Comprehensive Characterization of the Coding and Non-coding Single Nucleotide Polymorphisms in the Tumor Protein p63 (TP63) Gene Using In Silico Tools.

Shamima Akter
  George Mason University

Shafaat Hossain
  University of Dhaka

Md. Ismail Hosen
  University of Dhaka

Hossain Uddin Shekhar (✉ hossainshekhar@du.ac.bd)
  University of Dhaka

Research Article

Abstract

Single Nucleotide Polymorphisms (SNPs) help to understand the phenotypic variations in humans. Numerous studies have examined the association of SNPs with various complex diseases. Researchers have identified the association of SNPs of genes through Genome-wide association study (GWAS). A number of GWAS have identified a loci located in the TP63 gene to be significantly associated with the risk of urinary bladder cancer. However, there is not any study characterizing the SNPs located at the TP63 gene for their functional and structural significance. Hence, the study aimed to comprehensively characterize SNPs in the human TP63 gene for their functional and structural significance. We investigated and evaluated the genomic variations affecting the expression, structure, and function of the TP63 protein. The study systematically retrieved nsSNPs information for the TP63 gene from the dbSNP database. We screened and analyzed both nsSNPs and non-coding SNPs in TP63 protein using a wide variety of computational tools to find the risk of pathogenicity. A total of 17 nsSNPs were identified using the 13 bioinformatics tools (i.e., SIFT, CADD, PROVEAN, PolyPhen2, PANTHER, PhD-SNP, SNP&GO, I-Mutant 2.0, ClinVar, Mutpred2, ConSurf, HOPE, and Mutation 3D) along with domain analysis. These pathogenic mutations cause a decrease in protein stability according to I-Mutant2.0. HOPE predicted 17 SNPs to have significant effect on TP63 protein structure and function. 12 nsSNPs were found in highly conserved position in TP63 inferring the damaging effect on the structure and function of the protein. Swiss PDB Viewer showed loss of hydrogen bonds and increased energy due to the SNPs. Molecular docking showed the reduction of the binding affinity of proteins for DNA and loss of hydrogen bonds. Six non-coding SNPs were found in miRNA binding sites in gene showing the effect on protein regulation using PolymiRTS and five non-coding SNPs were identified in single tissue expression quantitative trait loci (eQTL) in lung tissue, heart tissue (LV), and cerebral hemisphere (Brain) according to GTEx portal. The characterization of nsSNPs and non-coding SNPs will support researchers to focus on TP63 gene loci and ascertain their association with certain diseases.

1. Introduction

Human Tumor Protein (TP63) gene is 4944 bp (4.94 kb) long and located in chromosome 3 at 3q28 locus. It has 12 isoforms listed in UniProt database.TP63 isoform1 gene encodes TP63) protein of 680 amino acid long. TP63 gene encodes for a transcription factor (TF) that regulates gene expressions in multiple pathways, notably in tumorigenesis and development[1].TP63 protein functions as a transcriptional activator or repressor which binds to DNA in a sequence-specific manner [2] and shares sequence similarity with the tumor suppressor p53 family. Mutations in TP63 are known to cause ectodermal dysplasia in humans. Recent studies have identified and characterized the association of SNPs in human TP63 protein with various disorders as well as different cancers, i.e., lung cancer, urinary bladder cancer, other diseases; i.e., heart disease, adults syndrome [3,4]

In human genome, SNPs are the most common type of genetic variation occurring both in coding and non-coding region of a gene. However, not all SNPs are associated with human diseases. Coding region SNPs are known as nsSNPs or non-synonymous SNPs which have greater impact on the protein structure, function, stability, and solubility through amino acid replacement in protein sequence [5,6]. These nsSNPs can be categorized as either 'pathogenic/deleterious' causing disease phenotypes, or 'tolerated/neutral' causing no effect on protein structure, function or stability [7,8]. The genetic basis of human diseases can be understood better by differentiating deleterious nsSNPs from tolerated nsSNPs that will assist in discovering new therapeutic targets for diseases. Therefore, detection of functionally deleterious nsSNPs in TP63 protein may provide important information regarding multiple diseases, particularly different types of cancers. Different computational tools for predicting deleterious SNPs and their role in protein function, stability and structure maintenance are extensively used [9–12]. Researchers have determined several harmful nsSNPs of human CHK2 gene using computational tools to predict the structural and functional impact on the protein[13]. Similarly, in another study, nsSNPs in human hTERT gene were examined through in silico study to characterize their functional and structural impacts comprehensively [14]. Researchers predicted deleterious effects of nsSNPs of human RASSF5 gene on protein structure and function employing in silco analysis[15].

These studies reveal the effectiveness and usage of computational tools to inspect and confirm genetic variants.

The utilization of various computational methods could be an effective screening and validation method to study SNPs of a specific protein [16]. Therefore, the present study was aimed to determine various nsSNPs of the human TP63 gene using SIFT, CADD, PROVEAN, PolyPhen2, PANTHER, PhD-SNP, SNP&GO, I-Mutant 2.0, ClinVar, Mutpred2, ConSurf, HOPE, and Mutation 3D. In addition, we also decided to (1) detect the effect of nsSNPs on binding affinity with ligand by molecular docking using PyRx, Haddock, Chimera 1.14, (2) Visualize the effects of nsSNPs using Swiss PDB Viewer, and (3) analyze noncoding SNPs in TP63 gene through RegulomeDB, PolymiRTS, and GTEx portal analysis (Figure 1).

2. Materials And Methods

The flowchart showing different steps of this study are presented in Figure 1.

2.1 SNP Data Retrieval

In this study, the human TP63 gene was investigated in the Ensemble genome browser (https://uswest.ensembl.org/index.html). For this study, the transcript encoding the full length TP63 protein (680 a.a; https://www.uniprot.org/uniprot/Q9H3D4) was selected to retrieve the SNPs data [17]. After downloading the SNP data set from the Ensemble genome browser, then we filtered the SNPs from dbSNP database (dbSNP NCBI: https://www.ncbi.nlm.nih.gov/snp/?term=chek2) for analysis.

2.2 Functional analysis nsSNPs

SIFT (Sorting Intolerant From Tolerant, https://sift.bii.a-star.edu.sg/) is used to detect the deleterious nsSNPs. SIFT can differentiate functionally neutral amino acid changes from functionally deleterious ones to predict polymorphisms and mutants [18]. This software presumes that major amino acids will be retained, shifts at particular positions appear to be predicted as deleterious in proteins. If the normalized probability is less than tolerance value (0.05), then substitutions are considered as "deleterious" and equal or higher than 0.05 is considered as "tolerated". rsIDs of each variant of human TP63 protein (UniProt ID: Q9H3D4) and individual amino acid substitutions serve as input in the SIFT.

PolyPhen2(Polymorphism Phenotyping v2, http://genetics.bwh.harvard.edu/pph2/) uses Naive Bayes to classify and predict the functional impacts of allele modifications [19]. For each variant, PolyPhen2 estimates the PSIC (position-specific independent count) based on site-specific sequence conservation along with the difference of scores between the native and mutant variant. PolyPhen2 classifies the SNPs into 3 different classes: (1) benign, (2) probably damaging, or (3) possibly damaging. The input on the PolyPhen2 web server includes the FASTA sequence of human TP63 protein and the individual amino acid substitutions.

CADD (Combined Annotation-Dependent Depletion, https://cadd.gs.washington.edu/) [20] is an SNPs predicting algorithm that prioritizes causal variants in polymorphism studies. CADD integrates multiple annotations in a single metric by comparing variants that have been survived in a natural selection to simulated mutations. CADD prioritizes causal variations in both research and clinical settings. Chromosomal locations of the human TP63 variants are given as input to the CADD web server.

SNPs&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html) uses functional annotations of proteins to assess the impact of single amino acid substitutions [21]. SNPs&GO analysis provides the prediction of SNPs using three tools i.ee SNP&GO, PhD-SNP and PANTEHR. SNPs&GO utilizes support vector machines and input includes the sequence or its three-dimensional protein structure, target variations, and Gene Ontology (GO) term functionality. The output of the algorithm provides the probabilities of each SNP to have human disease association.

