Research article

Identification of High risk nsSNPs in Human *TP53* Gene Associated with Li–Fraumeni Syndrome: An *In Silico* Analysis Approach

Mujahed I. Mustafa¹, Naseem S. Murshed², Mazen A. Elbasher³, and Abdelrafie M. Makhawi¹

1- Department of Biotechnology, University of Bahri, Khartoum, Sudan.

2- Department of Microbiology, International University of Africa, Khartoum, Sudan.

3- Department of Chemistry, University of Bahri, Khartoum, Sudan.

Correspondence should be addressed to Mujahed I. Mustafa; mujahedibrahim44@gmail.com

Abstract

Background: Li–Fraumeni syndrome (LFS) is a cancer–prone conditions caused by a germline mutation of the *TP53* gene on chromosome 17p13.1. It has an autosomal dominant pattern of inheritance with high penetrance. **Purpose:** The aim of this study is to identify the high-risk pathogenic nsSNPs in *PT53* gene that could be involved in the pathogenesis of Li–Fraumeni syndrome. **Methods:** The nsSNPs in the human *PT53* gene retrieved from NCBI, were analyzed for their functional and structural consequences using various *in silico* tools to predict the pathogenicity of each SNP. SIFT, Polyphen, PROVEAN, SNAP2, SNPs&Go, PHD-SNP, and P-Mut were chosen to study the functional inference while I-Mutant 3.0, and MUPro tools were used to test the impact of amino acid substitutions on protein stability by calculating $\Delta\Delta G$ value. The effects of the mutations on 3D structure of the *PT53* protein were predicted using RaptorX and visualized by UCSF Chimera. **Results:** A total of 845 *PT53* nsSNPs were analyzed. Out of 7 nsSNPs of *PT53* three of them (T118L, C242S, and I251N) were found high-risk pathogenic. **Conclusion:** In this study, out of 7 predicted high-risk pathogenic nsSNPs, three high-risk pathogenic nsSNPs of *PT53* gene were identified, which could be used as diagnostic marker for this gene. The combination of sequence-based and structure- based approaches is highly effective for pointing pathogenic regions.

Keywords: *In Silico* analysis; *TP53*; Diagnostic markers; Li–Fraumeni syndrome; single nucleotide polymorphisms.

1- Introduction:

Li–Fraumeni syndrome (LFS) is a cancer–prone conditions caused by a germline mutation of the TP53 (tumor protein 53) gene on chromosome 17p13.1 [1, 2]. first described in 1969 by Li and Fraumeni [3]. It has an autosomal dominant pattern of inheritance with high penetrance. LFS related typically with sarcomas, premenopausal breast cancer, brain tumors, and adrenocortical carcinomas [3-6], which it makes it one of the most devastating cancer predisposition syndromes [7]. LFS has long been seen as a rare syndrome but recent studies proves otherwise. The prevalence of LFS is about 1:5000 with a high interregional variance [7].

Two published algorithms are utilized to identify patients at risk of LFS who would benefit from molecular testing, the classical LFS criteria and the Chompret criteria [8, 9]. A *TP53* germline mutation was found in 29%and 35% of French and American families, respectively, who fulfilled the 2001 Chompret criteria [10, 11].

80% of the LFS cases had been reported due to germline *TP53* mutations. since approximately 20 % of LFS families do not exhibit TP53 mutations [4, 12, 13].The underlying genetic mutations in these families remain to be discovered. Yet, de novo mutations occur in ~10%–20% of LFS cases [3, 14].

Recent studies have identified large gene deletions in some families negative for point mutations [15, 16] Despite intensive search, no other gene has been significantly associated with LFS. Reports that germline mutations in the *CHEK2* gene may predispose to LFS have not been substantiated [17, 18]. Germline mutations in *CHEK2*, including the common 1100delC frameshift mutation in exon 10, are currently considered as low-penetrance mutations associated with a three to four fold increase in the risk of breast cancer [19], which is one of the most frequent cancer in adult patients with LFS [20-22].

The PT53 protein is a transcription factor expressed in most cell types, upregulating the transcription of target genes involved in cell cycle arrest, DNA repair, apoptosis, and senescence, in response to DNA damage [23]. The vital role of TP53 mutation in causing LFS is highlighted by the point that mice lacking functional trp53 alleles develop lymphomas and sarcomas. Knock-in mice with germline missense mutations in trp53 simulating LFS develop more destructive tumors as matched with mice lacking trp53 alleles, suggesting that dominant gain of-function effects [24].

