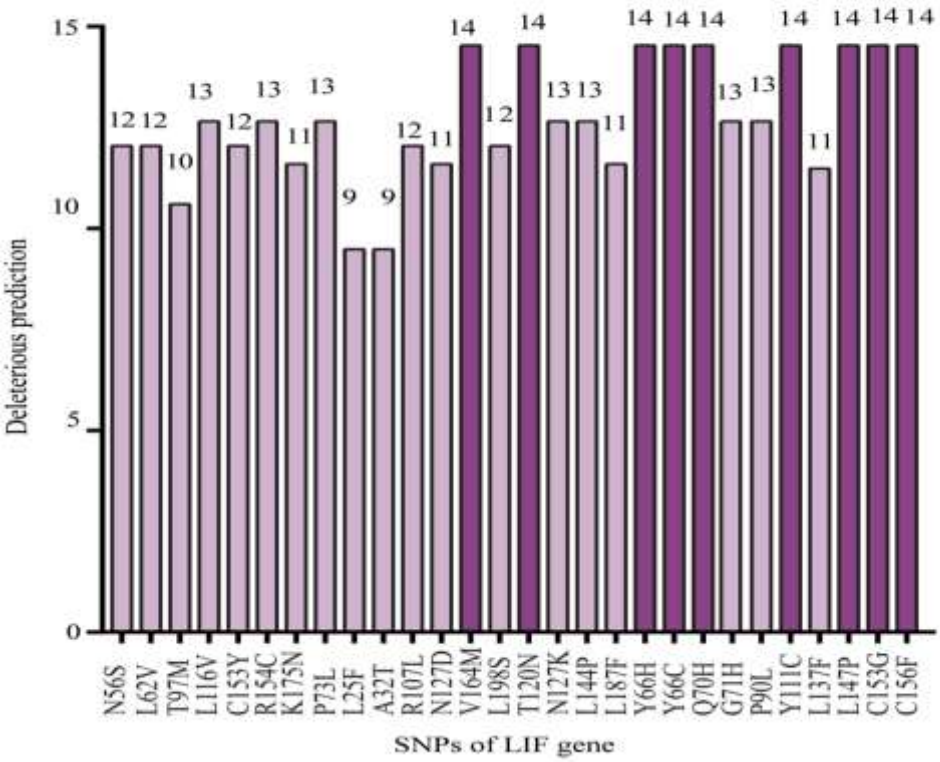

Figure 4. Significant nsSNPs verified by 15 tools. The dark purple lines showing significant SNPs.