PROVEAN (Protein Variation Effect Analyzer, http://provean.jcvi.org/index.php) is a server that assesses functional impact of a substituted amino acid or insertion−deletion mutation on a protein across organisms. PROVEAN can provide high-throughput analysis at both genomic and protein levels for human and mouse variants [22]. For each variant, PROVEAN calculates a score from the alignment of homologous sequences, and a score of -2.5 is considered to be damaging. Input in the PROVEAN web server was the FASTA sequence of human TP63 protein and changes in single amino acid(nsSNPs).

ClinVar (https://www.ncbi.nlm.nih.gov/clinvar) helps to analyze human genetic variations along with annotation of variant disease associations. ClinVar aggregates the known variant interpretations and makes them publicly available on ClinVar database [23]. The ClinVar database was searched by individual amino acid changes to identify disease associated variants.

MutPred2 (http://mutpred.mutdb.org/) enhances pathogenic amino acid substitution prioritization, predicts potentially disease-causing molecular pathways, and returns interpretable distributions of the pathogenicity score on individual genomes [24]. Input in the MutPred2 web server   was the FASTA sequence of human TP63 protein and changes in single amino acid(nsSNPs).

2.3. Structural Analysis

I-Mutant2.0 (http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi) is an algorithm that uses support vector machine to provide an assessment of protein stability change for the variations in single nucleotides [25]. I-Mutant2.0 uses either protein structure or protein sequence for predictions and estimates the shift in protein stability, and simultaneously predicts the corresponding DeltaDeltaG values. The protein sequence of human TP63 protein and individual amino acid substitutions were given as input in the I-Mutant 2.0 web server.

HOPE (http://www3.cmbi.ru.nl/hope/) predicts functional and structural impacts of point mutations. HOPE incorporates data from multiple information sources including measurements of 3D protein coordinates using services of WHAT IF Web, sequence annotations at UniProt database, and predictions of DAS service [26]. Data stored from these sources are used to classify the effects of a mutation on both protein function and three-dimensional structure using a decision scheme. With the help of text, statistics, and animations, HOPE produces a report that is simple to use and easily understandable.

As an algorithm and Web server, Mutation3D (http://mutation3d.org) detects changes in amino acid within protein three dimensional structures. Mutation3D can distinguish functional and nonfunctional variations [27]. The mutation3D Web server enables users to examine the substitution in a number of common formats while offering easy access to examine mutation clusters derived from 6,811 cancer sequencing studies reported from over 975,000 somatic mutations.

We performed the conservancy analysis through ConSurf web server [28]   (https://consurf.tau.ac.il/) to determine the evolutionary conservation of amino acid residues in wild type protein and identify the nsSNPs in those conserved residues.

2.4   Structural Effect Analysis

I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) were used to predict the structure of TP63. We did not find any crystal structures of full length TP63 protein in the protein data bank (PDB). I-TASSER predicted the full-length 3D structure of TP63. I-TASSER uses LOMETS which is a multiple threading alignments approach to construct full–length atomic models through iterative simulations of structure assembly. I-TASSER also uses BioLiP which is a protein function database to infer the function of the protein. To determine the accuracy of the model, both confidence score and sequence identity are considered. Confidence score explains true homology based on the similarity between

query and template sequence. At least 30 % (>30 preferable) sequence similarity or identity between template and query sequence is required to predict model with high accuracy.

Phyre2 predicted the 2RMN (PDB ID:2rmn) structure which is a DNA binding domain of TP63.2RMN was also obtained in PDBSum server (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/FindSequence.pl) searching with full sequence of TP63 and generated the full PROCHECK analyses with the RAMACHANDRAN plot (Figure S5). PROCHECK statistics and plot has been supplied in supplementary section. This 3D structure was extensively used to generate wild type and mutant peptide structures of nsSNPs for Molecular docking analysis.

Swiss PDB Viewer (v4.1.0) (https://spdbv.vital-it.ch/) is used to visualize the wild type and mutated protein to obtain the accuracy of the results. Using the software, the effects of nsSNPs/mutations can be observed in protein's 3D structure, i.e., the gain or loss of hydrogen bonds or clash between atoms or residues.

Molecular docking was performed to assess the impacts on the interactions between DNA and TP63 due to each nsSNPs/mutation as DNA is a direct ligand of TP63 protein. Binding affinity is the measurement of the interactions between molecule and ligand. Docking had been performed using virtual screening tool- PyRx (https://pyrx.sourceforge.io/)[29]. From PDB, 3QYN, a complex of DNA binding domain of TP63 and DNA chains was retrieved, DNA structure was extracted, and a single chain was chosen to perform docking analysis using USCF Chimera1.14 (https://www.rbvi.ucsf.edu/chimera/). In DNA-Protein docking, DNA chain was used as a macromolecule, peptides of wild type protein, and mutant proteins were used as ligands. All wild type and mutant peptides of TP63 were generated using build structure tool in Chimera 1.14[30] with energy minimization. Then in PyRx, DNA molecule, and ligands were uploaded and subsequently converted into PDBQT(.pdbqt) format using AutodockVina[31]. Further DNA-protein docking for TP63 was performed in HADDOCK web server[32](https://wenmr.science.uu.nl/).The binding interactions were visualized in Chimera 1.14.

2.5 Analysis the Functional Consequences of non-coding SNPs

RegulomeDB (https://www.regulomedb.org/regulome-search/) offers an annotation of regulatory SNPs, integrates the information from experimental data sets, computational predictions and manual annotations from ENCODE to attribute scores to variants in order to distinguish the functional SNPs from a wide pool [33]. rsIDs of the individual variants were submitted in the RegulomeDB database to assess the consequences of non-coding SNPs.

PolymiRTS database3.0 (http://compbio.uthsc.edu/miRSNP/) incorporates data from ligation, hybrids sequence, and crosslink experiments in order to analyze miRNA−mRNA interactions [34]. PolymiRTS database analyzes the functional consequences of SNPs in miRNA target sites and seed regions. rsIDs of the individual variants were submitted in the PolymiRTS database to assess the consequences of non-coding SNPs.

GTEx or the Genotype-Tissue Expression project (https://gtexportal.org/home/) the association between genetic alterations and gene expression in human tissues in human tissues. GTEx relates these regulatory mechanisms to traits and diseases [35]. rsIDs of the individual variants were submitted in the GTEx database to assess the consequences of non-coding SNPs.

3. Results

3.1 TP63 SNP data

The SNP data for the TP63 gene contained a total of 64,144 SNPs with transcript ID: ENST00000264731.8. Among these, 455 SNPs were missense or nonsynonymous SNPs (nsSNPs), 242 were synonymous SNPs, 680 SNPs were located at the UTR, 62727 at the Introns, 11 frameshifts, 6 in frame, and 11 stop gained SNPs. Different categories of SNPs are presented in Figure S1.

3.2 Prediction of Functionally Important nsSNPs in the TP63 gene

A variety of tools have been selected for computational analysis for nsSNPs of TP63. Initially, SIFT, PolyPhen2, and CADD were used for the analysis. Out of 455 nsSNPs, 194 SNPs were predicted to be deleterious by SIFT. While using PolyPhen2, only probably damaging criteria was selected and found 121 nsSNPs those were further analyzed by CADD tool. In CADD, likely deleterious criteria have been chosen, after filtering, 28 nsSNPs were predicted to be deleterious (Table S1) and selected for further analysis.

28 nsSNPs predicted as deleterious in Section 4.2 were subjected to PROVEAN analysis. The TP63 protein sequence from the NCBI database was the input sequence for PROVEAN. A cutoff score of -2.5 was selected for deleteriousness. Out of 28 nsSNPs, 23 SNPs were predicted to be deleterious and remaining 5 SNPs were detected as neutral (Table S2). 28 nsSNPs were also analyzed in ClinVar web server. 9 SNPs were evaluated for pathogenicity, 3 SNPs were found to be likely pathogenic, 3 SNPs had uncertain significance and the rest SNPs were not found in the ClinVar repository (Table S2).