Disease-causing single nucleotide polymorphisms (SNPs) are frequently found to arise at evolutionarily conserved regions; these have a key role at structural and functional levels of the protein [25, 26]. The capability to calculate whether a particular SNP is deleterious or not is very important for the prognosis of disorder. SNPs may or may not affect protein function [27]. Therefore, it is necessary to understand the

association between the SNPs and its phenotypic influence which might be useful in analyzing the reasons of numerous diseases or disorders. The single most frequent mutation is p.R337H, an atypical variant of incomplete penetrance which is present at high prevalence in the population of Southern Brazil due to a widespread founder effect [28-30].

Lately, screening of SNPs by the computational approach is very hot topic in revealing the mutations that are affects protein structure and function [31]. The nsSNPs of the *PT53* gene has not been considered to date to observe their functional and structural effects on the molecular level. To address this matter, this study therefore used several bioinformatics tools for the identification of pathogenic SNPs in the coding region of *PT53* gene, which could be used as diagnostic markers for this devastating syndrome.

2 Materials and Methods:

2.1 nsSNPs Data retrieval:

The nsSNPs data were retrieved from SNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) while the protein reference sequence (ID: P04637) was retrieved from Uniprot database [32].

2.2 Prediction of nsSNPs Pathogenicity:

2.2.1. SIFT:

It is the first in silico functional analysis that calculates whether an amino acid alteration affects protein function or not. Sorting Intolerant From Tolerant (SIFT) scores < 0.05 are expected to be damaging altered amino acid, otherwise it is considered to be tolerant [33].

2.2.2. Polyphen:

It is a trained machine learning to predict whether an amino acid replacement affects protein function and structure or not, by calculating position-specific independent count (PSIC) for each SNP at a time. There are 2 outputs whether probably damaging (values are more frequently 1) and possibly damaging or benign (values range from 0 to 0.95) [34].

2.2.3. PROVEAN:

It is an online in silico functional analysis tool that calculates whether an amino acid replacement has an influence on the organic function of a protein stranded on the alignment-based score. If the PROVEAN score ≤ -2.5 , the protein variant is expected to have a “deleterious” effect, whereas if the PROVEAN score is > -2.5 , the variant is expected to have a “neutral” effect [35].

2.2.4. SNAP2:

It is a trained functional analysis tool that differentiates between effect and neutral SNPs by taking various features into validation. SNAP2 got an accuracy of 83%, which has 2 expectations: effect (positive score) or neutral (negative score). It is considered an important and substantial enhancement over other methods [36].

2.2.5. SNPs&Go:

It is a trained machine learning based on the technique to precisely calculates the deleterious associated alterations from protein sequence. SNPs&Go collects in a unique framework information derived from protein sequence, evolutionary information, and function as coded in the Gene Ontology terms and underperforms other available predictive method (PhD-SNP) [37].

2.2.6. P-Mut:

It is a web-based tool for the explanation of SNP alternates on proteins, which allows the rapid and precise calculation (80%) of the compulsive features of each SNP grounded on the practice of neural networks [38].

2.3. Protein Stability Prediction:

2.3.1. I-Mutant 3.0:

I-Mutant is a support vector machine (SVM)- based tool. I-Mutant predicts whether the protein mutation stabilizes or destabilizes the protein structure by calculating free energy change by coupling predictions with the energy-based FOLD-X tool [39].

2.3.2. MUPro:

It is a structural analysis online tool for the calculation of protein stability variations in arbitrary SNPs. The value of the energy change is expected, and assurance mark between -1 and 1 for evaluating the assurance of the expectation is calculated. A score of <0 means the mutant decreases the protein stability; conversely, a score of >0 means the mutant increases the protein stability [40].

2.4. Evolutionary conservation analysis of nsSNPs

ConSurf server It is a web server that offers evolutionary conservation summaries for proteins of known structure in the protein data bank. ConSurf spots the parallel amino acid sequences and runs multi-alignment methods. The conserved amino acid across species flags its position using specific algorithm [41].