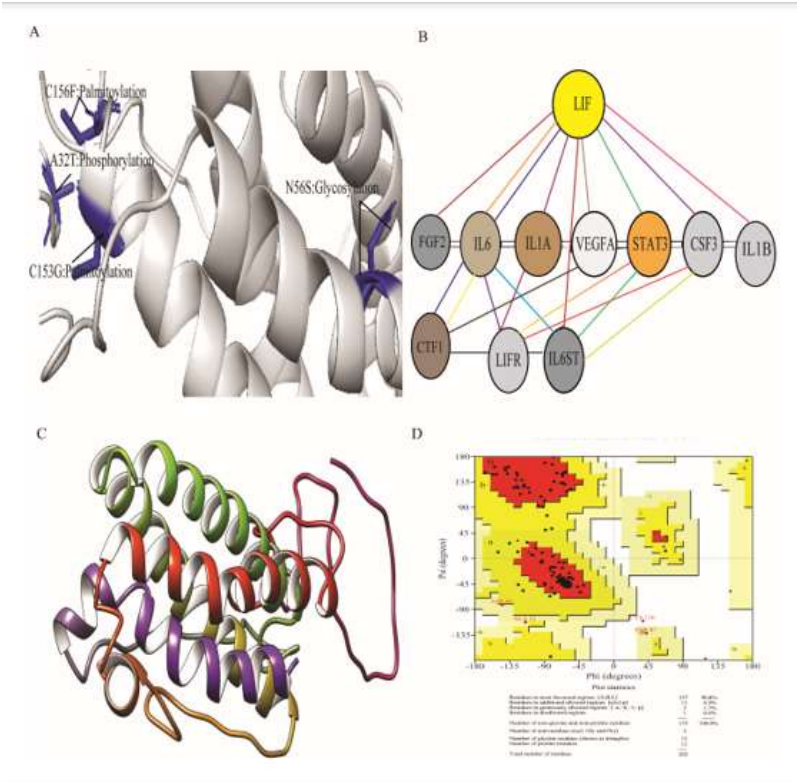

Figure 5. Identification of Post translational modification sites through Musitedeep (A), Protein-Protein interactions by Cytoscape (B), 3-D structure of LIF protein (C) and Ramachandran plot of LIF protein (D).

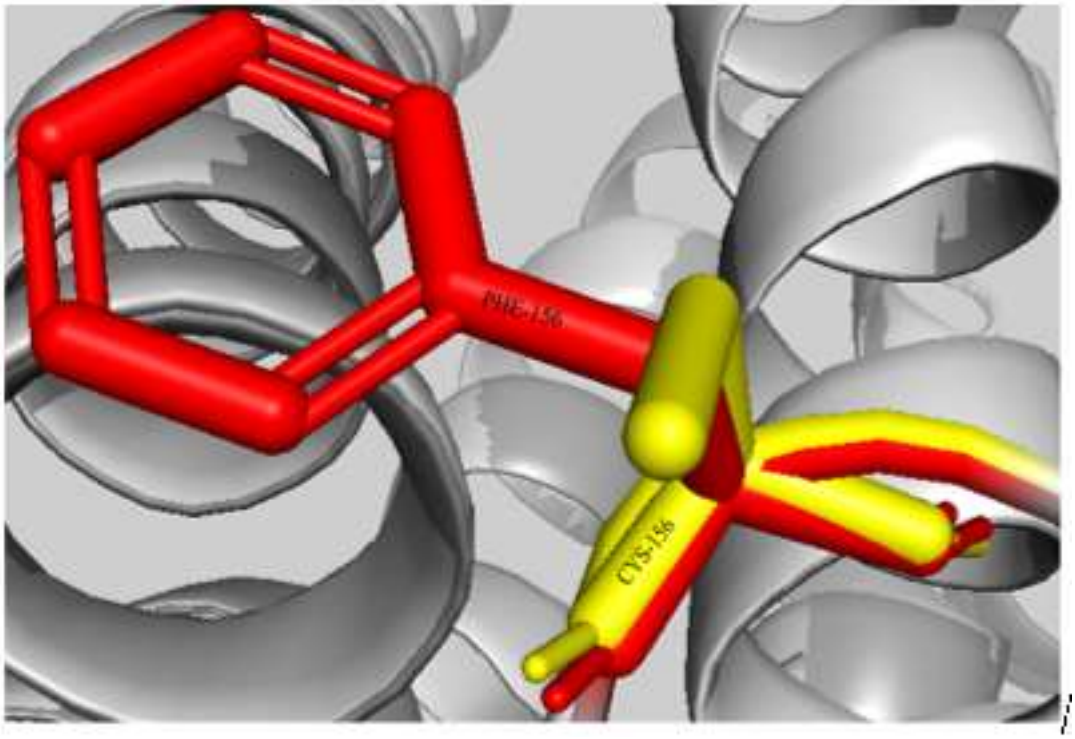

Figure 6. Superimposition structure of wild LIF protein over C156F mutant.