SNP	rsID(dbSNP)	Domain function	Methods							
			Mutpred2		PROVEAN		ClinVar	PhD-SNP		
			Mdscore	Mutpred2 impact	Score	Impact	Result	Prediction	Probability	Predicti
R408C	rs1282887680	Oligomerization	0.662		-7.064	Deleterious	not found	Disease	0.771	Disease
R408H	rs751698974	Oligomerization	0.508		-4.461	Deleterious	not found	Disease	0.767	Disease
R376C	rs757536818	Interaction with HIPK21	0.355		-3.65	Deleterious	not found	Disease	0.716	Neutral
C347F	rs1064793282	DNA binding domain	0.932	Gain of Strand (Pr = 0.27 | P = 0.03)	-10.073	Deleterious	Pathogenic	Disease	0.91	Disease
D351G	rs121908844	DNA binding domain	0.863		-6.41	Deleterious	Pathogenic	Disease	0.836	Disease
D355N	rs1553857889	DNA binding domain/ Interaction with HIPK21	0.706		-3.512	Deleterious	Pathogenic	Neutral	0.337	Disease
G349E	rs866267914	DNA binding domain	0.852	Gain - Intrinsic disorder P = 0.04 Loss of Strand P = 0.02	-7.342	Deleterious	Pathogenic	Disease	0.661	Disease
R266Q	rs121908849	DNA binding domain	0.807	Loss of Strand P = 0.02) Altered Stability P = 0.01	-3.612	Deleterious	Pathogenic	Disease	0.806	Disease
R318H	rs121908840	DNA binding domain	0.725	Loss - ADP-ribosylation at R318 P = 0.03	-4.645	Deleterious	Pathogenic	Disease	0.885	Disease
R319H	rs886039442	DNA binding domain	0.742		-4.627	Deleterious	Pathogenic	Disease	0.865	Disease
R337Q	rs113993967	DNA binding domain	0.861	Gain - Strand P = 0.02) Gain - ADP-ribosylation at R338 P = 0.05) Gain - Pyrrolidone carboxylic acid at R337 P = 0.02)	-3.618	Deleterious	Pathogenic	Disease	0.8	Disease
R343Q	rs121908841	DNA binding domain	0.801	Gain of Strand P = 0.03); Altered Stability P = 0.02)	-3.663	Deleterious	Pathogenic	Disease	0.855	Disease
R379C	rs761885185	Interaction with HIPK21	0.515	Loss- Intrinsic disorder P = 0.02); Loss - Phosphorylation at T382 P = 0.01); Loss - Acetylation at K375 | P = 0.01); Altered Disordered interface P = 0.04)	-2.648	Deleterious	Uncertain significance	Neutral	0.424	Disease
R379H	rs765502786	Interaction with HIPK21	0.312		-1.476	Neutral	Uncertain significance	Neutral	0.227	Neutral

L562R	rs774221257	SAM	0.896	Altered Transmembrane protein P = 9.7e-05.	-2.328	Neutral	not found	Disease	0.787	Disease
				Altered Ordered interface P = 0.02.						
				Altered Stability P = 0.03.						
				Loss - Sulfation at Y564 P = 0.03						
R647H	rs774550896	Transactivation inhibition	0.834	Altered Metal binding P = 2.9e-03).	-2.062	Neutral	not found	Disease	0.761	Disease
				Altered DNA binding P = 1.2e-03);						
				Altered Disordered interface P = 0.04).						
				Loss - Proteolytic cleavage at R643 P = 0.02).						
				Altered Transmembrane protein P = 0.03).						
				Altered Stability P = 0.04)						
R655Q	rs764601563	Transactivation inhibition	0.656	Altered Disordered interface P = 0.04.	-1.246	Neutral	not found	Disease	0.755	Disease
				Altered Metal binding P = 0.03.						
				Altered DNA binding P = 0.03.						
				Altered Transmembrane protein P = 0.05)						

All 28 SNPs were also analyzed with SNPs&GO to obtain the disease associated variations (SNPs) in protein sequence. SNP & GO simultaneously performed the analysis for three different tools: PhD-SNP, PANTHER, SNPs&GO, and provided results separately for these methods. So, for each SNP three different predictions with scores were obtained. Out of 28 SNPs, 21 nsSNPs were predicted as disease causing by all the three methods with high probability scores (Table 1). Overall ,9 nsSNPs were found to be significantly deleterious or pathogenic or disease-causing from the analysis with SIFT, PolyPhen2, CADD, PROVEAN, ClinVar, and SNPs&GO (Figure 2)

All the 28 nsSNPs were further analyzed using MutPred2.0 to find the structural/functional effect such as gain or loss of a definite structure or function in domains of a protein due to point mutation. Through this analysis, 7 SNPs were identified as causing no gain or loss of structure in protein that describes no effect due to mutation. Other 21 ns SNPs have some sort of gain or loss of structural part or a specific protein function due to a particular point mutation (Table 1). There are several different types of gain or loss of structure and function or altered function for protein TP63. Among them, loss of strand, gain of strand or intrinsic disorder, altered ordered interface, altered DNA binding, altered stability, altered metal binding, loss of phosphorylation or ribosylation are notable with p-value significance and with a high MutPred2 score (Table 1)

3.3 Domain Analysis from UniProt

Two domain such as (1) SAM domain, (2) DNA binding domain and four regions (1) Transcription activation, (2) Interaction with HIPK21, (3) Oligomerization, and (4) Transactivation inhibition regions are listed in the UniProt database (https://www.uniprot.org/uniprot/Q9H3D4) for TP63 protein. These domains and regions have specific length or regions in the protein. These regions and domains are summarized in supplementary Figure S2. Further, out of 28 nsSNPs,17

nsSNPs were selected for later analysis by following two concurrent steps – (1) prioritized the domain and regions nsSNPs to perform downstream analysis, (2) predicted as either pathogenic or deleterious using the tools (Section 4.2 and Section 4.3)

3.4 Structural analysis

I-Mutant 2.0 Analysis

Selected 17 nsSNPs of TP63 protein were subjected to protein stability analysis in I-Mutant 2.0 online server (Table 2). All 17 nsSNPs were predicted to decrease the stability in TP63 protein through I-Mutant 2.0 analysis.

HOPE Analysis

17 SNPs were subjected to run in the HOPE analysis. For each of the SNP, there is a change in size, charge, and hydrophobicity of the amino acid residue in the respective positions. Size changes to bigger or smaller, the charge is lost or gained, all these changes will be resulting in loss of interactions with other molecules or residues. Hydrophobicity is also increased resulting in loss of hydrogen bonds that will disturb the correct folding. Only a partial 3D-structure of TP63 protein was available at protein data bank (PDB) and ID is 3QYN. As only 9 SNPs were present in partial structure, HOPE created images for the 9 SNPs only, showing the wild type and mutated residues at one position with two different colors (Figure 3).

Results from HOPE analysis are presented in Table 3. In all cases, the mutations are in a domain that is important for binding with other molecules, such as residue of regulatory domain, or DNA binding domain or activity domain. These mutations will induce loss of those connections due to changes in amino acid properties. Hence, regulation of the protein, transfer of signal from binding domain to activity domain and binding of other molecules will be hampered, resulting in impeding the protein function overall.

Table 2
Effect of nsSNPs on protein stability using I-mutant 2.0

SNP	DDG	pH	Temp.	Stability
R376C	-0.97	7	25	Decrease
R408C	-1.01	7	25	Decrease
R408H	-1.38	7	25	Decrease
C347F	-0.48	7	25	Decrease
D351G	-1.64	7	25	Decrease
D355N	-1.49	7	25	Decrease
G349E	-1.48	7	25	Decrease
R266Q	-1.03	7	25	Decrease
R318H	-1.3	7	25	Decrease
R319H	-1.38	7	25	Decrease
R337Q	-0.91	7	25	Decrease
R343Q	-0.99	7	25	Decrease
R379C	-0.41	7	25	Decrease
R379C	-0.14	7	25	Decrease
L562R	-1.84	7	25	Decrease
R647H	-1.97	7	25	Decrease
R655Q	-1.96	7	25	Decrease

Evolutionary conservation analysis

From the conservation analysis, it has been identified that out of 17 SNP positions in protein, 12 residues (R266, R319, R337, R343, R376, R408, G349, D351, D355, R647, R655) are highly conserved, functional, and exposed (Figure 4A). Four residues (R318, R379) are medium conserved: two are functional and exposed, one is structural and buried, one is buried, and the remaining one residue (L562R) is not conserved (Figure 4B). This analysis showed that nsSNPs in highly conserved residues are deleterious for TP63 protein's structure and function.

