Bulletin of the National
Research Centre

RESEARCH

Computational analysis uncovers the deleterious SNPs along with the mutational spectrum of p53 gene and its differential expression pattern in pan-cancer

Saruar Alam[1†] , Mohammad Sayem[2†] , Martiniano Bello[3] , Sadia Islam Mou[1] , Nairita Ahsan Faruqui[4] ,
Faruk Hossain[5] and Md. Kamrul Hasan[6,7*]

Abstract

Background: A variety of accessible data, including those of single-nucleotide polymorphisms (SNPs) on the human p53 gene, are made widely available on a global scale. Owing to this, our investigation aimed to deal with the detrimental SNPs in the p53 gene by executing various valid computational tools, including—Filter, SIFT, PredictSNP, Fathmm, UTRScan, ConSurf, SWISS-MODEL, Amber 16 package, Tm-Adjust, I-Mutant, Task Seek, GEPIA2 after practical and basic appraisal, dissolvable openness, atomic progression, analyzing the energy minimization and assessing the gene expression pattern.

Results: Out of the total 581 p53 SNPs, 420 SNPs were found to be missense or non-synonymous, 435 SNPs were in the three prime UTR, and 112 SNPs were in the five prime UTR from which 16 non-synonymous SNPs (nsSNPs) were predicted to be non-tolerable while PredictSNP package predicted 14. Concentrating on six bioinformatics tools of various dimensions, a combined output was generated, where 14 nsSNPs could exert a deleterious effect. We found 5 missense SNPs in the DNA binding domain's three crucial amino acid positions, using diverse SNP analyzing tools. The underlying discoveries were fortified by microsecond molecular dynamics (MD) simulations, TM-align, I-Mutant, and Project HOPE. The ExPASy-PROSITE tools characterized whether the mutations were located in the functional part of the protein or not. This study provides a decisive outcome, concluding the accessible SNPs' information by recognizing the five unfavorable nsSNPs—rs28934573 (S241F), rs11540652 (R248Q), rs121913342 (R248W), rs121913343 (R273C), and rs28934576 (R273H). By utilizing Heatmapper and GEPIA2, several visualization plots, including heat maps, box plots, and survival plots, were produced.

Conclusions: These plots disclosed differential expression patterns of the p53 gene in humans. The investigation focused on recognizing the detrimental nsSNPs, which augmented the danger posed by various oncogenesis in patients of different populations, including within the genome-wide studies (GWS).

†Contributed equally as first author: Saruar Alam and Mohammad Sayem

*Correspondence: kamrul.hasan11@northsouth.edu;
kamrulhasanhridoy205@gmail.com

[6] Department of Biochemistry and Molecular Biology, Tejgaon College,
National University of Bangladesh, Gazipur 1704, Bangladesh
Full list of author information is available at the end of the article

Background

Single Nucleotide Polymorphism (SNP) is marked as the most prevalent form of genetic mutation in humans. About 93% of the human genes include at least a single SNP (Chakravarti 2001). SNPs contribute to most of the variations among individuals, making each person unique. SNPs can be in the coding, non-coding, and intergenic regions between two genes (Carninci et al. 2005; Liu et al. 2006). Although non-coding SNPs are phenotypically neutral, nsSNPs can influence phenotype by altering protein sequences (Chakravarti 2001; Carninci et al. 2005; Liu et al. 2006; Ng and Henikoff 2006). Moreover, nsSNPs alter the amino acids in their corresponding protein, which could have a deleterious effect on the structure and function (Dryja et al. 1990; Smith et al. 1994; Singh et al. 2020). They are associated with various human diseases and disorders. Several studies confirm the association of nsSNPs with their susceptibility to infection and the progression of autoimmune diseases and inflammatory disorders (Dryja et al. 1990; Smith et al. 1994; Singh et al. 2020; Barroso et al. 1999; Chasman and Adams 2001; Lander 1996). About 50% of the mutations implicated with hereditary genetic disorders are nsSNPs (Kelly and Barr 2014; Radivojac et al. 2010). As a result, many researchers focus on nsSNPs in cancer biology, precisely, cancer-causing genes.

Mutations in the tumor suppressor gene, p53, account for ~ 50% of human cancers (Doniger et al. 2008; Finlay et al. 1989; Baker et al. 1990; Hamzehloie et al. 2012; Zhang et al. 2020). p53 is a critical regulator of tissue homeostasis (Baugh et al. 2018; Diller et al. 1990; Chng et al. 2007), which further binds to stabilize DNA as a tetramer, leading to the regulation of genes. Consequently, this helps mediate critical cellular processes, including cell-cycle arrest, DNA repair, senescence, and apoptosis (Kastenhuber and Lowe 2017; Riley et al. 2008). The regular allele of p53 encodes a 53-kD nuclear phosphoprotein that plays an important role in controlling cell proliferation (Eliyahu et al. 1989; Ahuja et al. 1989; Takahashi et al. 1989; Bressac et al. 1990; Matsuda et al. 2005). However, in human tumors, point mutations, rearrangements, allelic loss, and deletions are found in the p53 gene (Hamosh et al. 2004; Sherry et al. 2001; UniProt Consortium 2007). Together with the changes in oncogenes and tumor suppressor genes, these abnormalities consist of a network of mutations that leads to malignancy. Despite the importance of p53, no computational studies have been reported that detect the deleterious nsSNPs in the p53 gene. Therefore, we carried out an in silico analysis of the p53 gene to characterize the deleterious mutations in this current investigation. Our study encompasses- (1) retrieving SNPs in the p53 gene from available databases, (2) allocating deleterious nsSNPs to their phenotypic effects based on the sequence and structure-based homology, and identifying the regulatory nsSNPs responsible for altering the patterns of splicing and gene expression, (3) predicting the role of the substitution of the amino acid on the secondary structures based on solvent stability and accessibility, and (4) predicting the effect of mutations in the domain structures.
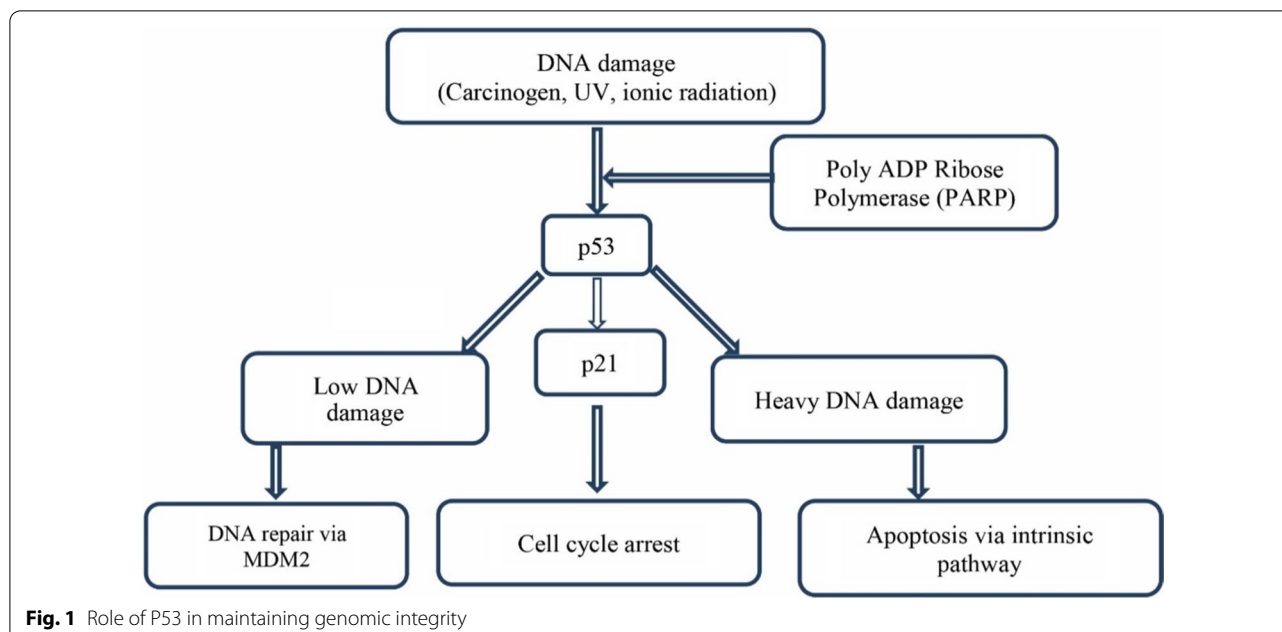
Methods

The flowchart expounds on our study's process of identifying and characterizing detrimental SNPs in the p53 gene. The structural and functional consequences have been analyzed upon missense mutation (Fig. 1). The workflow is given in Fig. 2.