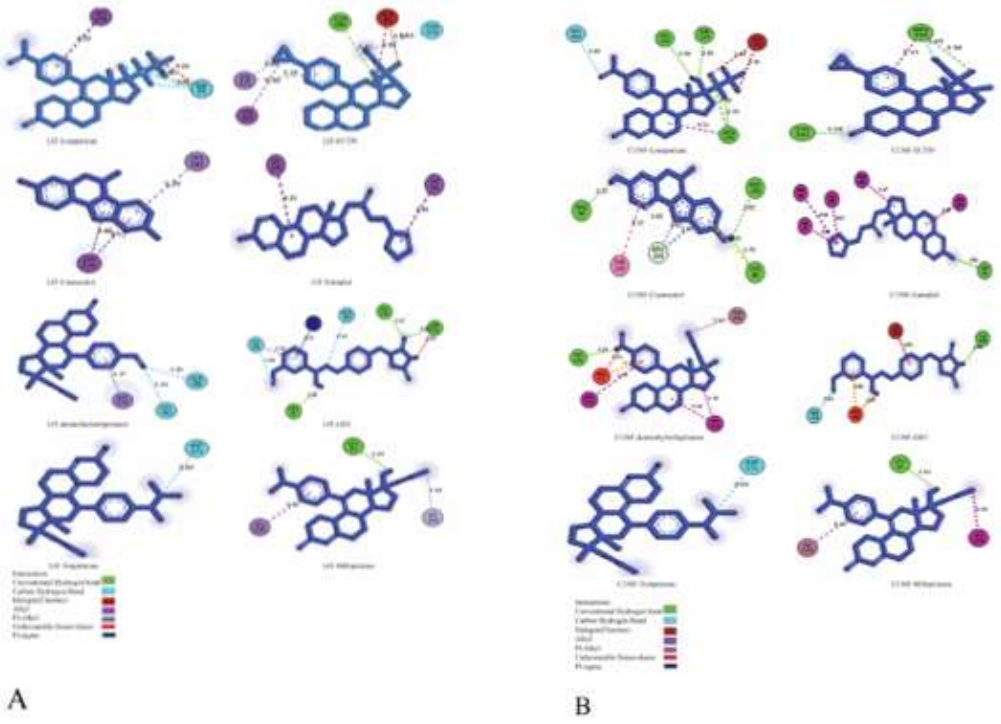

Figure 7. A) Showing the protein-ligand interactions for native structures. Residues were distributed based on their ten types of interactions including conventional hydrogen bonds, carbon hydrogen bonds, pi-charged, pi-sigma, pi-pi/pi-pi T shaped and alkyl, pi-alkyl and unfavorable donor, B) Showed the protein-ligand interactions for mutant structures. Residues were distributed based on their types of interactions including conventional hydrogen bonds, carbon hydrogen bonds, pi-charged, pi-sigma, pi-pi/pi-pi T shaped and alkyl, pi-alkyl and unfavorable donor.