Mutation 3D Analysis

Out of 17 SNPs, 11 nsSNPs (the remaining 6 nsSNPs were already identified as deleterious in several prior studies) were selected to perform 3D analysis using Mutation 3D tool where the whole TP63 protein sequence was the input sequence with 11 SNPs [36,37]. This analysis shows the three-dimensional structure of 153-388 residue length of TP63 protein with 5 nsSNPs (mutation). These SNPs are shown in the structure which is available in PDB with 100% sequence similarity and three SNPs are present in the DNA binding domain of TP63.Other mutations are in TP63 protein in two different domains: one is p53 tetramer domain and another one is SAM-2 domain (Figure S2 and S3).

3.5 Structural effect of nsSNPs

Molecular docking

Docking results showed the binding affinity of mutated TP63 protein/peptides (ligands) towards DNA. Out of 9 nsSNPs, 5 SNPs (R266Q, R319H, C347F, G349E, D351G) showed to decrease binding affinity compared to wild type peptides while interacting with DNA (Table 4). The interactions patterns of the docked DNA-Protein complexes were visualized and studied in USCF Chimera (Figure 6 & Figure 7). G349E nsSNP showed decrease in binding affinity with DNA. Wild type peptide (G349) showed binding affinity of (-) 6.4 kcal/mol which was reduced to (-) 5.8 KJ/mol due to mutation.

Swiss PDB Viewer

The predicted whole 3D structure was observed in PymolViewer. 17 nsSNPs were incorporated as point mutation using mutagenesis tool with high probability score from rotamer library. Both wild type and mutated whole 3D structures were loaded in Swiss PDB Viewer separately. Calculated energy minimization score for whole predicted structure was ~ (-) 22,000 KJ/mol for each of R266, R319, R343, G347, G349, and E351(residues showed decrease in binding affinity in molecular docking). Total energy: ~ (+)35,000KJ/mol - ~ (+)37,000 KJ/mol was found for mutated whole structure after energy minimization for same position (nsSNPs) in wild type protein (Table 5). 5 hydrogen bonds are lost due to cumulative mutations (Figure 5).

Table 3
Prediction of structural and functional consequences of nsSNPs in TP63 through Hope

SNP	Size of residue		Charge of residue		Mutant hydrophobicity	Effect on hydrophobicity	Protein folding	Loss of interactions	Interactions with domains and other molecules
	Wild type	Mutant	Wild type residue	Mutant					
R376C	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R408C	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R408H	Large	Small	Positive	Neutral	High	N/A	Affected	Yes	Disturbed
C347F	Small	Large	Neutral	Neutral	high	Yes	Affected	Yes	Disturbed
D351G	Large	Small	negative	Neutral	High	Yes	Affected	Yes	Disturbed
D355N	Large	Small	negative	Neutral	High	Yes	Affected	Yes	Disturbed
G349E	Small	Large	Neutral	Negative	High	Yes	Affected	Yes	Disturbed
R266Q	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R318H	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R319H	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R337Q	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R343Q	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R379C	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R379C	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
L562R	Small	Large	Neutral	Positive	High	Yes	Affected	Yes	Disturbed
R647H	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed
R655Q	Large	Small	Positive	Neutral	High	Yes	Affected	Yes	Disturbed

3.6 Analysis of Non-Coding SNPs

The gene variants of TP63 with transcript ID ENST00000264731.8 were retrieved from the Ensemble database (Ensemble genome browser). The SNP source was NCBI- dbSNP database. The total number of non-coding SNPs (intron, 5 prime UTR, 3 prime UTR variants) are 836 with a Global MAF range of 0.05-0.5.

RegulomeDB

836 SNPs were selected for RegulomeDB analysis. The filtering process led to 21 SNPs based on proposed ranking criteria described in RegulomeDB site (Table S3). Out of 23 SNPs, 20 were intron variants, 1 was 5 prime UTR, and the rest 2 were 3 prime UTR variants.

Finding eQTLs using GTEX Analysis

The filtered 23 SNPs from RegulomeDB analysis were further analyzed in GTEX portal to identify the single tissue eQTLs. Out of 23 noncoding SNPs, 9 SNPs (rs4488809, rs677493, rs6794898, rs79155799, rs4687090, rs1913722, rs78233713, rs73199799, rs35592567) were obtained for 13 eQTLs. In addition, Tissue specific eQTLs were observed for other genes too. The TP63 eQTLs are as follows- eQTLs: Cells - Cultured fibroblasts express PHP2 gene with SNPs: rs35592567, rs1913722, rs73199799, rs35592567, rs1913722, rs73199799. eQTL; Brain - Cerebellar Hemisphere express rs78233713 in IL1RAP gene (Table 6). Violin plots of single tissue eQTLs are presented in the Figure S4.

PolymiRTS

A total of 836 SNPs was analyzed with PolymiRTS tool. Only 44 SNPs with miRbase ID and affecting miRNA site were identified primarily. Those SNPs have ancestral allele, changed allele, conservation score which reflects the occurrence of miRNA site in other vertebrates with a significant context+ score change. These SNPs are classified into four functional groups: 1) D describes the disruption of a conserved miRNA site, 2) N describes the disruption of a non-conserved miRNA site, 3) C denotes the creation of a new miRNA site, 4) O denotes without determination of an ancestral allele (Table S4). The results show that all the miRNA target sites for miRNA predicted to be disrupted by SNPs in IFNAR2 are obtained from CLASH experiment data(N). In the next step, 44 SNPs were filtered considering the functional class of C, D, and conservation score of 19-20. Finally, only 6 SNPs were obtained. One SNP with a highly significant p-value for eQTLs has been found in Lymphoblastoid tissue. This SNP has been previously published in a study[38]. Also, several SNPs were found to be associated with different diseases and cancers and the diseases associations of those SNPs have been published in different studies (Table S5).

Table 4
Binding affinity of 9 nsSNPs using Molecular docking.

Wild type		Mutant	
Residue	Binding affinity	Residue	Binding affinity
R266	-6.2	Q266	-5.9
R318	-5.3	H318	-6.2
R319	-5.8	H319	-5.4
R337	-5.5	Q337	-5.8
R343	-5.8	Q343	-5.8
C347	-6	F347	-5.6
G349	-6.4	E349	-5.8
D351	-5.6	G351	-5.4
D355	5	N355	-5.8

4. Discussion

We performed a comprehensive characterization of the SNPs located at the TP63 gene using a combination of various In-Silico tools. We obtained the prediction of damaging/pathogenic effects for 17 nsSNPs on TP63 protein structure, stability, and function through analysis with different computational tools. In recent studies, different diseases or syndromes associated with coding SNPs or nsSNPs in different isoforms of TP63 protein have been identified and validated. These SNPs are natural variants resulting from different mutations in the identical position. Several nsSNPs (R 243 Q/W, R266Q, C208Y, S311N, R318C, R318H, R319C/H/S, R343Q/W, C345R, C347S, P348S and D351G/H) were identified and described to be associated with Ectrodactyly, ectodermal dysplasia, and cleft lip/palate syndrome 3 (EEC3) [39]. R318H is also associated with Rapp-Hodgkin syndrome (RHS) [37,39]. Split hand/foot malformation 4 (SHFM4) is associated with K232E and R219H which are also natural variants [36]. In addition, I549T, S580P are associated with RHS [36] and R352G is associated with Orofacial cleft8 (OLC8). F55A, W59A, and L62A are associated not only with the disease but also abrogates transcriptional activity and interact with transactivation inhibition domain [40]. Y543F abolishes ubiquitination [41]. S129L and S184L were reported to be associated with head and neck cancer, A187P with lung carcinoma, and Q204 with cervical cancer [42]. One nsSNP, N6H has been identified from TP63 isoform 2, 4, 6, 8, 10 and 12 and was found to be associated with the adult syndrome.