Retrieval of SNP datasets

The human p53 gene was retrieved from web-based data sources, such as the Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/omim/). The SNPs' information, both protein accession number and SNP ID of the p53 gene, was retrieved from the NCBI dbSNP (Database of Single Nucleotide Polymorphism) (Hamosh et al. 2004), and the protein FASTA sequence was retrieved from UniProt (http://www.uniprot.org/) (Sherry et al. 2001).

Analysis of functional consequences of nsSNPs

An online tool known as Sorting Intolerant From Tolerant (SIFT) was employed to identify the deleterious nonsynonymous SNPs of the p53 gene (UniProt Consortium 2007). This program assumes that primary amino acids will be conserved in the protein family and that changes at particular positions are potentially harmful (Ng and Henikoff 2003; fathmm - Analyze Cancer-Associated Variants 2018). During mutagenesis studies in humans, SIFT can easily differentiate between functionally neutral or innocuous and detrimental polymorphisms (UniProt Consortium 2007). The SIFT program's algorithm uses SWISSPROT, nr, and TrEMBL databases to find homologous sequences with a query. The rsIDs (identification number) of each SNP of the human p53 gene obtained from NCBI were submitted to SIFT as a query for homology searching. The SIFT score ≤ 0.05 was set to indicate

Alam *et al. Bulletin of the National Research Centre* (2022) 46:191

Page 3 of 17

Fig. 1 Role of P53 in maintaining genomic integrity

the deleterious effect of a non-synonymous mutation on protein function.

Characterization of functional nsSNPs

For characterization of functional nsSNPs, we used PredictSNP web server (Ashkenazy et al. 2010). It was constructed from three independent datasets by eliminating all inconsistencies, duplicities, and mutations before assessment. The standard dataset comprising over 43,000 mutations was taken for the impartial assessment of eight well-known prediction tools: nsSNPAnalyzer, MAPP, PANTHER, PolyPhen-1, PhD-SNP, PolyPhen-2, SIFT, and SNAP. The six best-performing tools were shared into an accord classifier PredictSNP, resulting in drastically better prediction implementation and simultaneous time returned results for all mutations, therefore corroborating that the unanimity prediction denotes an accurate and vigorous alternative compared to the predictions delivered by individual tools (Ashkenazy et al. 2010).

Prediction of cancer-promoting mutations

Some mutations may have an association with cancer. To predict the cancer-associated SNPs of the p53 gene, we used the Functional Analysis through Hidden Markov Models (Fathmm) webserver (http://fathmm. biocompute.org.uk/cancer.html). Fathmm allows combining sequence conservation within hidden Markov models (HMMs). Fathmm server is a high-throughput web server often employed to identify phenotypic, molecular, and functional consequences of protein

variants on coding and non-coding regions. Fathmm employs unweighted, sequence/conservation-based, and weighted algorithms combined with sequence conservation with pathogenicity weights. In the Fathmm server, the default prediction threshold is set at -0.75, where a prediction with a score less than this value predicts that the mutation is considered to be potentially associated with cancer. Cancer-promoting mutations are detrimental to our bodies. These types of mutations play a critical role in cell-cycle regulation, and mutations falling in the conserved region can also depress the nature of the domain.

Identification of functional SNPs in conserved regions

Functional amino acids remain conserved throughout evolution. Evolutionarily conserved amino acid residues in the p53 protein were identified by the ConSurf web server (http://consurf.tau.ac.il/2016/index_proteins.php) by using a Bayesian algorithm (conservation scores: 1–4 variable, 5–6 intermediate, and 7–9 conserved) (Bendl et al. 2014; Pesole et al. 2002). Protein FASTA sequence was submitted, and the conserved regions were predicted, shown by means of coloring scheme and conservation score of the amino acids. It also predicts the functional and structural residues of the protein. For further analysis, highly conserved amino acids at high-risk nsSNP sites were selected.
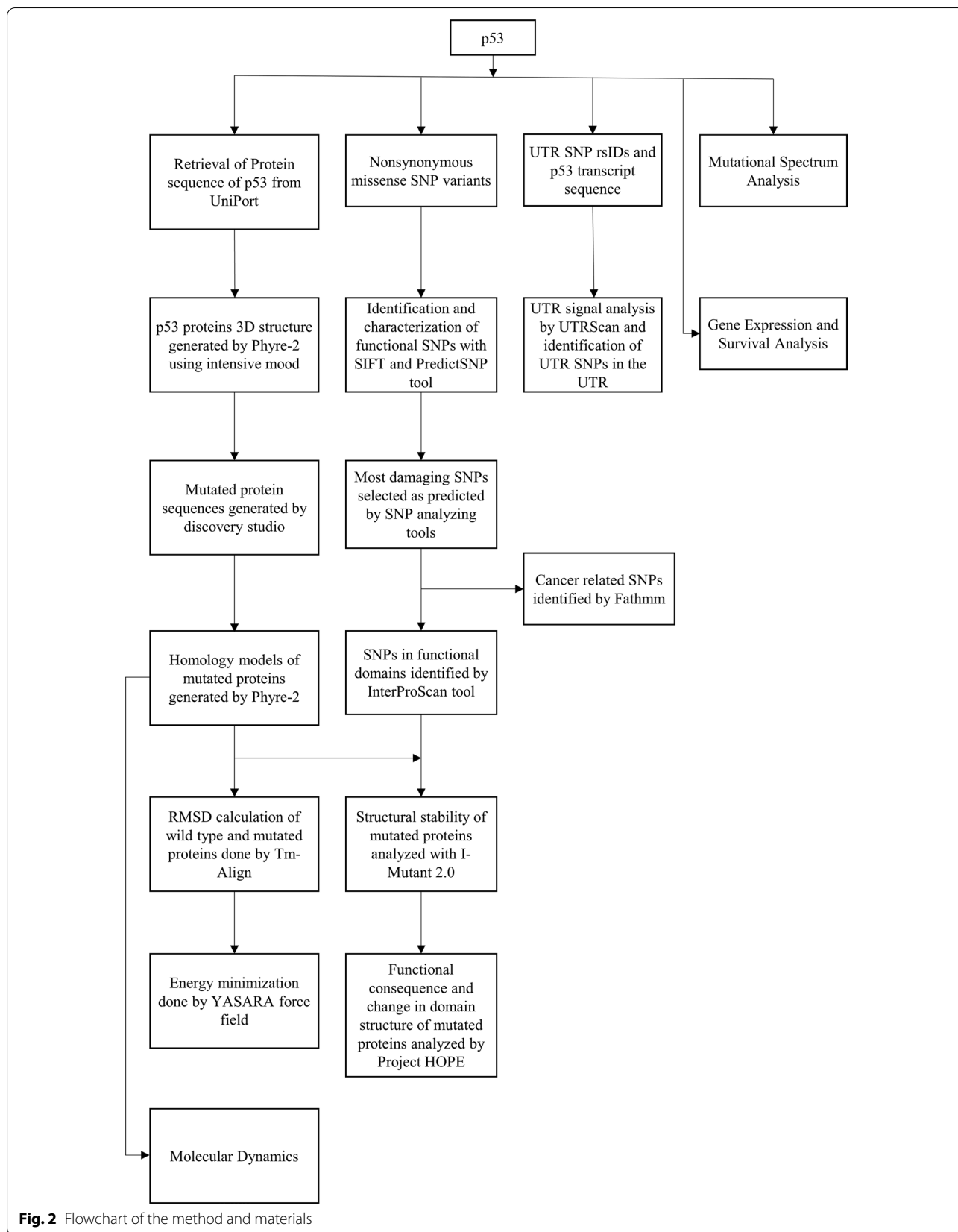
Fig. 2 Flowchart of the method and materials

Scanning of UTR SNPs

Untranslated regions (UTRs) play vital roles in the post-transcriptional instruction and regulation of gene expression, which comprise the modulation of the transport of mRNAs out of the nucleus and translation competence, subcellular localization, and constancy (EMBL-EBI 2018). To find the functional SNPs, we employed UTRScan, a web server (Pesole et al. 2001) for alignment matching. UTRScan searches nucleotide (RNA, tRNA, DNA) or protein sequences to find UTR motifs and locate motifs that distinguish 3′UTR and 5′UTR sequences (in specific sequences). The UTRSite Database defines such motifs as a compilation of functional sequence arrangements in the 5′- or 3′-UTR sequences (Grillo et al. 2010; Zhang et al. 2013). If an SNP with a different nucleotide at each UTR is found to have dissimilar working patterns, this UTR SNP is expected to impact the mRNA stability. To perform this, 5′- and 3′-UTR SNPs from NCBI were submitted in FASTA format, and the results showed predicted UTRs at the specific region.