2.5. Protein secondary structure prediction and interactions:

PSIPRED was also used to predict and validates the secondary structure of *PT53* protein. The PSIPRED server developed by PSI-BLAST based on specific matrices. The input was the protein sequence in FASTA format, while the output was Protein secondary structures [42].

2.6. 3D structure prediction, visualization and physiochemical changes of substitutions:

2.6.1. RaptorX:

It have been used to generate a 3D structural model for wildtype *PT53*. The FASTA format sequence of *PT53* protein was retrieved from UniProt; it was then used as an input to predict the 3D structure of human *PT53* protein [43].

2.6.2. UCSF Chimera:

It is a visualization tool of 3D structure prototype. A predicted model was formed by RaptorX to visualize and compare the wild and mutant amino acids using UCSF Chimera [44].

2.6.3. HOPE:

It is a server to search protein 3D structures by data mining from several sources databases. It was used to analyze the physiochemical changes induced by the mutations and to confirm the outcomes that we achieved earlier. The inputs were the protein sequence and a single mutation, while the outputs were the physiochemical changes triggered by the mutations [45].

Figure 1: Graphic representation of *TP53* gene work flow.

3. Results:

3.1. Prediction of deleterious SNPs of *TP53* gene:

The total number of SNPs in the coding region of *TP53* gene that were retrieved from dbSNP database were 845 nsSNPs. These SNPs were submitted into diverse functional analysis tools (**Figure 1**). 127 out of 845 nsSNPs were found to be affect by SIFT, 114 damaging SNPs (5 possibly damaging and 120 probably damaging) by using Polyphen 189 were found to be deleterious by PROVEAN and 157 were predicted deleterious by SNAP2 (**Table 1**). The quad-positive deleterious SNPs from the previous 4 analysis tools, out of 845 SNPs there were 7 predicted pathogenic SNPs, and then were submitted into SNPs&Go, PHD-SNP, and P-Mut to get more accurate prediction that could effects on the functional influence; the triple positive results from the three tools were found to be 4 SNPs (**Table 2**).

3.2. Prediction of protein structural stability:

The protein stability analysis with these 4 SNPs (T118L, K120, C242S, and I251N) was performed by using I-Mutant and MUPro, which discovered that all SNPs decreased the protein stability except K120P, which increased the protein stability. All further analysis was hold for these SNPs (T118L, C242S, and I251N).

Figure 2: Venn diagram representing the effects of high-risk nsSNPs that shared between SNPs&Go, PHD-SNP, and P-Mut tools.

Figure 3: Venn diagram representing the effects of high-risk nsSNPs of PT53 on protein stability shared between I-Mutation 3.0, and MUPro servers.

3.3. Evolutionary conservation analysis of nsSNPs:

The conservation regions of TP53 protein had been predicted by Consurf server; the result shows 3 SNPs (T118L, C242S, and I251N) located in highly conserved regions.

3.4. Protein secondary structure prediction and interactions:

The secondary structure of TP53 was evaluated by PSIPRED server. The results revealed a mixture of coil, alpha helix and extended strand. The random coil was observed to be the main secondary structural motif (63.61%), followed by Alpha helix (19.34%) and extended strand (17.05%) as generated by PSIPRED (Figure 5).

3.5. 3D structure prediction, visualization and physiochemical changes of substitutions:

In order to study the biophysical properties of these mutations, Project HOPE server was used to serve this purpose, RaptorX was used to predict a 3D structure model for TP53 protein; while UCSF Chimera was used to visualize the amino acids change (Figure 2). In (Figure 3): (T118L): the amino acid Threonine changes to Leucine at position 118. The mutant residue is bigger than the wild-type residue. The mutant residue is more hydrophobic than the wild-type residue. The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein.

In (Figure 4): (C242S): the amino acid Cysteine changes to Serine at position 242. The wild-type residue is more hydrophobic than the mutant residue. The mutation will cause loss of hydrophobic interactions in the core of the protein.

In (Figure 5): (I251N): the amino acid Isoleucine changes to Asparagine at position 251. The mutant residue is bigger than the wild-type residue. The wild-type residue is more hydrophobic than the mutant residue. The wild-type residue is very conserved.