Table 5
Energy score of wild type and mutated TP63 protein

Protein (TP63)	Residue	Total Energy score (KJ/mol)	Residue Energy (Individual, KJ/mol)
Wild type	R266	-22028	-302.095
	R319	-22045	-257.839
	R343	-22065.8	-288.588
	C347	-22072.7	4.263
	G349	-22077.6	16.674
	D351	-22101	-41.772
Mutant	Q266	37571.2	-204.358
	H319	37504.11	-7.031
	Q343	37182.07	-195.391
	F347	37147.81	8.314
	E349	35738.79	-7.963
	G351	35720.33	11.934

It has been observed from the study that all nsSNPs did not have a significant deleterious impact on protein activity- many SNPs are either deleterious or pathogenic, some are neutral or show insignificant influence. It is critical to (1) to evaluate the vulnerability of particular SNPs to pathogens and (2) to emphasize on SNPs which are accountable for systemic and functional effects on proteins i.e.TP63. Moreover, the use of a single bioinformatic tool for predicting the pathogenic effect of nsSNPs is not trustworthy or reliable as screening with single tool may result in unsatisfactory outcome with low confidence [43]. Therefore, we aimed to utilize various In-Silico tools to evaluate the missense and noncoding SNPs.

In this study, initially 403,549 SNPs of the humanTP63 gene have been identified from the dbSNP (archive) database of NCBI in non-coding, coding, and regulatory categories including all isoforms. SNPs of the coding region induce variation in amino acids resulting in protein structure alteration and contributing to disease susceptibility. Therefore, we have decided to study missense SNPs or nsSNPs using several bioinformatics tools. Prior studies have suggested that the best reported performing tools are SIFT, PolyPhen2, and CADD [44,45]

In our study total 17 nsSNPs were identified and out of 17, 9 SNPs were found to be significantly deleterious or pathogenic or disease-causing from the analysis with SIFT, PolyPhen2, CADD, PROVEAN, ClinVar, MutPred2 and SNPs&GO. In addition, these 9 nsSNPs are present in the DNA binding domain of TP63. Furthermore, due to the presence of these SNPs, there is a decrease in protein stability predicted by I-mutant 2.0. HOPE analysis reports show the changes in amino acid properties (i.e., size, charge, and hydrophobicity), and the loss of interactions with DNA binding domain, protein regulatory domain, or other molecules in other domains. These activities are predicted to alter the function of protein TP63.

Table 6
eQTLs prediction of non-coding SNPs in GTEx portal

Gene Symbol	Variant ID	SNP ID (noncoding)	P-Value	NES	Single Tissue eQTL
TP63	chr3_189638472_T_C_b38	rs4488809	6.50E-07	0.23	Lung
TP63	chr3_189638472_T_C_b38	rs4488809	0.000017	0.13	Nerve - Tibial
TP63	chr3_189664468_A_G_b38	rs6774934	0.000055	-0.34	Heart - Left Ventricle
TP63	chr3_189664468_A_G_b38	rs6774934	0.000065	-0.2	Nerve - Tibial
TP63	chr3_189672911_G_C_b38	rs6794898	0.000013	0.2	Lung
TP63	chr3_189710792_T_G_b38	rs79155799	1.8E-07	-0.17	Nerve - Tibial
TP63	chr3_189721190_A_G_b38	rs4687090	7.10E-11	-0.19	Nerve - Tibial

The remaining 8 SNPs (R408C.R408H, R376C, R379C/H, L562R, R547H, and R655Q) were chosen for additional analysis due their presence in domains which are functionally important for different activities i.e., such as transactivation inhibition, oligomerization, and interaction with HIPK21. HIPK21 phosphorylates SIRT1 which is involved in several biological processes i.e., cancer management, cell death control, cellular stress reaction, metabolic control [46,47]. Mutation 3D analysis reveals three domains of TP63: (1) DNA binding domain, (2) SAM domain, and (3) p53 _tetramer_sf (Figure S3). PROVEAN and PhD-SNP predicted R379C/H, L562R, R547H, and R655Q as neutral, while ClinVar predicted these 8 SNPs to either have uncertain significance or not reported in their database. SIFT, PolyPhen2 and CADD evaluated all these 8 SNPs as deleterious at the beginning. I-mutant predicted these 8 SNPs to cause decrease in protean stability and SNPs&GO predicted them to be disease-causing. In the HOPE report, these SNPs are described as changing in amino acid properties and loss of interactions with other molecules and functional regions, thus subsequently interrupting the protein function.

nsSNPs that have already been found to be associated with different inherited disorders and diseases in several studies have been described and annotated in the UniProt database. From our analysis and selection of SNPs, 6 SNPs (R266Q, R318H, R319H, R337Q, R343Q, D351G) have been previously reported to be associated with disorders and known to the scientific community. Thus, 11 yet unreported nsSNPs with no known disease association have been detected in this study. These nsSNPs are R376C, R408C, R408H, C347F, D355N, G349E, R379C,

R379H, L562R, R647H, R655Q. Three-dimensional analysis of TP63 protein with mutation 3D tools for all 11 SNPs shows the location of these mutations in protein (Figure S2 and S3).

We predicted the three-dimensional structure of the whole TP63 protein using I-TASSER. Next, we generated the mutated protein structure by incorporating 17-point mutations based on the identified 17 nsSNPs in PyMol viewer. Both wild-type and mutated whole structures were analyzed in Swiss PDB Viewer. We found that total energy minimization score for selected each of five residues (R266, R343, C347, G349, and D351) was very low ~(-) 22,000 KJ/mol in wild type protein. This score correlates with any native protein since stable protein has very low energy. However, the total energy minimization score for each of five nsSNPs (R266, R343, C347, G349, and D351) was very high; ~(+)37,000 KJ/mol in mutated structure (Table 5). Figure 5 also shows loss of hydrogen bonds due to mutations(cumulative).

Molecular docking analysis of DNA-protein was also performed in this study using PyRx and HADDOCK web server that showed the effect of nsSNPs in binding interactions between DNA and TP63 protein. Binding affinity was reduced for 5 mutated peptides (R266Q, R319H, C347F, G349E, D351G) generated based on nsSNPs in the DNA binding domain (Table 4). Analysis of docked complexes of TP63 peptides and DNA (wildtype and mutated) in Chimera 1.14 showed the loss of hydrogen bonds for R319H and G349E (Figure 6 & Figure 7). Therefore, the above structural effects of nsSNPs on energy score, hydrogen bond loss, and binding affinity reductions established the pathogenicity for nsSNPs.

PolymiRTS analysis of all non-coding SNPs resulted in 44 SNPs with the sequence of the miRNA binding site (Table S4). 6 SNPs (based on conservation score) having miRNA binding site were also retrieved. These SNPs can alter miRNA binding sites where miRNA will bind to control the function of TP63. Our study also found that TP63 variants have been identified in various types of cancer including bladder cancer (Table S5)

Our study of noncoding SNPs of the gene variants through PolymiRTS reveals several disease associations with SNPs which have been published in a handful of studies. The SNPs with lowest P-value are rs4488809, rs10937405, rs17505102, rs710521 (Table S5). rs4488809 identified in the GWAS analysis confirms new lung cancer susceptibility loci in never-smoking women in Asia [48] and two different loci at 13q12.12 and 22q12.2 in the Han Chinese population [49]. Rs10937405 was identified for lung adenocarcinoma in the Japanese population and correlated with lung adenocarcinoma susceptibility in Japanese and Korean populations [4,46,47,50]. rs17505102, another intron region SNP of TP63 protein, was found to be associated with childhood acute lymphoblastic leukemia [51]. The intron SNP rs710521 was identified as susceptible at multiple loci and was found to be associated with urinary bladder cancer from different GWA studies. [52–54].Two common bladder cancer associated variants in Europe are rs9642880 in 8q24.21 and rs710521 in 3q28. [55].