Identification of a deleterious mutation in the functional domain

The functional domain of the product of the p53 gene was identified using InterProScan (http://www.ebi.ac.uk/InterProScan/). InterProScan connects diverse protein signature identification methods from the Inter-Pro consortium associate databases into one resource. A web-based version of InterPro is accessible for academic and profitable organizations from the EBI. The InterProScan tool allows scanning protein sequences received in FASTA format for matches against the InterPro protein signature databases. After analyzing the deleterious mutation from the SIFT mutation among them, it was identified in the functional domain of the p53 protein.

Modeling of the mutated protein

The three-dimensional structure (3D) of p53 was obtained from the Protein Data Bank (PDB entry 6XRE), and the missing region was constructed using the SWISS-MODEL server (Waterhouse et al. 2018). The p53 model was submitted to molecular dynamics (MD) simulations through 500 ns (ns). The most populated conformer through MD simulations was employed to construct mutants, and the mutations were implemented using PyMOL (Discovery Studio 2018). MD simulations were carried out using the Amber 16 package (DeLano 2002) and the ff14SB force field (Case et al. 2005). Systems were inserted in a dodecahedral box of 1 nm between the protein and the edge of the box using the TIP3P water model and

neutralized utilizing Cl^- and Na^+ counter-ions. The solvated protein was submitted through energetic minimization using the steepest descent method through 1000 steps. Following this, the system was equilibrated by 1 ns at 300 K, where the solvent was kept to desist, but the protein atoms were restrained. MD simulations of 500 ns were performed for the wild-type protein and 100 ns for all the mutants created from the p53 protein. MD simulations were run considering an NPT ensemble, and the time step for the MD simulations was 2 femtoseconds (fs). The temperature was set at 310 K using the V-rescale algorithm (Duan et al. 2003), and the pressure was set at 1 bar using Parrinello-Rahman (Duan et al. 2003). The LINCS algorithm (Pronk et al. 2013; Hess et al. 1998) and the SETTLE algorithm (Krieger and Vriend 2015; Miyamoto and Kollman 1992) constrain all bonds, including hydrogen atoms and water molecules. The particle mesh Ewald method (Eastman and Pande 2010; Darden et al. 1993) was used to treat the long-range electrostatic forces, whereas van der Waals forces were treated using a cutoff of 1.2 nm. Trajectories were analyzed with the cpptraj module of Amber 16 (DeLano 2002).

Energy minimization and RMSD calculation of the protein models

Using the TM-align algorithm, most populated conformers obtained through clustering analysis were used to estimate RMSD between wild type and mutants. TM-align combines the TM-score rotation matrix and dynamic programming (DP) to identify the best structural alignment between protein pairs. This server was used for the RMSD calculation of the protein structures (Baker 2017). YASARA-minimization server was employed to perform the energy minimization of the most populated conformers of both wild type and mutants obtained through MD simulations. YASARA (Yet Another Scientific Artificial Reality Application) is a modeling and simulation software with multiple applications. YASARA-minimization server uses YASARA force field for energy minimization that can optimize the damage of the mutant proteins and thus precisely calculates the reliable energy. To perform this task, the PDB file of the wild-type and mutant proteins was inserted as input data, and the result was also additionally examined (Zhang and Skolnick 2005).

Effect of mutation in protein stability

I- Mutant 3.0 server was employed to predict the alteration in stability upon representative mutations. I-Mutant is a high-throughput support vector machine (SVM)-based tool server. The server can automatically predict the alteration in the stability of the structure by examining the structure of the protein sequence. I-Mutant 3.0

can be utilized as a classifier to predict the sign of stability with mutations and a regression calculator to predict the distinction in Gibbs free energy. The resulting DDG value indicates the difference between the Gibbs free energy of mutated protein and the wild-type protein in kcal/mol (Krieger et al. 2009).

Prediction of structural effects upon mutation

Project HOPE (Project Have your Protein Explained) was employed to understand the effect of the amino acid substitutions (Capriotti et al. 2005). HOPE server was utilized for molecular dynamics simulation to observe the effect of the mutations on the structure of p53. This web server performed a BLAST against the PDB, built a homology model of the query protein through YASARA (if applicable), and collected 3D structure data from WHAT IF web services. Subsequently, the sequence from the UniProt database was retrieved, and features like an active site, motifs, domains, and so forth were also shown. Finally, to predict the protein features, Distributed Annotation System (DAS) servers were utilized to exchange annotations on genomic and protein sequences (HOPE 2018).

Mutational spectrum analysis

Several analysis techniques, e.g., heat map, differential analysis, and survival analysis, were performed to retrieve the gene expression level of P53 in different types of cancer. For mutational spectrum analysis, data regarding mutation type in P53 were retrieved from cBioPortal for Cancer Genomics (https://www.cbioportal.org/). cBioPortal is an open-access resource for exploring large-scale datasets, where data were from both extensive consortium efforts (e.g., TCGA) and individual laboratories. Those data were organized in Microsoft Excel, and the heat map was generated in Heatmapper (Venselaar et al. 2010). The heat map used column Z score to compare the expression level among p53 mutation types in specific cancers, where $+4$ was the highest expression (red), and -4 was the lowest expression (blue).

Differential analysis and survival analysis

We employed GEPIA2 (Gene Expression Profiling Interactive Analysis) datasets for differential analysis, investigating a gene's differential expression patterns and comparing those with TCGA and GTEx data (Babicki et al. 2016). Here, the signature score, generated by GEPIA2, was gauged by the mean value of log2 (TPM + 1) of each gene of Th-1 like signature gene set, with a cutoff of 1 and *p* value cutoff of 0.01. Data were normalized by the overall survival method, with a 95% confidence interval. The group cutoff was selected median with a cutoff

high of 50% and a cutoff low of 50%. The statistical significance level was considered as *p* value ≤ 0.05.

Results

SNP database from for p53

The polymorphism data are available in several databases. NCBI dbSNP houses extensive data for different genes, and NCBI has the largest database that helps analyze single nucleotide polymorphisms. 581 SNPs have been found for cellular tumor antigen p53 (NCBI Reference Sequence: NP_000537.3). Among them, 420 SNPs were found to be missense or non-synonymous. Among 581, there were 435 SNPs in 3′UTR and 112 in 5′UTR regions. Only the missense, 3′ (3 prime), and 5′ (5 prime) UTR SNPs have been selected for further analysis.

Prediction of detrimental non-synonymous SNP

An online tool, SIFT, was employed to analyze an amino acid's conservancy by sequence homology; this helps to determine the conservation of a specific position of an amino acid in a protein. SIFT aligns paralogous and orthologous proteins' amino acid sequences while determining the effect of an amino acid replacement, which helps to analyze its functional importance and physical characteristics. SIFT takes rsIDs as input. Our analysis found 16 missense SNPs among 420 to be predicted deleterious using the SIFT web server (Table 1).