Table 1: Pathogenic SNPs in *TP53* gene as predicted by four computational tools:

SUB	SIFT prediction	Score	Polyphen2 prediction	Score	PROVEAN prediction	Score	SNAP2 Prediction	Score
K120P	Deleterious	0	Probably damaging	1	Deleterious	-5.644	effect	75
I251N	Deleterious	0	Probably damaging	1	Deleterious	-6.591	effect	13
T118L	Deleterious	0	Probably damaging	1	Deleterious	-5.028	effect	40
G117M	Deleterious	0	Probably damaging	1	Deleterious	-4.834	effect	34
F113V	Deleterious	0	Probably damaging	1	Deleterious	-6.155	effect	48
F113R	Deleterious	0	Probably damaging	1	Deleterious	-7.975	effect	75
C242S	Deleterious	0	Probably damaging	1	Deleterious	-9.759	effect	85

Table 2: The most damaging SNPs in *TP53* gene as predicted by three different online tools:

SUB	SNPs&Go Prediction	RI	Probability	PHD-SNP Prediction	RI	Probability	P-Mut prediction	Probability
K120P	Disease	5	0.763	Disease	7	0.839	Disease	0.86 (91%)
I251N	Disease	7	0.835	Disease	7	0.863	Disease	0.86 (91%)
T118L	Disease	5	0.741	Disease	5	0.744	Disease	0.78 (88%)
G117M	Neutral	1	0.448	Neutral	2	0.405	Disease	0.56 (81%)
F113V	Neutral	0	0.484	Disease	7	0.854	Disease	0.78 (88%)
F113R	Neutral	2	0.379	Disease	7	0.854	Disease	0.86 (91%)
C242S	Disease	8	0.893	Disease	6	0.802	Disease	0.91 (93%)

Table 3: Shows protein stabilities of *TP53* nsSNPs by I-Mutant 3.0 and MUPro:

SUB	I-Mutant Prediction	RI	DDG Value Prediction	MUPro Prediction	MUPro Score
K120P	Increase	6	0.01	Decrease	-0.67801266
I251N	Decrease	5	-1.99	Decrease	-1.9395329
T118L	Decrease	6	-0.73	Decrease	-0.54246124
C242S	Decrease	7	-1.13	Decrease	-2.8407208

The conservation scale:

Figure 4: The conserved amino acids across species in TP53 protein were determined using ConSurf. *e*: exposed residues according to the neural-network algorithm are indicated in orange letters. *b*: residues predicted to be buried are demonstrated via green letters. *f*: predicted functional residues (highly conserved and exposed) are indicated with red letters. *s*: predicted structural residues (highly conserved and buried) are demonstrated in blue letters. The black boxes indicate the high-risk nsSNPs.

Figure 5: Secondary structure of *PT53* showing beta helix and coil. The results highlight a mix distribution of coil and alpha helix, as generated by PSIPRED server. The black boxes indicate the high-risk nsSNPs.

Figure 6: The 3D structure of PT53 protein models by RaptorX.

Figure 7: (T118L): The amino acid Threonine changes to Leucine at position 118. The green color indicates the wild amino acid, the red color indicates the mutant one, while the gray color indicates the rest of TP53 protein structure. The 3D observations were done by using UCSF Chimera.

Figure 8: (C242S): the amino acid Cysteine changes to Serine at position 242.

Figure 9: (I251N): the amino acid Isoleucine changes to Asparagine at position 251.

Figure 10: Our strategy to classify the most pathogenic nsSNPs .

4- Discussion:

In this study, various sequence-based approaches were performed, because they are more superior to structure-based approaches. Structure-based approaches are restricted to a known 3D structure; on the other hand, sequence-based approaches can be employed in proteins with unknown 3D structures. A combination of multiple predictors revealed better predictions in many recent reports for identifying deleterious nsSNPs in many genes [46-52]. Typically, at least five of *in silico* tools should be considered increasing the consensus on the effect of SNPs [53]; nevertheless, 14 different computational algorithms were engaged to categorize the most pathogenic nsSNPs from *PT53* gene. In the present study a total of three pathogenic SNPs (T118L, C242S, and I251N) were identified after extensive computational analysis.