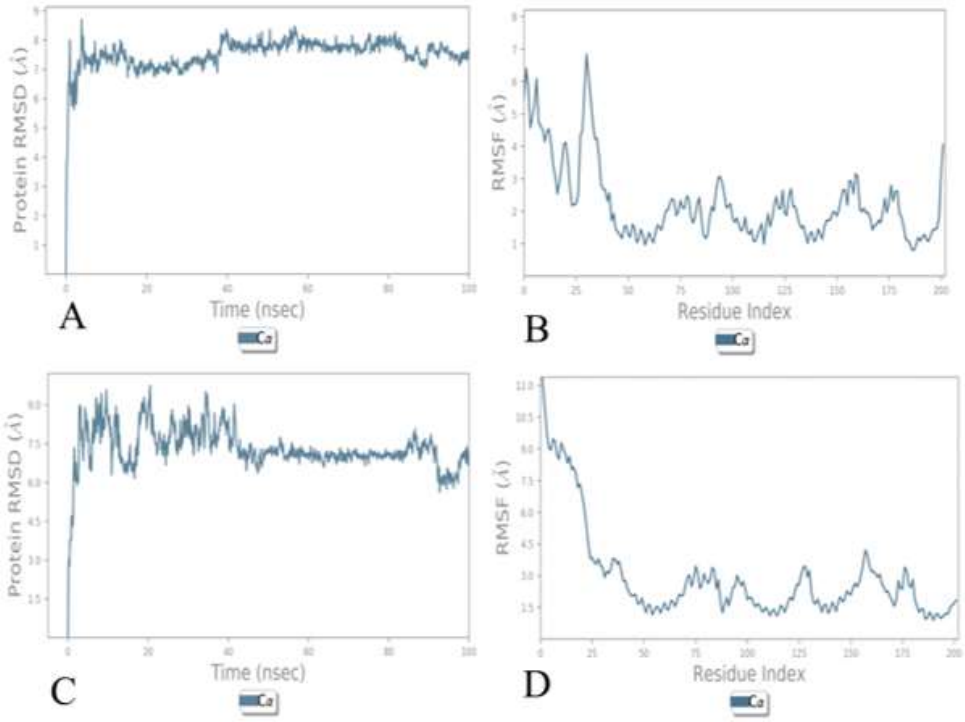

Figure 8. A and B showing RMSD and RMSF of wild protein, whereas, C and D depict RMSD and RMSF of mutant C156F protein.

Table 1. Identification of nsSNPs pathogenicity in LIF.

		SIFT		PolyPhen		CADD	Polyphen2		Condel		PROVEAN		SNAP2	
dbSNP	**Mutations**	**prediction**	**Score**	**Prediction**	**Score**	**Score**	**Prediction**	**Score**	**Label**	**Score**	**Prediction**	**Score**	**Prediction**	**score**
rs775324532	C156F	Deleterious	0	Probably Damaging	1	27	Probably damaging	0.998	Deleterious	0.6017010 2	Deleterious	-8.6	effect	93
rs760902711	C153G	Deleterious	0	Probably Damaging	1	28	Probably damaging	1	Deleterious	0.5992744 2	Deleterious	-11	effect	75
rs127750482 3	L147P	Deleterious	0	Probably Damaging	1	28	Probably damaging	1	Deleterious	0.5805059 7	Deleterious	-6.3	effect	86
rs1006881400	L137F	Deleterious	0	Probably Damaging	1	26	Probably damaging	0.999	Deleterious	0.5875659 5	Deleterious	-3.8	effect	71
rs142459670 9	Y111C	Deleterious	0	Probably Damaging	1	25	Probably damaging	1	Deleterious	0.5952046 9	Deleterious	-6.3	effect	69
rs779219330	P90L	Deleterious	0	Probably Damaging	1	28	Probably damaging	1	Deleterious	0.5884722 1	Deleterious	-8.2	effect	74
rs122417622 6	G71E	Deleterious	0	Probably Damaging	1	27	Probably damaging	1	Deleterious	0.5895663 2	Deleterious	-6.5	effect	77