Table 1 Prediction of deleterious SNP by SIFT tool

SNP	Position	Amino acid change	Prediction	Score	Median
rs11540652	248	R248Q	Damaging	0.00	2.11
rs11540654	110	R110L	Damaging	0.04	2.13
rs11540654	110	R110P	Damaging	0.04	2.13
rs17849781	278	P278A	Damaging	0.00	2.11
rs28934571	249	R249S	Damaging	0.00	2.11
rs28934573	241	S241F	Damaging	0.00	2.11
rs28934574	282	R282W	Damaging	0.00	2.14
rs28934576	273	R273H	Damaging	0.00	2.11
rs28934577	257	L257Q	Damaging	0.00	2.11
rs28934578	175	R175H	Damaging	0.00	2.11
rs28934873	133	M133T	Damaging	0.00	2.13
rs28934874	151	P151T	Damaging	0.02	2.11
rs55832599	267	R267W	Damaging	0.00	2.11
rs112431538	285	E285K	Damaging	0.01	2.14
rs121913343	273	R273C	Damaging	0.00	2.11
rs121913342	248	R248W	Damaging	0.00	2.11

Table 2 Further analysis of 16 SIFT predicted nsSNPs by PredictSNP (online tool)

SNP	Mutation	Predict SNP (%)	MAPP (%)	PhD-SNP (%)	PolyPhen-1 (%)	Polyphen-2 (%)	SIFT (%)	SNAP (%)	nsSNPAnalyzer (%)	PANTHER (%)
rs11540654	R110L	76	72	82	59	47	46	50	63	48
rs11540654	R110P	87	59	88	59	63	46	56	63	57
rs28934873	M133T	51	77	61	67	76	79	50	63	56
rs28934874	P151T	87	48	88	59	55	79	72	63	–
rs28934578	R175H	87	63	88	59	56	79	72	63	–
rs28934573	S241F	87	77	88	74	81	79	89	63	–
rs11540652	R248Q	87	59	88	59	81	53	81	65	–
rs121913342	R248W	87	77	88	74	81	79	85	65	–
rs28934571	R249S	87	77	88	74	81	53	81	63	–
rs28934577	L257Q	87	78	88	74	81	79	72	63	–
rs55832599	R267W	87	77	88	74	65	79	81	63	–
rs121913343	R273C	87	91	88	74	81	79	85	63	–
rs28934576	R273H	87	77	88	59	65	53	62	65	–
rs17849781	P278A	87	77	88	59	81	53	81	63	–
rs28934574	R282W	87	77	88	74	81	79	72	63	–
rs112431538	E285K	87	84	77	74	68	79	62	65	–

Table 3 Cancer-associated SNP where the threshold value is − 0.75. Scores having − 0.75 predicted as having their association with cancer

SNP	Position	Substitution	Prediction	Score
rs11540652	248	R248Q	Cancer	− 9.67
rs11540654	110	R110L	Cancer	− 9.06
rs11540654	110	R110P	Cancer	− 9.05
rs17849781	278	P278A	Cancer	− 9.85
rs28934571	249	R249S	Cancer	− 9.67
rs28934573	241	S241F	Cancer	− 9.61
rs28934574	282	R282W	Cancer	− 9.73
rs28934576	273	R273H	Cancer	− 9.71
rs28934577	257	L257Q	Cancer	− 10.07
rs28934578	175	R175H	Cancer	− 9.93
rs28934873	133	M133T	Cancer	− 9.08
rs28934874	151	P151T	Cancer	− 9.88
rs55832599	267	R267W	Cancer	− 9.73
rs112431538	285	E285K	Cancer	− 9.29
rs121913343	273	R273C	Cancer	− 9.73
rs121913342	248	R248W	Cancer	− 9.73

Analysis of SIFT predicted deleterious SNPs

In SIFT analysis, 16 SNPs were found to be detrimental. These 16 SNPs were also analyzed in an online SNP analyzing tool known as PredictSNP, a package system where other SNP analyzing methods have been assembled. Protein sequences in FASTA formats were used as inputs in PredictSNP, and following this, the SIFT result mutation was done in the PredictSNP. The impact of the mutation was analyzed by choosing nsSNPAnalyzer, MAPP, PANTHER, PolyPhen-1, PhD-SNP, PolyPhen-2, SNAP, and SIFT tools. PredictSNP shows the result of different SNPs, mentioning which percentage are either characterized as detrimental or neutral; the percentage value indicates the confidence of the result. SNPs predicted as neutral by more than one tool have been excluded from our study (Table 2).

Identification of cancer-associated SNPs from predicted deleterious SNPs

SNPs predicted as deleterious were then analyzed using Fathmm to determine whether they were associated with cancer. Our analysis revealed that every SNP predicted as detrimental was also predicted as having their association with cancer. In the Fathmm server, the default prediction threshold is − 0.75, where a prediction with a score less than this indicates that the mutation is potentially associated with cancer (Table 3). Our result finds that the predicted score is much lower than the threshold, indicating a much higher potentiality of relating these SNPs to cancer (Table 3).

Identification of functionally important SNPs in the conserved regions

Some amino acids are crucial for the function of a protein. Essential amino acids contributing toward specific functions tend to be evolutionarily conserved. To identify the evolutionarily conserved amino acids and

Table 4 Summary of amino acids (conservation profile) in that corresponds to region with high-risk nsSNPs. CS: conservation score, color (6–9 = conserved, 5 = average and 1–4 = variable); (s): predicted structural site; (f): predicted functional site

Residue position	Amino acid	CS	Color	B/E	function
110	R	1.057	1	E	S
133	M	− 1.019	9	B	S
151	P	− 0.801	8	E	F
175	R	− 0.989	9	E	F
241	S	− 1.162	9	E	F
248	R	− -1.070	9	E	F
249	R	− 0.990	9	E	F
257	L	− 0.951	9	B	S
267	R	− 0.944	9	E	F
273	R	− 1.073	9	E	F
278	P	− 1.030	9	E	F
282	R	− 1.029	9	E	F
285	E	− 0.878	8	E	F

the proteins, we used the online tool ConSurf. ConSurf results are tabulated, where conserved amino acids in a specific position are shown with CS and color value, where the lower the CS and the higher the color value, the higher the conservancy (Table 4).

Functional SNPs in UTR identification

3′UTR regions significantly affect gene expression due to the defective ribosomal RNA translation or RNA half-life. 5′UTR also plays an important role in mRNA stabilization. The UTRscan server analyzed 77 3′UTR

Table 6 Functional domains and their position in p53 protein

Amino acid position in the protein	Domain name
5–29	p53 transactivation domain (p53_TAD)
95–288	p53 DNA binding domain
319–357	p53 tetramerization domain

SNPs and 129 5′UTR SNPs of the p53 gene (Table 5). The UTRscan server looks for patterns in the UTR database for regulatory region motifs and, according to the given SNP information, predicts if any matched regulatory region is damaged (Tang et al. 2019; Pesole and Liuni 1999). UTRscan found 8 UTRsite motif matches in the p53 transcript. A total of 141 matches were found for 5 motifs.

Prediction of a deleterious mutation in the functional domain of p53

InterProScan tool analysis found 3 functional domains within the p53 gene. InterProScan takes the FASTA format of protein sequences as input and scans for matching protein sequences against the InterPro protein signature databases. InterProScan also predicts the function of the residues, provides the consensus amino acid for protein function, and determines whether the predicted deleterious mutation SNP is necessary for a function or not. We can authenticate our prediction if

Table 5 Result shows UTR regions in the p53 transcript from the UTRScan server

Signal name	UTR region	Match total	Position in transcript
uORF (Upstream open reading frames)	5′	44	–
	3′	25	–
GY-BOX	5′	9	–
	3′	6	–
IRES (Internal ribosome entry site**)**	5′	17	–
	3′	11	–
TOP (Terminal oligopyrimidine)	5′	4	–
	3′	9	–
UNR-bs	3′	7	–
PAS (Polyadenylation Signal)	3′	6	–
K-BOX	5′	2	13–70 1–8
ARE2	3′	1	409–444

Alam *et al. Bulletin of the National Research Centre* (2022) 46:191

Page 9 of 17

Table 7 Functional residues in different functioning sites of p53 protein

Functioning site	Functioning residue with position
Zinc binding site	176C, 179H, 238C, 242C
Dimerization site	177P, 178H, 179H, 181R
DNA binding site	239N, 241S, 248R, 273R, 275C, 276A, 277C, 280R

the predicted deleterious SNP is found in the amino acid in the functional domain or functioning residue in Table 6. The InterProScan result is tabulated in Table 7.