All three SNPs were also found to be deleterious in common with I-Mutant 2.0 and MUPro servers by destabilizes the protein stability. Furthermore, SNPs that were conserved across the evolutionary perspective are more significant than those that were not conserved, because they are structurally and functionally important for the protein. For this purpose, the ConSurf server was used for evolutionary conservation study. This analysis showed that T118, C242 and I251 were highly conserved in the *PT53* which can directly disturb the protein function and could affect the protein interactions with other proteins.

Our analysis consequently achieves that these SNPs might cause damage to the *PT53* protein by disturbing its stability. Nevertheless, the 3D structure of the protein shows a vital role in sympathetic the general effects of nsSNPs on protein function. Thus, 3D protein modeling was achieved by the UCSF Chimera following the improvement of the model to observe the mutational impact on *PT53* protein.

In most cases, proteins perform biological functions as temporary or permanent complexes by interacting with other macromolecules. Therefore, mutations in or near some special amino acids that contribute to the functional spatial conformation are at a high risk of causing pathologies. According to the analysis from HOPE, The mutated residues are located in a domain that is important for binding of other molecules. The mutated residues are in contact with a regulatory domain. These mutations could affect this interaction and thereby disturb regulation of the protein.

Also we searched the related published literature to compare our outcomes that had been found by *in silico* method with other approaches, in (I251N) SNP, is present in 0.03% of AACR GENIE (American Association for Cancer Research) cases. This mutation was previously reported to cause lung adenocarcinoma, colon adenocarcinoma, esophageal adenocarcinoma, adenocarcinoma of the gastroesophageal junction, and ampulla of Vater carcinoma [54]. The second variant (C242S) was previously reported to cause Acute Myeloid Leukemia [55]; while for the last variant (T118L), we didn't found any associated clinical studies.

Although it is more reliable to differentiate deleterious SNPs through experiments, it takes much time to execute tests on all variants. Different approaches offer a certain degree of reliability for hazard prediction. By using our computational analysis methods, mainly focuses on 3 high-risk nsSNPs (T118L, C242S, and I251N) from 845. These 3 nsSNPs comes from an systematic analysis summarized in **(figure 10)**. Certainly, there are several limitations of this study; the main limitation is that it focuses on coding region of *PT53* gene using a combination of bioinformatics analysis tools. The second limitation is that the reported causative nsSNPs were limited in number. The third is that there are some similarities of the disease prediction tool between diverse computational analysis tools because most of them were based on changes in conserved residues over time.

Nevertheless, *in silico* approach remains a cost-effective technique to achieve a swift analysis concerning the predictable effect of variants.

5-Conclusion:

In this study, out of 7 predicted pathogenic nsSNPs, Three high-risk pathogenic nsSNPs of *PT53* gene were identified, the combination of sequence-based and structure-based approaches is highly effective for pointing pathogenic regions. Yet, computational approach can't replace clinical experiments, and their outcomes should be confirmed by further molecular and clinical studies.

Data Availability

All data underlying the results are available as part of the article, and no additional source data are required.

Conflicts of Interest:

The authors declare that they have no conflicts of interest.

Funding Statement:

The authors received no financial support for the research, authorship, and/or publication of this article.

Authors' Contributions:

MIM: conceptualization, formal analysis, investigation, methodology, validation, visualization, and writing (original draft); NSM & MAE: formal analysis, investigation, and validation. AMM: data curation, conceptualization, project administration, supervision, and writing (review and editing). All authors have read and approved the final manuscript.

Acknowledgments:

The authors acknowledge the Deanship of Scientific Research at University of Bahri for the supportive cooperation.

References:

1. Ariffin, H., et al., *Whole-genome sequencing analysis of phenotypic heterogeneity and anticipation in Li-Fraumeni cancer predisposition syndrome*. Proc Natl Acad Sci U S A, 2014. **111**(43): p. 15497-501.
2. Asdahl, P.H., R.P. Ojha, and H. Hasle, *Cancer Screening in Li-Fraumeni Syndrome*. JAMA Oncol, 2017. **3**(12): p. 1645-1646.

3. Correa, H., *Li-Fraumeni Syndrome*. J Pediatr Genet, 2016. **5**(2): p. 84-8.
4. Malkin, D., *Li-fraumeni syndrome*. Genes Cancer, 2011. **2**(4): p. 475-84.
5. Schneider, K., et al., *Li-Fraumeni Syndrome*, in *GeneReviews*(®), M.P. Adam, et al., Editors. 1993, University of Washington, Seattle

Copyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.: Seattle (WA).