Our study with non-coding SNP analysis also identified 23 SNPs (20 intron, one 5' prime and two 3 prime UTR) though RegulomeDB analysis (Table S3). Studying gene expression and gene regulation on the tissue level of an organ is an important approach to interpret the variant's effect on a particular tissue type. These 23 SNPs was analyzed in the Genotype-Tissue Expression (GTEx) portal and identified 13 single tissue eQTLs. Only 5 SNPs (rs4488809, rs6774934, rs6794898, rs79155799, and rs4687090) in TP63 gene were found to have single tissue eQTLs (Table 6) with various tissue types: lung, heart, skin, nerve cells and cultured fibroblasts. The violin plots of eQTLs showed normalized expression of TP63 gene with a specific SNP. Figure S4 shows the proof of the mutation in the tissue and indicates the alteration of the regulation of the geneTP63 affecting regulation of protein products.

The study has limitations too. First, the study was based on publicly available data. Hence, achieving more clinical information for the SNPs is challenging. Second, we found a small number of causative nsSNPs, and laboratory experiments were not performed, therefore the effect of nsSNPs on stability, structure, and function of TP63s, was hardly determined. However, this study will work as benchmark for further validation of the TP63 SNPs for their disease association.

5. Conclusion And Future Directions

In this study, we harnessed a combination of in silico tools to identify a total of 17 nsSNPs which can alter the structure and function of TP63 protein. The strength of our In silico prediction of deleterious or functionally important SNPs were further validated by the fact that 6 out of these 17 SNPs have also been shown to be associated with different diseases in prior studies. Hence, the remaining 11 previously unreported SNPs which are predicted to be pathogenic using various computational tools are crucial due to their impact on the structure and function of TP63. Energy minimization, visualization, and molecular docking analysis supported us to verify the deleteriousness of these SNPs in TP63 protein. The non-coding SNP analysis reveals miRNA binding site disturbance which may dysregulate the expression of TP63 protein. GTEx analysis identified the single tissue eQTLs showing the expression of TP63 with non-coding SNPs in several tissue types: lung, nerve, heart, skin and brain. To confirm the association of these 11 nsSNPs, 6 noncoding PolymiRTS analyzed SNPs and 5 noncoding GTEx analyzed SNPs with different diseases or cancers, experimental analysis needs to be carried out in future i.e., performing the functional analysis in cell-lines.

Declarations

Contributions

H.U.S and I.H revised the manuscript and supervised the whole work. All authors read and agreed to the submitted version of the manuscript.

Funding:

Not applicable.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Conflicts of Interest:

The authors declare no conflict of interest.

References

1. Monti, P. *et al.* P63 modulates the expression of the WDFY2 gene which is implicated in cancer regulation and limb development. *Bioscience Reports* **39**, 1–15 (2019).
2. Ghioni, P. *et al.* Complex Transcriptional Effects of p63 Isoforms: Identification of Novel Activation and Repression Domains†. *Molecular and Cellular Biology* **22**, 8659–8668 (2002).
3. Kiemeney, L. A. *et al.* A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nature Genetics* **42**, 415–419 (2010).
4. Miki, D. *et al.* Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nature Genetics* **42**, 893–896 (2010).
5. Gassoum, A., Abdelraheem, N. E. & Elsadig, N. Comprehensive Analysis of rsSNPs Associated with Hypertension Using In-Silico Bioinformatics Tools. *OALib* **03**, 1–24 (2016).
6. Hossain, S. & Hosen, M. I. Dissecting the structural and functional impact of SNPs located in the spinal muscular atrophy associated gene SMN1 using in silico analysis. *Gene Reports* **15**, 100388 (2019).
7. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes [published erratum appears in Nat Genet 1999 Nov;23(3):373]. *Nature Genetics* **22**, 231–238 (1999).
8. Czarny, P. *et al.* Single-nucleotide polymorphisms of uracil-processing genes affect the occurrence and the onset of recurrent depressive disorder. *PeerJ* **2018**, (2018).
9. Desai, M. & Chauhan, J. B. Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. *Computational Biology and Chemistry* **70**, 7–14 (2017).
10. Divanshu, G., Lekshmi, M. & Shanthi, V. In silico studies of deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) of NRL gene. *Network Modeling Analysis in Health Informatics and Bioinformatics* **3**, 1–7 (2014).
11. Nimir, M. *et al.* In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene. *F1000Research* **6**, 243 (2017).
12. Zhang, M., Huang, C., Wang, Z., Lv, H. & Li, X. In silico analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the human GJA3 gene associated with congenital cataract. *BMC Molecular and Cell Biology* **21**, 1–13 (2020).
13. Badgujar, N. V., Tarapara, B. V. & Shah, F. D. Computational analysis of high-risk SNPs in human CHK2 gene responsible for hereditary breast cancer: A functional and structural impact. *PLoS ONE* **14**, 1–18 (2019).
14. Doss, C. G. P., Chakraborty, C., Rajith, B. & Nagasundaram, N. In silico discrimination of nsSNPs in hTERT gene by means of local DNA sequence context and regularity. *Journal of Molecular Modeling* **19**, 3517–3527 (2013).
15. Hossain, M. S., Roy, A. S. & Islam, M. S. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. *Scientific Reports* **10**, 1–14 (2020).
16. Vanbokhoven, H., Melino, G., Candi, E. & Declercq, W. P63, a story of mice and men. *Journal of Investigative Dermatology* **131**, 1196–1207 (2011).
17. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308–311 (2001).
18. Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research* **40**, W452–W457 (2012).
19. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248–249 (2010).
20. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research* **47**, D886–D894 (2019).
21. Capriotti, E., Fariselli, P., Calabrese, R. & Casadio, R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics* **21**, ii54–ii58 (2005).

22. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).

23. Landrum, M. J. *et al.* ClinVar: Improvements to accessing data. *Nucleic Acids Research* **48**, D835–D844 (2020).

24. Pejaver, V. *et al.* Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature Communications* **11**, (2020).

25. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research* **33**, W306–W310 (2005).

26. Venselaar, H., te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548 (2010).

27. Meyer, M. J. *et al.* mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome. *Human Mutation* **37**, 447–456 (2016).

28. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research* **38**, W529–W533 (2010).

29. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. in *Chemical biology* 243–250 (Springer, 2015).

30. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605–1612 (2004).

31. Vina, A. Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading Trott, Oleg; Olson, Arthur J. *J. Comput. Chem* **31**, 455–461 (2010).

32. De Vries, S. J. *et al.* HADDOCK versus HADDOCK: new features and performance of HADDOCK2. 0 on the CAPRI targets. *Proteins: structure, function, and bioinformatics* **69**, 726–733 (2007).

33. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790–1797 (2012).

34. Bhattacharya, A., Ziebarth, J. D. & Cui, Y. PolymiRTS Database 3.0: Linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Research* **42**, D86–D91 (2014).

35. Lonsdale, J. *et al.* The genotype-tissue expression (GTEx) project. *Nature genetics* **45**, 580–585 (2013).

36. Ianakiev, P. *et al.* Split-hand/split-foot malformation is caused by mutations in the p63 gene on 3q27. *American Journal of Human Genetics* **67**, 59–66 (2000).

37. Bougeard, G., Hadj-Rabia, S., Faivre, L., Sarafan-Vasseur, N. & Frébourg, T. The Rapp-Hodgkin syndrome results from mutations of the TP63 gene. *European Journal of Human Genetics* **11**, 700–704 (2003).

38. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).

39. Van Bokhoven, H. *et al.* p63 gene mutations in EEC syndrome, limb-mammary syndrome, and isolated split hand-split foot malformation suggest a genotype-phenotype correlation. *American Journal of Human Genetics* **69**, 481–492 (2001).

40. Serber, Z. *et al.* A C-Terminal Inhibitory Domain Controls the Activity of p63 by an Intramolecular Mechanism. *Molecular and Cellular Biology* **22**, 8601–8611 (2002).

41. Li, Y., Zhou, Z. & Chen, C. WW domain-containing E3 ubiquitin protein ligase 1 targets p63 transcription factor for ubiquitin-mediated proteasomal degradation and regulates apoptosis. *Cell Death and Differentiation* **15**, 1941–1951 (2008).

42. Borremans, B., Hobman, J. L., Provoost, a & Brown, N. L. Cloning and Functional Analysis of the. *Society* **183**, 5651–5658 (2001).