Comparative modeling of high-risk Non-synonymous SNPs and MD simulations

We used SWISS-MODEL to get the 3D structure of human p53 protein with the predicted SNPs. Initially, the stability of the p53 protein was evaluated through one microsecond (μs). Analysis of root-mean-squared deviation (RMSD) and radius of gyration (Rg) values considering backbone atoms showed that p53 protein reached constant RMSD (Fig. 3A) and Rg (Fig. 3B) values after 0.6 μs with an average Rg value of 22.3 ± 0.3. Based on this result, clustering analysis was performed over the equilibrated simulation time (last 0.4 μs) to obtain the most populated conformer. Subsequently, the following non-synonymous mutations in the DNA binding domain were introduced: S241F, R248Q, R248W, R273H, and R273C using PyMOL, to obtain the p53 mutants. Analysis of the mutants shows that the five mutants reached stable RMSD (Fig. 3C) and Rg (Fig. 3D) values between 20 and 30 ns. Average Rg values showed the following values R248Q (23.9 ± 0.2), R273C (22.2 ± 0.2), R273H (23.6 ± 0.3), S241F (23.4 ± 0.2), and R248W (21.5 ± 0.2). This analysis indicates that R248Q, R273H, and S241F systems experience an increment of the hydrodynamic radius compared with the wild-type protein, whereas the R248W showed a small decrease and R273C maintained a similar radius to the wild-type protein.

Root-mean-square fluctuation (RMSF) analysis over the equilibrated simulation time (Fig. 3C) showed that the regions with the highest mobility are localized between the N-terminal region and residue 40 and between residues 60 and 90. Both these regions correspond to a long loop with four small α-helices. The lowest mobility of the region between the N-terminal region and residue 40 was observed for R248W, which also showed the lowest average Rg value, suggesting that this region could be responsible for the differences in Rg values.

Clustering analysis was performed over the equilibrated simulation time (last 70 ns) to obtain the most populated conformers. Wild type and mutants were then subjected to the YASARA energy minimization server for energy minimization. Energy minimization results showed decreased free energy for all mutant models compared to the wild-type models. The results are shown in Table 8. RMSD was calculated using the Tm-align tool, where the results were shown to be between 3.0 and 4.0 Å. These outcomes demonstrate a critical change in the protein structure that can alter its natural function.

After mutation of the wild type, it was found that in every case, energy after minimization was much higher (more positive) for the mutants than the conventional wild type, indicating that these mutations destabilize the structure of the protein. In case of a mutation in the R273H and R273C domains, changing the position of the amino acid arginine by histidine or cysteine affects the structure of the protein more than the other mutations.

Prediction of protein structural stability

We used the neural network-based routine tool I-Mutant 3.0 to study the potential change in protein stability upon mutations. This tool took the input of the mutated protein models derived from the PHYRE-2 server in PDB format. I-Mutant 3.0 creates results taking the help of the ProTherm database. This database housed extensive experimental data on free energy alterations due to mutations. In addition, this tool predicts the score of free energy change due to mutations, incorporating the energy-based online tool FOLD-X. This increases the precision to 93% on one-third of the database if the FOLD-X analysis is incorporated with I-Mutant (Datta et al. 2015). Models with the following mutations: S241F, R248Q, R248W, R273H, and R273C were subjected to the server to predict DDG stability and RSA calculation. The result shows that every mutation decreased the stability of the protein. Mutation R273H was responsible for the lowest DDG value (-1.62 kcal/mol), followed by R273C (-1.52 kcal/mol). DDG values for other mutations ranged from -0.51 kcal/mol to -0.93 kcal/mol; these negative DDG values decreased protein stability. The results are shown in Table 9.

Analysis of structural effect upon mutation in DNA binding domain

The InterProScan tool was used to find the functional domain in p53 protein and map the predicted deleterious mutations in these domains to speculate the changes they might cause in the domain structures. Among the predicted 14 detrimental SNPs revealed by different SNP analyzer tools, we found 5 missense SNPs in the 3 crucial amino acids located in a domain responsible for DNA binding. These amino acids are
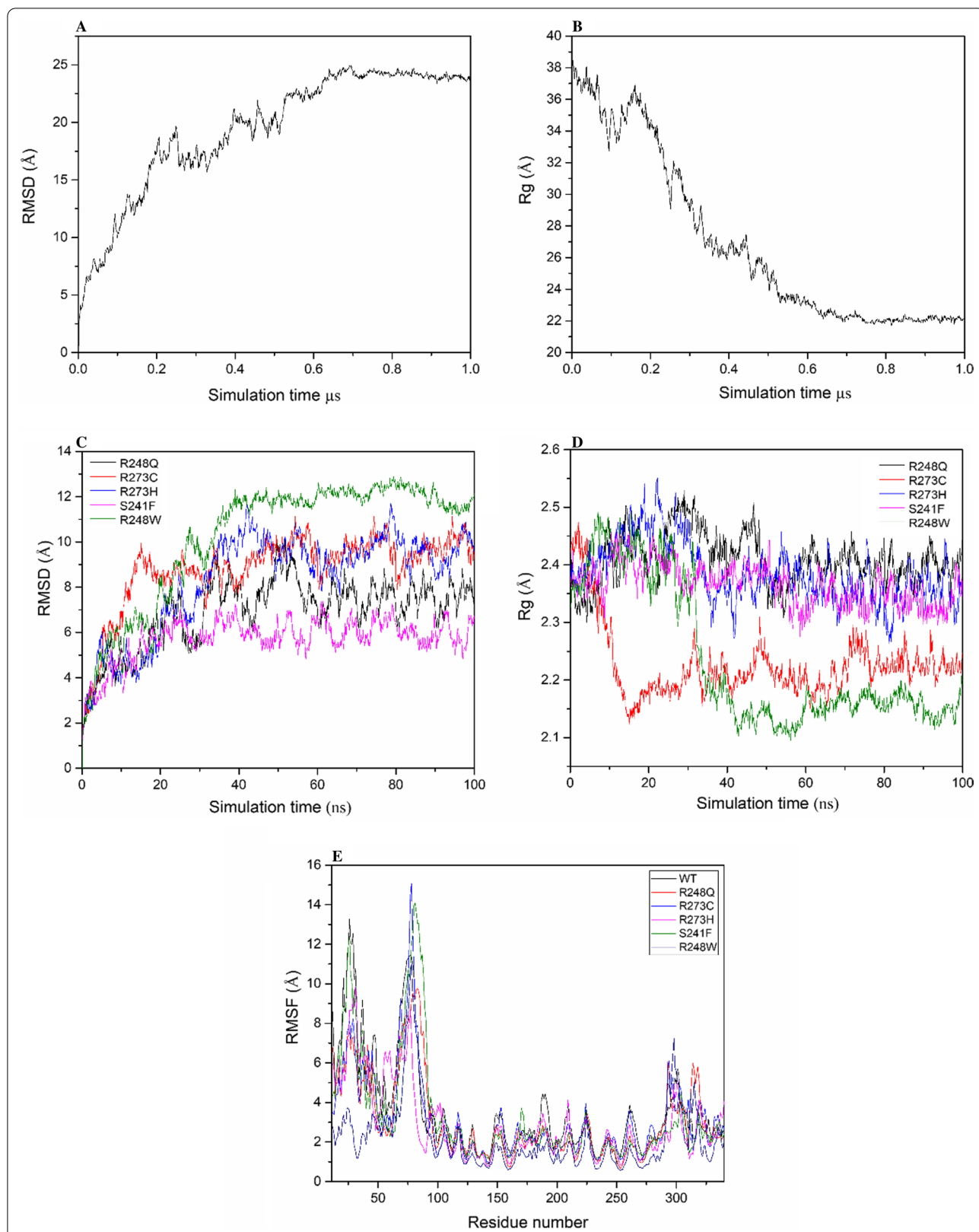
Fig. 3 RMSD, Rg, and RMSF analysis of the wild-type and mutant p53 protein. RMSD (**A**) and Rg (**B**) analysis of wild-type p53. RMSD (**C**), Rg (**D**), and RMSF (**E**) analysis of the wild type and mutants of p53 protein

Alam *et al. Bulletin of the National Research Centre*   (2022) 46:191

Page 11 of 17

Table 8 RMSD (Å) values and total free energy after energy minimization of both wild-type and mutated protein models

Mutated models	Energy after minimization (kj/mol)	RMSD (Å)
Wild Type	− 161,568.4	
S241F	− 157,568.4	4.32
R248Q	− 159,349.4	4.27
R248W	− 158,244.0	3.74
R273H	− 156,160.6	4.68
R273C	− 155,923.0	3.60

Table 9 I-mutant predictions for selected nsSNPs

Mutation	Sign of DDG	DDG value prediction Kcal/mol	RI	RSA
S241F	Decrease	− 0.51	7	47.8
R248Q	Decrease	− 0.93	9	88.6
R248W	Decrease	− 0.58	5	88.6
R273H	Decrease	− 1.62	9	20.5
R273C	Decrease	− 1.52	8	20.5

essential for the functional activity of the domain. Therefore, a mutation in this amino acid position could change the protein structure and function. We observe the effect on the structure of p53 due to these 5 missense SNPs using an online tool, HOPE.