6. Frebourg, T., et al., [*Li-Fraumeni syndrome: update, new data and guidelines for clinical management*]. Bull Cancer, 2001. **88**(6): p. 581-7.
7. Dutzmann, C.M., et al., [*Update on Li-Fraumeni syndrome*]. Pathologie, 2019. **40**(6): p. 592-599.
8. Tinat, J., et al., *2009 version of the Chompret criteria for Li Fraumeni syndrome*. J Clin Oncol, 2009. **27**(26): p. e108-9; author reply e110.
9. Bougeard, G., et al., *Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers*. J Clin Oncol, 2015. **33**(21): p. 2345-52.
10. Bougeard, G., et al., *Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families*. J Med Genet, 2008. **45**(8): p. 535-8.
11. Gonzalez, K.D., et al., *Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations*. J Clin Oncol, 2009. **27**(8): p. 1250-6.
12. McBride, K.A., et al., *Li-Fraumeni syndrome: cancer risk assessment and clinical management*. Nat Rev Clin Oncol, 2014. **11**(5): p. 260-71.
13. Achatz, M.I. and G.P. Zambetti, *The Inherited p53 Mutation in the Brazilian Population*. Cold Spring Harb Perspect Med, 2016. **6**(12).
14. Chompret, A., et al., *P53 germline mutations in childhood cancers and cancer risk for carrier individuals*. Br J Cancer, 2000. **82**(12): p. 1932-7.
15. Bougeard, G., et al., *Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene*. Oncogene, 2003. **22**(6): p. 840-6.
16. Schiffman, J.D., et al., *Identification of a novel p53 in-frame deletion in a Li-Fraumeni-like family*. Pediatr Blood Cancer, 2008. **50**(4): p. 914-6.
17. Bell, D.W., et al., *Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome*. Science, 1999. **286**(5449): p. 2528-31.
18. Vahteristo, P., et al., *p53, CHK2, and CHK1 genes in Finnish families with Li-Fraumeni syndrome: further evidence of CHK2 in inherited cancer predisposition*. Cancer Res, 2001. **61**(15): p. 5718-22.
19. Weischer, M., et al., *CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls*. J Clin Oncol, 2008. **26**(4): p. 542-8.
20. Penkert, J., et al., *Breast cancer patients suggestive of Li-Fraumeni syndrome: mutational spectrum, candidate genes, and unexplained heredity*. Breast Cancer Res, 2018. **20**(1): p. 87.
21. Nandikolla, A.G., S. Venugopal, and J. Anampa, *Breast cancer in patients with Li-Fraumeni syndrome - a case-series study and review of literature*. Breast Cancer (Dove Med Press), 2017. **9**: p. 207-215.
22. Khincha, P.P., et al., *Reproductive factors associated with breast cancer risk in Li-Fraumeni syndrome*. Eur J Cancer, 2019. **116**: p. 199-206.
23. Reinhardt, H.C. and B. Schumacher, *The p53 network: cellular and systemic DNA damage responses in aging and cancer*. Trends Genet, 2012. **28**(3): p. 128-36.