43. Snps, P. & Homolog, M. *Computational Analysis of Deleterious Single Nucleotide*. vol. 6 (2016).

44. Hicks, S., Wheeler, D. A., Plon, S. E. & Kimmel, M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Human Mutation* **32**, 661–668 (2011).

45. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation* **32**, 358–368 (2011).

46. Conrad, E. *et al.* HIPK2 restricts SIRT1 activity upon severe DNA damage by a phosphorylation-controlled mechanism. *Cell Death and Differentiation* **23**, 110–122 (2016).

47. Bordone, L. & Guarente, L. Calorie restriction, SIRT1 and metabolism: Understanding longevity. *Nature Reviews Molecular Cell Biology* **6**, 298–305 (2005).

48. Manuscript, A. & Structures, T. 112Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia112. **6**, 247–253 (2009).

49. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nature Genetics* **43**, 792–796 (2011).

50. Shiraishi, K. *et al.* A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nature Genetics* **44**, 900–903 (2012).

51. Ellinghaus, E. *et al.* Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* **26**, 902–909 (2012).

52. Kiemeney, L. A. *et al.* A sequence variant at 4p16. 3 confers susceptibility to urinary bladder cancer. *Nature genetics* **42**, 415–419 (2010).

53. Kiemeney, L. A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature genetics* **40**, 1307–1312 (2008).

54. Quispe-Tintaya, W. ▢▢▢▢▢▢ HHS Public Access. *Physiology & behavior* **176**, 139–148 (2017).

55. Kiemeney, L. A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genetics* **40**, 1307–1312 (2008).

Figures

Figure 1
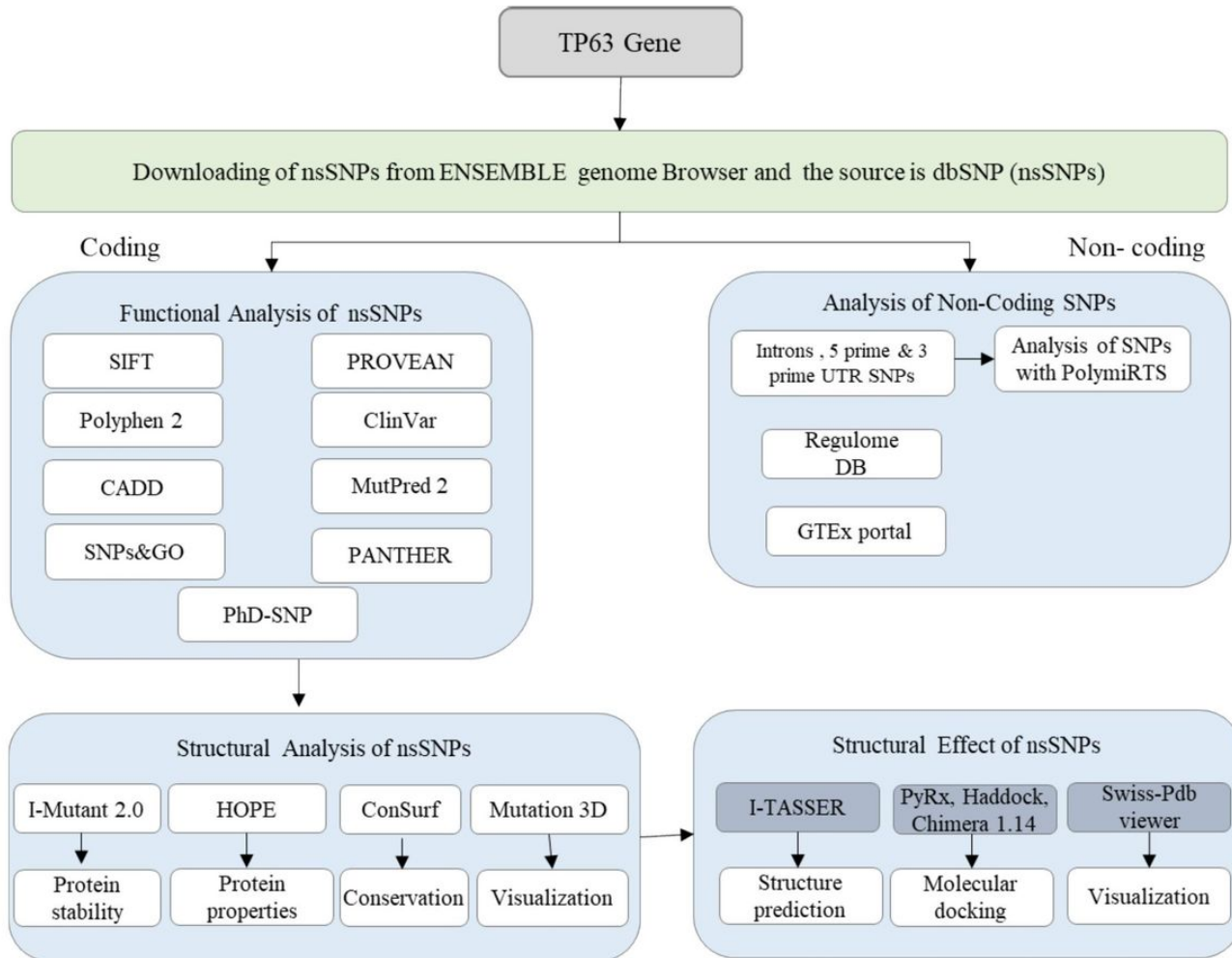
Schematic representation of computational tools for In Silico analysis of SNPs in TP63 protein.
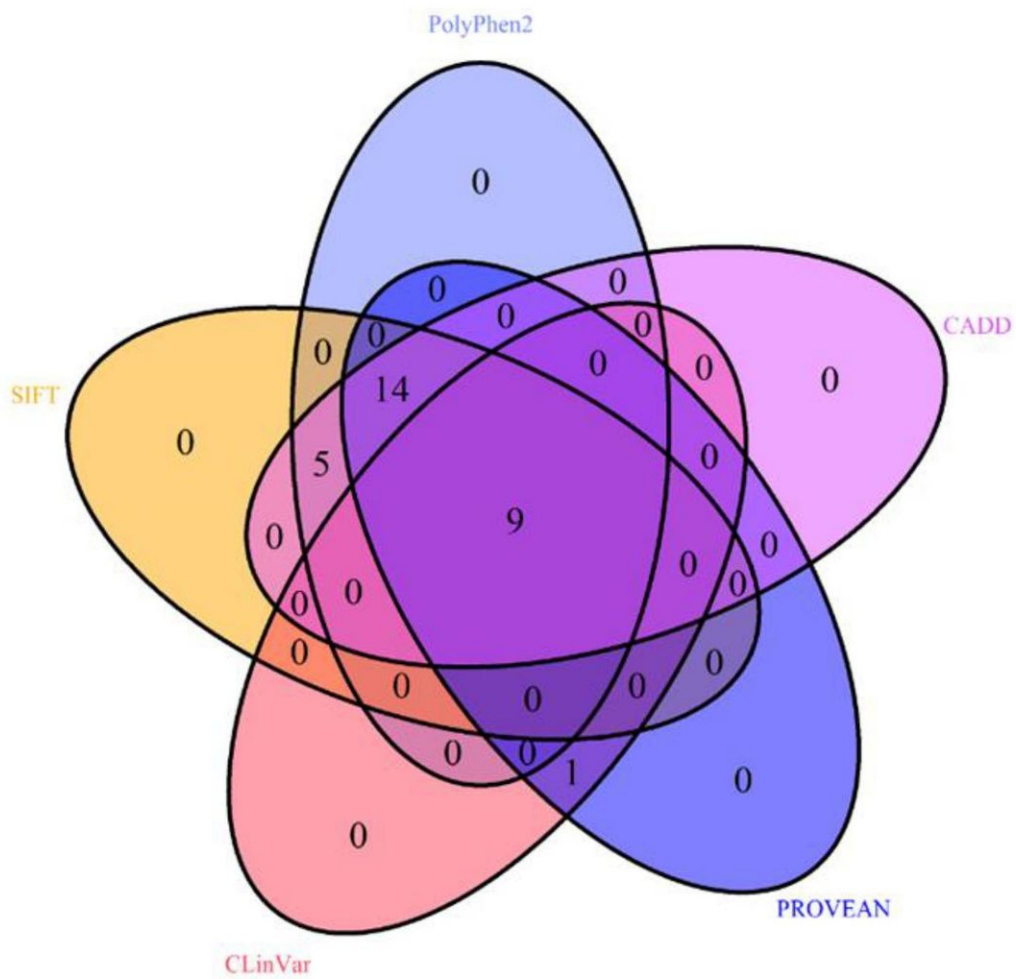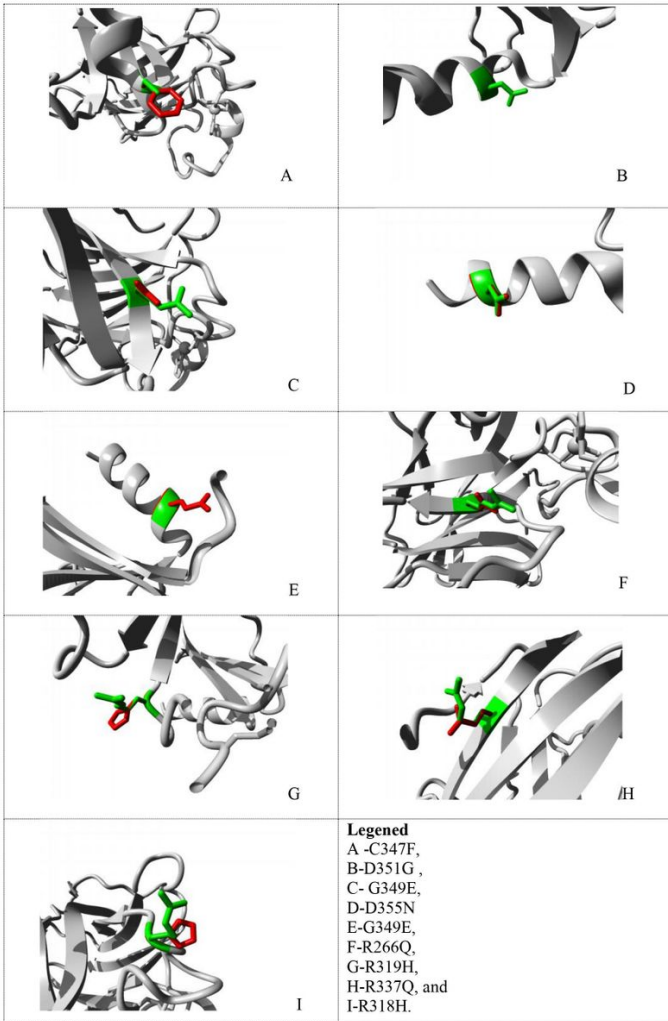
Figure 2