rs125457070 2	Q70H	Deleterious	0	Probably Damaging	0.99	23	Probably damaging	0.993	Deleterious	0.5854709 7	Deleterious	-4.1	effect	61
rs138834734 4	Y66C	Deleterious	0	Probably Damaging	1	33	Probably damaging	1	Deleterious	0.5819761 8	Deleterious	-6.9	effect	64
rs101483707 0	Y66H	Deleterious	0	Probably Damaging	1	32	Probably damaging	1	Deleterious	0.5816237 6	Deleterious	-4	effect	74
rs368411105	L187F	Deleterious	0.01	Probably Damaging	1	32	Probably damaging	0.999	Deleterious	0.5878505 5	Deleterious	-3.1	effect	55
rs758693208	L144P	Deleterious	0.01	Probably Damaging	0.99	25	Probably damaging	1	Deleterious	0.5411878 5	Deleterious	-4.1	effect	82
rs762890518	N127K	Deleterious	0.01	Probably Damaging	1	23	Probably damaging	0.999	Deleterious	0.5897151 7	Deleterious	-4.8	effect	15
rs143701849 6	T120N	Deleterious	0.01	Probably Damaging	1	26	Probably damaging	1	Deleterious	0.5874778 7	Deleterious	-3.3	effect	34
rs760089055	L198S	Deleterious	0.02	Probably Damaging	0.98	36	Probably damaging	0.997	Deleterious	0.5366640 8	Deleterious	-2.7	effect	30
rs132781312 6	V164M	Deleterious	0.02	Probably Damaging	1	25	Probably damaging	0.996	Deleterious	0.5982825 4	Deleterious	-2.5	effect	76
rs127381048 0	N127D	Deleterious	0.02	Probably Damaging	1	27	Probably damaging	0.999	Deleterious	0.5908662 7	Deleterious	-3.9	effect	30
rs148200166	R107L	Deleterious	0.02	Probably Damaging	0.97	25	Probably damaging	1	Deleterious	0.5900703 4	Neutral	-2.3	effect	71
rs868113406	A32T	Deleterious	0.02	Probably Damaging	0.98	24	Probably damaging	0.995	Neutral	0.4953598 2	Neutral	-1.4	effect	64
rs144294088 8	L25F	Deleterious	0.02	Probably Damaging	0.97	25	Probably damaging	0.99	Neutral	0.5119193 1	Neutral	-2.5	effect	64
rs140799590	P73L	Deleterious	0.03	Probably Damaging	1	29	Probably damaging	1	Deleterious	0.5908600 5	Deleterious	-8.2	effect	50
rs131364203 6	K175N	Deleterious	0.04	Probably Damaging	1	25	Probably damaging	1	Deleterious	0.5946574 4	Deleterious	-4.5	effect	63
rs533306784	R154C	Deleterious	0.04	Probably Damaging	1	32	Probably damaging	1	Deleterious	0.5742238 4	Deleterious	-4	effect	38
rs750628718	C153Y	Deleterious	0.04	Probably Damaging	1	26	Probably damaging	0.998	Deleterious	0.5990680 5	Deleterious	-9.9	effect	85
rs373784036	L116V	Deleterious	0.05	Probably Damaging	0.97	24	Probably damaging	0.962	Deleterious	0.5833187 8	Neutral	-2.1	effect	55
rs371771485	T97M	Deleterious	0.05	Probably Damaging	1	24	Probably damaging	1	Deleterious	0.5770430 4	Deleterious	-3.4	effect	42
rs758458009	L62V	Deleterious	0.05	Probably Damaging	0.97	27	Probably damaging	0.994	Deleterious	0.5789807 2	Neutral	-2.4	effect	7

rs7808 02848	N56 S	Dele- teri- ous	0. 05	Probably Damag- ing	0. 93	27	Probably damag- ing	0. 99 9	Dele- teri- ous	0.595 3402 2	Dele- teri- ous	- 4. 1	ef- fect	39

Table 2. Identification of disease associated nsSNPs in LIF.