In Fig. 4A, the wild-type residue has positively charged arginine (R). However, the mutation from arginine to glutamine (Q) at the 248th position makes the mutant neutral. In Fig. 4B, serine (S) mutated into phenylalanine (F) at 241th. The mutant residue is bigger and more hydrophobic than the wild type. In Fig. 4C, arginine (R) mutated into histidine (H) at 273rd position, and the mutant residue is more minor and neutral, whereas the wild type is positively charged. In Fig. 4D, the arginine (R) mutated into cysteine (C) at the 273rd position, and the mutant residue is more minor and neutral, but the wild type is positively charged. In Fig. 4E, arginine (R) mutated into tryptophan (W) at the 248th position. The mutant residue is more considerable and neutral, whereas the wild type is more significant and positively charged.
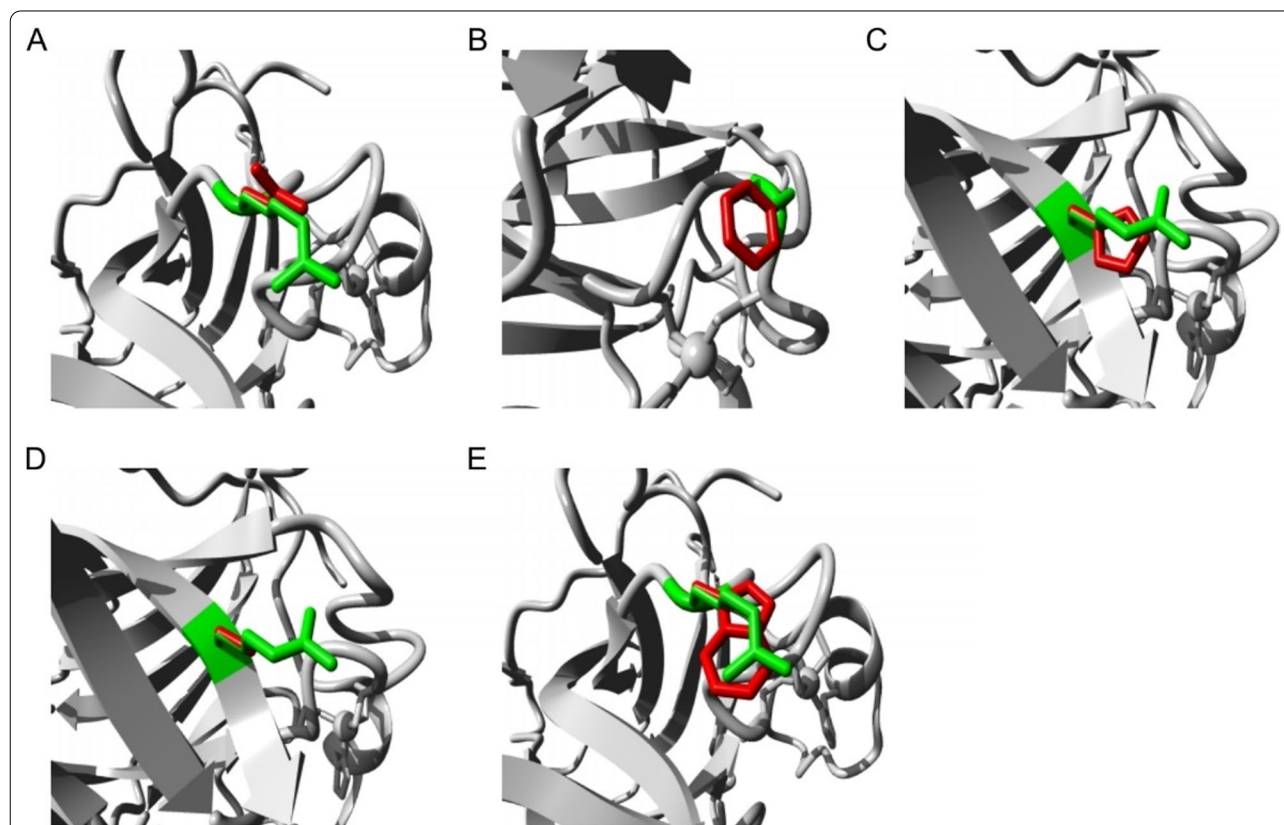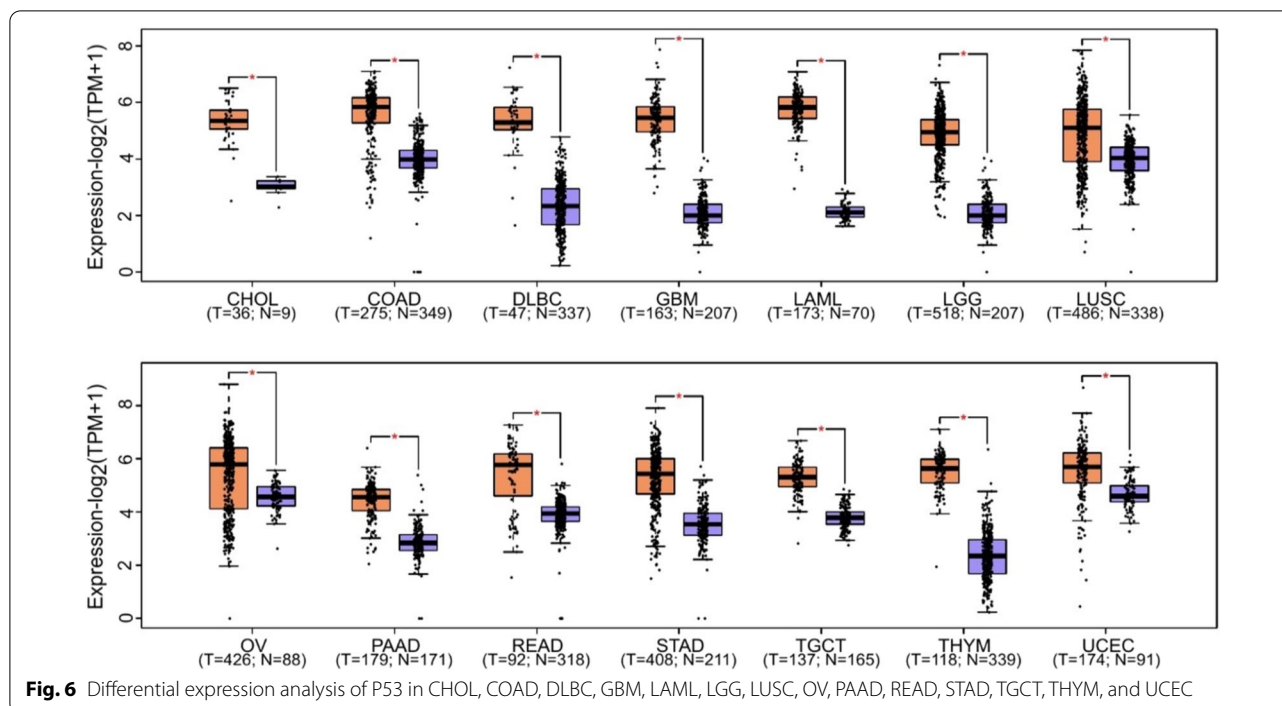

Fig. 4 Effect in the structure due to mutation **A** R248Q, **B** S241F, **C** R273H, **D** R273C, **E** 248W (The protein is colored gray, green indicates the side chains of wild type, and mutant residue is shown in red)