Venn diagram showing prediction of 9 nsSNP as deleterious by five bioinformatics tools (SIFT, Polyphen 2, CADD, PROVEAN, CLinVar)

Figure 3

Visualization of the mutation in structure of TP63 protein. The wild type of residue is labelled as green and mutated residue is labelled as red. Here, only side chain of the residue is shown in color, the whole protein is shown as grey.
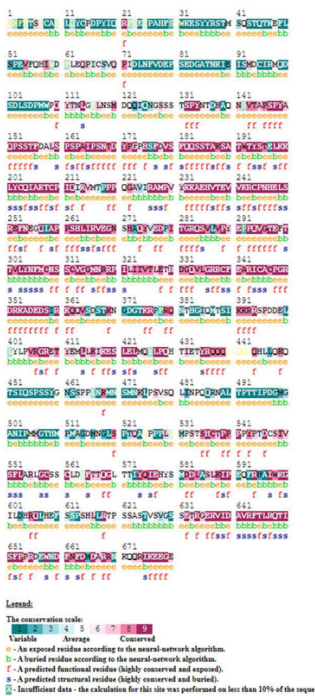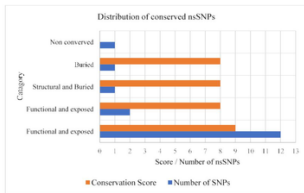
Figure 4

A:Conservation analysis of nsSNPs in Consurf B: Distribution of high-risk nsSNPs of TP63 as functional and exposed, highly or medium conserved or not conserved, and buried
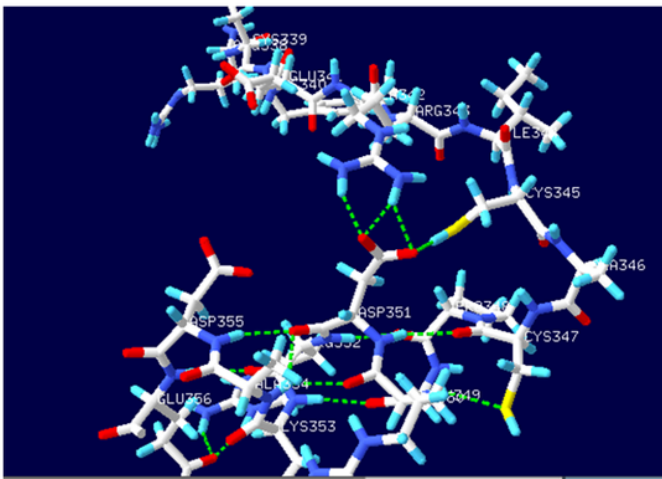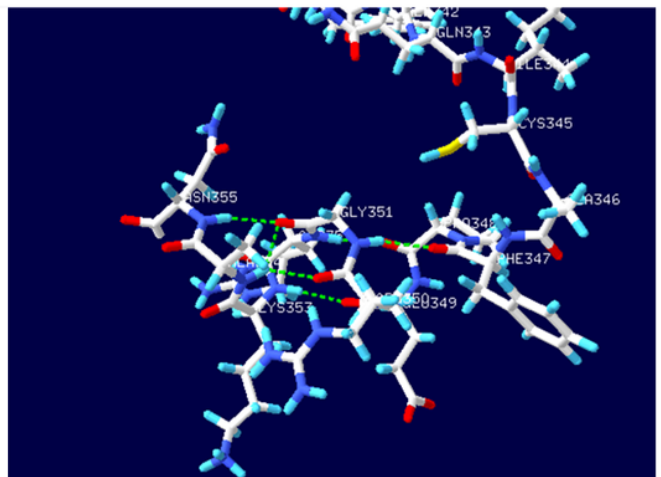
Figure 5

Structural effect of R343Q,C347F and G349E and D351G on A) wild type and B)mutated whole TP63 protein. A) R343 forms three hydrogen bonds with E351 denoted by green dashed lines and there are 11 H-bonds among residues from 343-355. B) Q343 does not form any H-bond with any neighboring residue as

Asp at 351 mutated to G(glycine) and there are only six H-bonds among residues from 343-355.Due to mutations in 5 residues in this region of protein (DNA binding domain),5 hydrogen bonds are lost impacting greatly in structure and function of TP63

Figure 6

Ligand binding effect of nsSNPs through molecular docking for G349E. a) Wild type or native protein binding with DNA reveals that Arg8 in peptide has three H-bonds with DNA backbone and DNA base, Lys9 has two H-bonds with DNA base and sugar, Arg6 forms one H-bond with DNA base, Cys1 forms one hydrogen bond b) Mutated residue Glu (E) in position 5 in peptide or 349 in TP63.Here, it is observed that Arg 8 does not form any hydrogen bond with DNA and Lys9 forms only one H-bond. This corelates with binding affinity reduction in docking

Figure 7

Ligand binding effect of nsSNPs through molecular docking for R319H. a) Wild type or native protein binding with DNA reveals that Arg4 in peptide (318 in protein) has four H-bonds with DNA base (Adenine), Arg5(319 in protein) has two H-bonds with DNA base and sugar, GLY6 forms one H-bond with DNA base b) Mutated residue His(H) is in position 5 in peptide or 319 in TP63.Here, it is observed that Arg 4 forms one hydrogen bond with DNA sugar and HIS9 forms only one H-bond. This finding co-relates with binding affinity reduction in docking result as there is a significant number of hydrogen bonds lose due to mutation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplimentaryTablesandFiguresShamimaShafaatIsmailHossain.docx