		SNP&GO	PHD-SNP		PMut		MetaSNP	
dbSNP	Muta- tions	Predic- tion	Predic- tion	RI	Predic- tion	Score	Predic- tion	Score
rs775324532	C156F	Disease	Disease	7	FALSE	0.4646	Disease	0.748
rs760902711	C153G	Disease	Disease	4	FALSE	0.4646	Disease	0.744
rs1277504823	L147P	Disease	Disease	7	FALSE	0.4933	Disease	0.708
rs1006881400	L137F	Neutral	Neutral	1	FALSE	0.4907	Neutral	0.243
rs1424596709	Y111C	Disease	Disease	5	FALSE	0.2325	Disease	0.756
rs779219330	P90L	Disease	Neutral	0	FALSE	0.4727	Disease	0.643
rs1224176226	G71E	Disease	Disease	6	TRUE	0.5016	Disease	0.693
rs1254570702	Q70H	Disease	Disease	6	FALSE	0.411	Disease	0.66
rs1388347344	Y66C	Disease	Disease	6	FALSE	0.4933	Disease	0.721
rs1014837070	Y66H	Disease	Disease	6	FALSE	0.4933	Disease	0.685
rs368411105	L187F	Disease	Neutral	5	FALSE	0.159	Neutral	0.185
rs758693208	L144P	Disease	Disease	7	FALSE	0.3125	Disease	0.708
rs762890518	N127K	Disease	Disease	4	FALSE	0.4626	Neutral	0.384
rs1437018496	T120N	Disease	Disease	5	FALSE	0.4014	Disease	0.67
rs760089055	L198S	Disease	Disease	3	FALSE	0.3125	Neutral	0.18
rs1327813126	V164M	Disease	Disease	2	FALSE	0.3183	Disease	0.691
rs1273810480	N127D	Disease	Disease	3	FALSE	0.3911	Neutral	0.21
rs148200166	R107L	Disease	Disease	7	FALSE	0.2662	Disease	0.724
rs868113406	A32T	Disease	Neutral	3	FALSE	0.142	Neutral	0.187
rs1442940888	L25F	Disease	Neutral	0	FALSE	0.255	Neutral	0.278
rs140799590	P73L	Disease	Disease	5	FALSE	0.38	Disease	0.636
rs1313642036	K175N	Disease	Disease	2	FALSE	0.2831	Disease	0.665
rs533306784	R154C	Disease	Disease	5	FALSE	0.164	Disease	0.692
rs750628718	C153Y	Disease	Disease	4	FALSE	0.3711	Disease	0.731
rs373784036	L116V	Disease	Disease	5	FALSE	0.3313	Disease	0.643
rs371771485	T97M	Disease	Neutral	3	FALSE	0.0978	Neutral	0.454
rs758458009	L62V	Disease	Disease	2	FALSE	0.3073	Neutral	0.163
rs780802848	N56S	Disease	Disease	6	FALSE	0.3945	Neutral	0.463

Table 3. Prediction of effect of nsSNPs on protein stability, amino acid conservation.

		MuPro		I-Mutant		I-Stable	Con- surf
dbSNP	Mutations	Prediction	Detal Delta	Prediction	RI	Prediction	score
rs775324532	C156F	Decrease stability	-0.84973772	Decrease	2	Decrease	9,b,s
rs760902711	C153G	Decrease stability	-1.7817825	Decrease	7	Decrease	9,b,s
rs1277504823	L147P	Decrease stability	-2.1453046	Decrease	5	Decrease	7,b
rs1006881400	L137F	Decrease stability	-1.0630089	Decrease	8	Decrease	9,b,s
rs1424596709	Y111C	Decrease stability	-0.93664478	Decrease	6	Decrease	7,b
rs779219330	P90L	Decrease stability	-0.1497019	Decrease	6	Decrease	9,e,f
rs1224176226	G71E	Increase stability	0.0751969	Decrease	3	Increase	8,e,f
rs1254570702	Q70H	Decrease stability	-0.9469835	Decrease	7	Decrease	9,e,f
rs1388347344	Y66C	Decrease stability	-0.8921418	Decrease	1	Decrease	7,b
rs1014837070	Y66H	Decrease stability	-1.4628795	Decrease	6	Decrease	7,b

rs368411105	L187F	Decrease stability	-0.6587313	Decrease	3	Decrease	6,b
rs758693208	L144P	Decrease stability	-2.3331058	Decrease	5	Decrease	3,b
rs762890518	N127K	Decrease stability	-1.1608516	Decrease	3	Decrease	8,e,f
rs1437018496	T120N	Decrease stability	-1.0394151	Decrease	8	Decrease	7,e
rs760089055	L198S	Decrease stability	-1.0669967	Decrease	9	Decrease	6,b
rs1327813126	V164M	Decrease stability	-0.8564677	Decrease	8	Decrease	9,b,s
rs1273810480	N127D	Decrease stability	-0.6367277	Increase	2	Increase	8,e,f
rs148200166	R107L	Decrease stability	-0.4619755	Decrease	8	Decrease	5,e
rs868113406	A32T	Decrease stability	-0.84281051	Decrease	4	Decrease	5,e
rs1442940888	L25F	Decrease stability	-0.5930456	Decrease	8	Decrease	4,b
rs140799590	P73L	Decrease stability	-0.0013742	Decrease	8	Increase	8,e,f
rs1313642036	K175N	Decrease stability	-0.8882829	Increase	5	Increase	3,e
rs533306784	R154C	Decrease stability	-0.7401985	Decrease	7	Decrease	5,e
rs750628718	C153Y	Decrease stability	-1.0369501	Increase	0	Increase	9,b,s
rs373784036	L116V	Decrease stability	-1.063049	Decrease	9	Decrease	7,b
rs371771485	T97M	Decrease stability	-0.0996769	Decrease	2	Increase	5,e
rs758458009	L62V	Decrease stability	-1.684606	Decrease	8	Decrease	8,b
rs780802848	N56S	Decrease stability	-0.5648586	Decrease	4	Increase	9,e,f