Alam *et al. Bulletin of the National Research Centre*   (2022) 46:191

Page 12 of 17

Fig. 5 Mutational spectrum analysis of 33 cancer subtypes
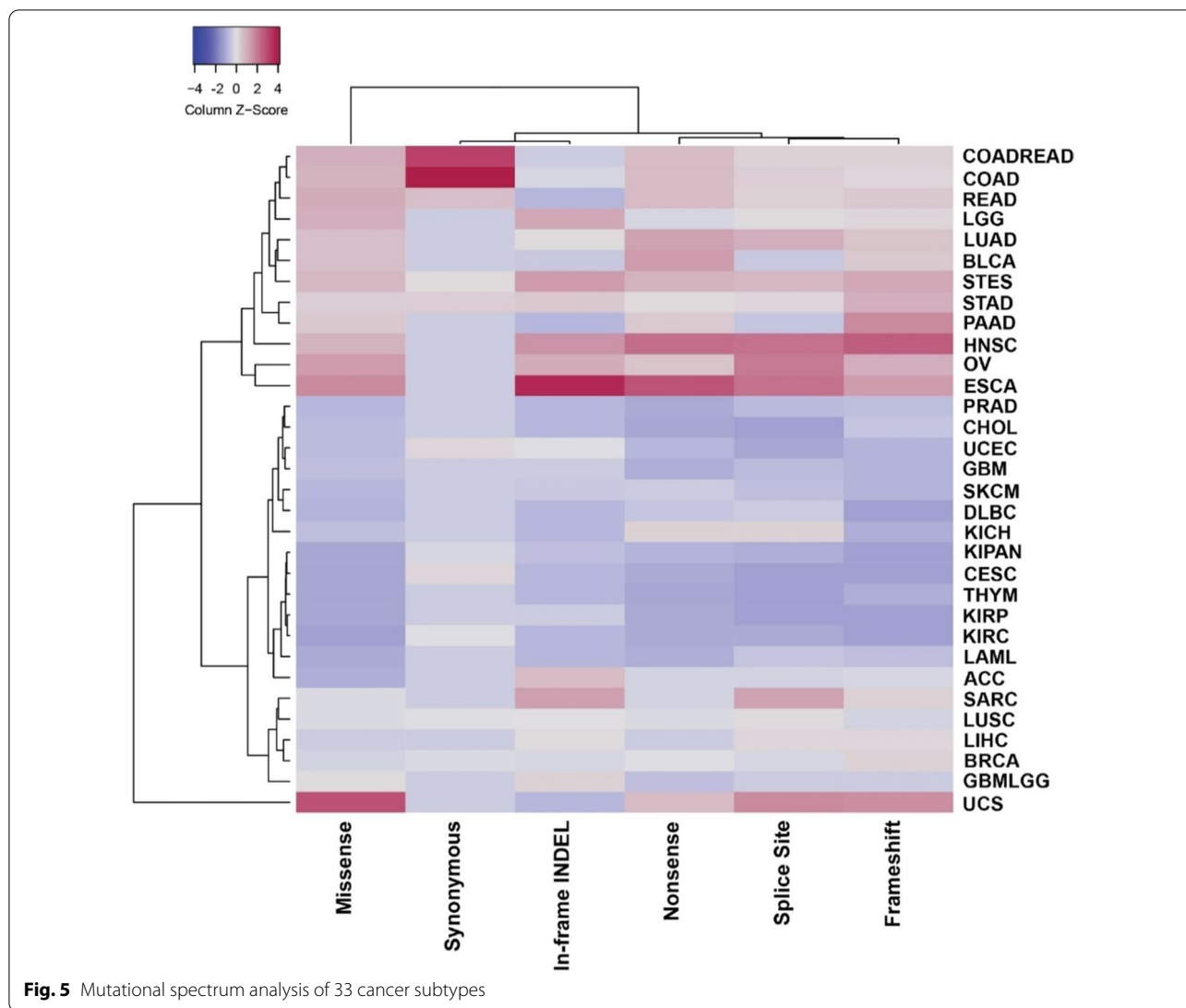
Evaluation of p53 gene mutation level in various cancer
For the mutational spectrum analysis, we investigated the expression of six mutation types of p53—missense, synonymous, in-frame indel, nonsense, splice site, and frameshift in 33 cancer subtypes including ACC (Adrenocortical carcinoma), BLCA (bladder urothelial carcinoma), BRCA (breast invasive carcinoma), CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma), CHOL (cholangiocarcinoma), COAD (colon adenocarcinoma), DLBC (lymphoid neoplasm diffuse large B cell lymphoma), ESCA (esophageal carcinoma), GBM (glioblastoma multiforme), HNSC (head and neck squamous cell carcinoma), KICK (kidney chromophobe), KIRC (kidney renal clear cell carcinoma), KIRP (kidney renal papillary cell carcinoma), LAML (acute myeloid leukemia), LGG (brain lower grade glioma), LIHC (liver hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), MESO (mesothelioma), OV (ovarian serous cystadenocarcinoma), PAAD (pancreatic adenocarcinoma), PCPG (pheochromocytoma and paraganglioma), PRAD (prostate adenocarcinoma), READ (rectum adenocarcinoma), SARC (sarcoma), SKMC (skin cutaneous melanoma), STAD (stomach adenocarcinoma), TGCT (testicular germ cell tumors), THCA (thyroid carcinoma), THYM (thymoma), UCEC (uterine corpus endometrial carcinoma), UCS (uterine carcinosarcoma), and UVM (uveal melanoma) (Fig. 5). Employing column Z score, significant upregulation was represented in red and downregulation in blue. The heat map generated the hierarchical clustering of cancer subtypes based on their level of similarity. COADREAD to ESCA subtypes were assorted in one cluster and PRAD to GBMLGG in another, leaving out UCS.

Fig. 6 Differential expression analysis of P53 in CHOL, COAD, DLBC, GBM, LAML, LGG, LUSC, OV, PAAD, READ, STAD, TGCT, THYM, and UCEC

GBMLGG and UCS showed different expression patterns despite being in adjacent positions. Significant upregulation of missense and frameshift mutation had been perceived in the UCS cancer subtype, which supports the findings that the highest mutation rates in p53 results in UCS (91.2%), followed by OV (83%) (Bhagwat 2010).

Differential gene expression analysis and correlation of p53 with the survival rate of patients

The investigation was proceeded further by assessing the transcription level of p53 in normal and tumor cells, where the blue box denoted normal cells, and cancer cells were marked by the orange box in the box plot. The result showed a significant difference between the expression level of standard and tumor cells in CHOL, COAD, DLBC, GBM, LAML, LGG, LUSC, OV, PAAD, READ, STAD, TGCT, THYM, and UCEC subtypes (Fig. 6). In these subtypes, the expression of p53 was upregulated in tumor cells, implying that p53 mutation has a strong association with the occurrence of malignancy. According to Perri et al., 2016, more than 50% of human carcinogenesis arises from the genetic alteration of the p53 gene (Wang and Sun 2017).

Survival analysis estimates the statistical probability of the survival period on-time event for cancer patients. The Kaplan–Meier method approximates the survival probability and visualizes the survival plots (Perri et al. 2016; Susmi et al. 2021). We compared the overall survival period between the high p53 and low p53 groups in different cancer subtypes. Only BRCA, COAD, LGG, and PRAD exhibited statistically significant outputs among the subtypes. The survival plots disclosed that the high expression of P53 was directly correlated with the high survival rate in LGG and COAD. Conversely, a higher survival rate was associated with low levels of P53 expression in BRCA and PRAD (Fig. 7).

Discussion

Single Nucleotide Polymorphisms or SNPs are the most common nucleic acid variations that result in differences among humans; SNPs are also responsible for many hereditary disorders due to amino acid substitutions. Though approximately 4 million SNPs could be found in the database, many SNPs do not cause disease-causing alterations to protein structure due to amino acid degeneracy, which consummately dispels mutations in critical functional regions. Genetic studies to differentiate the functionally neutral nature and disease-associated polymorphism have become a significant concern. Henceforth, SNPs that become dispersed throughout the genome often become excellent genetic markers. Most non-synonymous SNPs associated with the diseases are generally found in the exonic regions, but SNPs that occur in the intrinsic sites of gene disrupting regulatory regions ultimately affect the splicing process. With the increasing rate of reported SNPs in different databases, extensive population-based study becomes difficult due