24. Kim, M.P., Y. Zhang, and G. Lozano, *Mutant p53: Multiple Mechanisms Define Biologic Activity in Cancer*. Front Oncol, 2015. **5**: p. 249.
25. Wakui, M., *[Analysis of single nucleotide polymorphisms (SNPs)]*. Rinsho Byori, 2013. **61**(11): p. 1008-17.
26. Shaw, G., *Polymorphism and single nucleotide polymorphisms (SNPs)*. BJU Int, 2013. **112**(5): p. 664-5.
27. Riley, J.H., et al., *The use of single nucleotide polymorphisms in the isolation of common disease genes*. Pharmacogenomics, 2000. **1**(1): p. 39-47.
28. Garritano, S., et al., *Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of Southern Brazil: evidence for a founder effect*. Hum Mutat, 2010. **31**(2): p. 143-50.
29. Achatz, M.I., et al., *The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families*. Cancer Lett, 2007. **245**(1-2): p. 96-102.
30. Pinto, E.M. and G.P. Zambetti, *What 20 years of research has taught us about the TP53 p.R337H mutation*. Cancer, 2020. **126**(21): p. 4678-86.
31. Hu, J. and C. Yan, *Identification of deleterious non-synonymous single nucleotide polymorphisms using sequence-derived information*. BMC Bioinformatics, 2008. **9**: p. 297.
32. UniProt Consortium, T., *UniProt: the universal protein knowledgebase*. Nucleic Acids Res, 2018. **46**(5): p. 2699.
33. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function*. Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
34. Capriotti, E. and R.B. Altman, *Improving the prediction of disease-related variants using protein three-dimensional structure*. BMC Bioinformatics, 2011. **12 Suppl 4**(Suppl 4): p. S3.
35. Choi, Y., et al., *Predicting the functional effect of amino acid substitutions and indels*. PLoS One, 2012. **7**(10): p. e46688.
36. Hecht, M., Y. Bromberg, and B. Rost, *Better prediction of functional effects for sequence variants*. BMC Genomics, 2015. **16 Suppl 8**(Suppl 8): p. S1.
37. Calabrese, R., et al., *Functional annotations improve the predictive score of human disease-related mutations in proteins*. Hum Mutat, 2009. **30**(8): p. 1237-44.
38. López-Ferrando, V., et al., *PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update*. Nucleic Acids Res, 2017. **45**(W1): p. W222-w228.
39. Capriotti, E., P. Fariselli, and R. Casadio, *I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W306-10.
40. Cheng, J., A. Randall, and P. Baldi, *Prediction of protein stability changes for single-site mutations using support vector machines*. Proteins, 2006. **62**(4): p. 1125-32.
41. Ben Chorin, A., et al., *ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins*. Protein Sci, 2020. **29**(1): p. 258-267.
42. Buchan, D.W.A. and D.T. Jones, *The PSIPRED Protein Analysis Workbench: 20 years on*. Nucleic Acids Res, 2019. **47**(W1): p. W402-w407.
43. Wang, S., et al., *RaptorX-Property: a web server for protein structure property prediction*. Nucleic Acids Res, 2016. **44**(W1): p. W430-5.
44. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.

45. Venselaar, H., et al., *Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces*. BMC Bioinformatics, 2010. **11**: p. 548.
46. Mustafa, M.I. and T.A. Abdelhameed, *Novel Deleterious nsSNPs within MEFV Gene that Could Be Used as Diagnostic Markers to Predict Hereditary Familial Mediterranean Fever: Using Bioinformatics Analysis*. 2019. **2019**: p. 1651587.
47. Mustafa, M.I., et al., *In Silico Genetics Revealing 5 Mutations in CEBPA Gene Associated With Acute Myeloid Leukemia*. Cancer Inform, 2019. **18**: p. 1176935119870817.
48. Mustafa, M.I., et al., *Extensive In Silico Analysis of ATL1 Gene : Discovered Five Mutations That May Cause Hereditary Spastic Paraplegia Type 3A*. Scientifica (Cairo), 2020. **2020**: p. 8329286.
49. Mustafa, M.I., et al., *In silico analysis of the functional and structural consequences of SNPs in human ARX gene associated with EIEE1*. Informatics in Medicine Unlocked, 2020. **21**: p. 100447.
50. Yousefi, T., et al., *In silico analysis of non-synonymous single nucleotide polymorphism in a human KLK-2 gene associated with prostate cancer*. Meta Gene, 2019. **21**: p. 100578.
51. Mustafa, M.I., et al., *Novel Mutations within PRSS1 Gene that Could Potentially Cause Hereditary Pancreatitis: Using Bioinformatics Approach*. International Annals of Science, 2019. **8**(1): p. 96-113.
52. Mustafa, M.I., et al., *Identification of Novel Key Biomarkers in Simpson-Golabi-Behmel Syndrome (SGBS): Evidence from Bioinformatics Analysis*. International Annals of Science, 2019. **8**(1): p. 1-11.
53. Brown, D.K. and Ö. Tastan Bishop, *Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level*. Glob Heart, 2017. **12**(2): p. 151-161.
54. AACR Project GENIE: *Powering Precision Medicine through an International Consortium*. Cancer Discov, 2017. **7**(8): p. 818-831.
55. Dutta, S., et al., *Functional Classification of TP53 Mutations in Acute Myeloid Leukemia*. 2020. **12**(3).