Table 4. Wild and mutant structure validation and superimposition value.

			Procheck		TMAlign			
dbSNP	**Mutations**	**Errat quality factor%**	**Core %**	**Allowed %**	**Generously%**	**Disallowed%**	**TM score**	**RMSD**
Native LIF protein	Mutants	86.2857	90.8	6.9	1.7	0.6	-	-
rs775324532	C156F	75	90.2	8.7	0.6	0.6	0.96626	1.17
rs760902711	C153G	89.0173	89.5	9.3	0.6	0.6	0.97455	0.9
rs1277504823	L147P	76.5714	89.5	8.1	1.2	1.2	0.97907	0.81
rs1424596709	Y111C	79.661	89.6	7.5	2.3	0.6	0.97146	0.91
rs1254570702	Q70H	76.1364	91.9	6.4	1.2	0.6	0.97257	1.02
rs1388347344	Y66C	84.3931	89.6	8.7	1.2	0.6	0.9818	0.74
rs1014837070	Y66H	86.2069	90.8	8.1	0.6	0.6	0.97329	0.99
rs1437018496	T120N	83.908	89.6	9.2	0.6	0.6	0.97184	1.11
rs1327813126	V164M	80.9249	89.6	8.7	1.2	0.6	0.96824	1.1

Table 5. Showed the protein-ligand interactions for both native and mutant structures. Residues were distributed based on their ten types of interactions including conventional hydrogen bonds, carbon hydrogen bonds, pi-charged, pi-sigma, pi-pi/pi-pi T shaped and alkyl, pi-alkyl and unfavorable donor.

Ligands	protein	bindingaffinity	Convention-	Carbon-Hy-	Pi-Cha	Pi-Sig	Pi-Pi-	Pi-Pi-shaped	AlkylandPi-Alkyl	Pi-Sul	Unfavoura-	Halogens(fluorine)

			alHy-drogen-Bonds	dro-gen-Bonds	rg ed	m a	sh ap ed			p h ur	ble-donor-Donor	
Lona-parisan	LIF	-8		GLY83					LYS175			GLY83
	C156F	-7.5	VAL110,THR114,ARG145	HIS163					ARG145			GLY113
EC330	LIF	-7.8	GLN186	THR172					LYS175, LYS182			ASP88
	C156F	-7.1	ARG154,CYS153	ARG154					ARG154			
Cou-mestrol	LIF	-7.5							LYS182, LYS175			
	C156F	-6.4	ASN150,CYS40, GLU98			ARG154	TYR159	TYR 160	ARG154	CYS40		
Estra-diol	LIF	-7.3							LYS80, LYS175			
	C156F	-7.2	LEU44						ARG154, PHE156, ALA35,PRO29,ILE27			
Desmet hylmif-epris-tone	LIF	-7.1		LYS182,GLN186					LYS175			
	C156F	-6.8	SER149		ARG37				ARG145,VAL10, TYR168		ARG37	
GH1	LIF	-7.1	GLY174, SER173, LEU81	GLN186,LYS182		LYS175			LYS182		SER173	
	C156F	-6.6	VAL166	THR114	ARG145	VAL110					ARG145	
Tor-ipris-tone	LIF	-7.1		ASP171								
	C156F	-7		ASP171								
Mife-pristone	LIF	-7	LEU81						LYS182			
	156F	-6.9	LEU81						LYS175, LYS182			