Fig. 7 Survival analysis plot of p53 in LGG, BRCA, COAD, and PRAD expressing the correlation of p53 expression with patients' survival period (months)
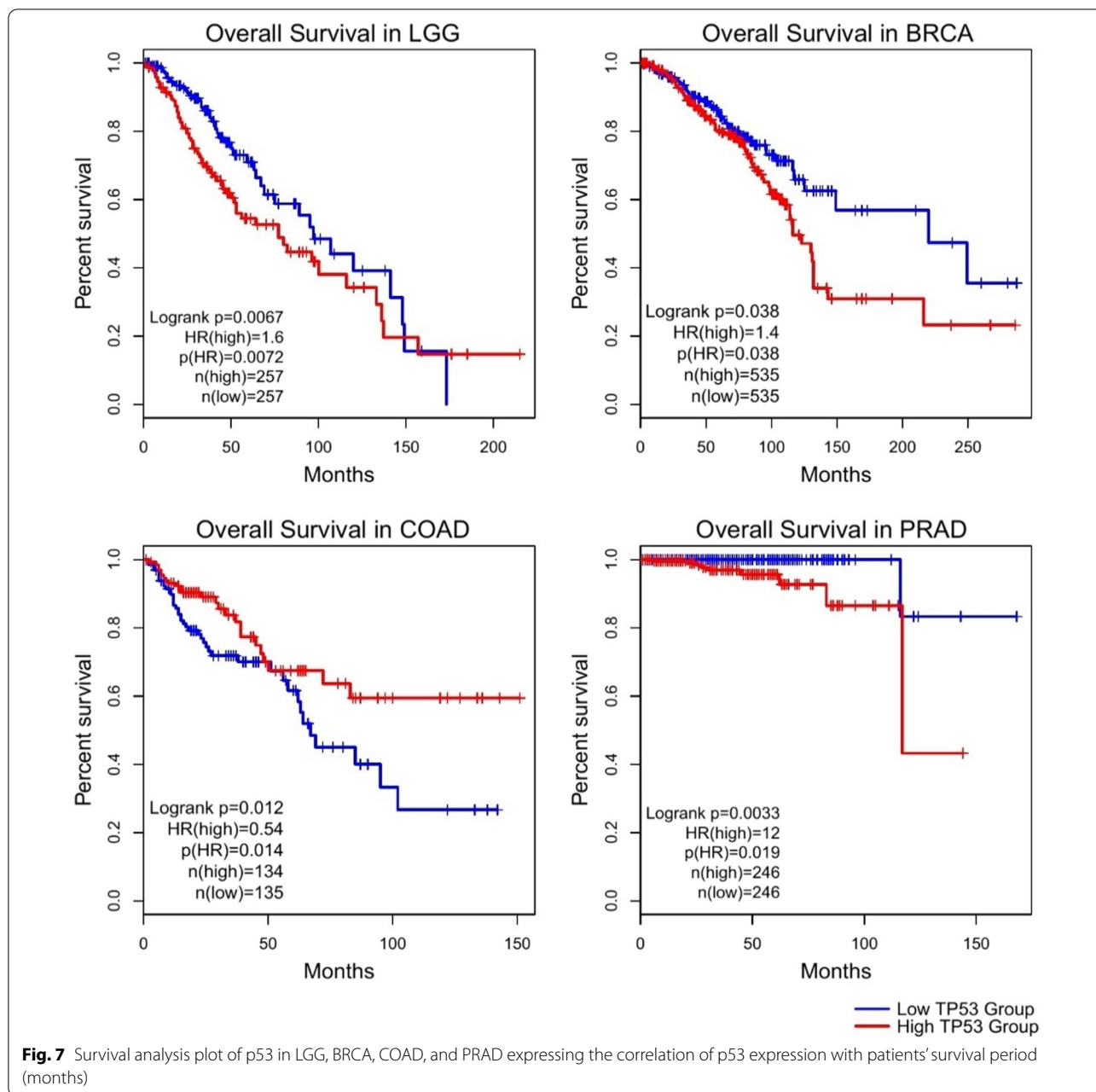
to the cost, and it remains tough to select a target SNP for the investigation while identifying the ones most prone to cause diseases. However, *an in silico* approach to detecting detrimental SNPs can be more helpful.

This study analyzed the SNP databases to seek SNPs that might be detrimental to p53, following data-driven methods. Search for nsSNPs against p53 resulted in 420 hits. The rsIDs were submitted to SIFT and PolyPhen-2 servers. SIFT and PolyPhen found 16 nsSNPs as non-tolerable and most likely damaging (Tables 1, 2). By performing the Fathmm test, we found 14 cancer-associated

SNPs. ConSurf helps to predict evolutionarily conserved amino acids and found 12 SNPs. We found three functional domains and their position in the p53 gene by analyzing them through the InterPro scan server. SWISS-MODEL allowed predicting the 3D structure, which was refined through one μs of MD simulations. Clustering analysis allowed obtaining the most populated conformer over the equilibrated simulation time, including mutations. The 5 non-synonymous SNPs in the DNA binding domain (S241F, R248Q, R248W, R273H, and R273C) were predicted with PyMOL. These mutants were also

submitted through MD simulations to obtain the most populated conformers over the equilibrated simulation time. YASARA energy minimization server showed decreased free energy for all the mutant models compared to the wild-type models. The least energy was minimized in case of mutation in the 273rd arginine amino acid position by histidine and cysteine, which affects the protein structure more than the other mutations.

Regarding human cancer, the p53 gene becomes the most frequently mutated gene, and the predominance of missense mutations is scattered over 200 codons (Ferreira and Patino 2016). p53 receives inputs from stress and abnormality sensors that function within the cell's intracellular operating systems; if the degree of damage to the genome is excessive or if the levels of nucleotide pools, growth-promoting signals, glucose, or oxygenation are suboptimal, p53 can potentially halt further cell-cycle progression until these conditions have been normalized (Vogelstein et al. 2000). Alternatively, in the face of alarm signals, indicating overwhelming or irreparable damage to such cellular subsystems, p53 can also trigger the process of apoptosis. Mutation in p53 results in the loss of regulation or over-proliferation (Fridman and Lowe 2003). Tumor cells evolve various strategies to limit or circumvent apoptosis. The most common one includes the loss of p53 tumor suppressor function, eliminating this critical damage sensor from the apoptosis-inducing circuitry (Vogelstein et al. 2000).

The functional domains of p53 have been subjected to extensive analysis. We found 5 different SNPs in the functional domains of the p53 gene that are deleterious by analyzing with different dry-laboratory tools. R248 and R273 residue have a role in the structural integrity of the functional domain (Greenblatt et al. 1994; Hainaut et al. 1997). The tetrameric p53 protein (which is a dimer of a dimer) binds to four repeats of a consensus DNA sequence. S241 connects with the phosphate backbone in the major groove (Greenblatt et al. 1994). Amino acid substitution in the sequence of the functional domain may lead to alterations of the protein structures (Cho et al. 1994).

Mutations of the p53 gene cause diverse types of cancer in humans. The research found that 70% of mutations occurred in the p53 gene in lung cancer-affected patients. 45%, 60%, 20%, 10–30%, 60%, 40%, 10%, 30%, 60%, and 60% of mutations found in p53 gene, respectively, in stomach, colon, liver, prostate, head/neck, esophagus, leukocytes, lymphocytes, ovary, and bladder cancer-affected patients (Walker and Levine 1996).

p53 was allowed to go through extensive gene expression analysis. Utilizing Heatmapper, the heat map was engendered to assess the expression level of different mutations of p53 in 33 cancer subtypes. GEPIA2 was employed for differential analysis and survival analysis. Differential analysis revealed a substantial upregulation of p53 in the tumor cells in CHOL, COAD, DLBC, GBM, LAML, LGG, LUSC, OV, PAAD, READ, STAD, TGCT, THYM, and UCEC. Most P53 mutations lead to oncogenic progression (Wang and Sun 2017; IARC TP53 Database 2018). Additionally, survival analysis estimated the interconnection of the survival period with the gene expression. In the case of LGG and COAD, the expression level is positively interrelated with survival probability, whereas BRCA and PRAD demonstrate a negative correlation of survival period with the gene expression. This study revealed that despite some correct assumptions, web-based tools need to be more precise in detecting deleterious SNPs, and population-based studies are necessary to identify and further test the predicted SNPs in different populations.

Conclusions

In this study, different SNP analyzing tools have been employed to analyze the available data from the NCBI dbSNP database for the tumor suppressor gene p53. The predicted deleterious SNPs were evaluated for their potentially detrimental effects on protein function and stability. Five SNPs were predicted to be deleterious—rs28934573 (S241F), rs11540652 (R248Q), rs121913342 (R248W), rs121913343 (R273C), and rs28934576 (R273H); they have the highest probability to make p53 functional by changing their structure and functional residues involved in the active site formation. Henceforth, it is very likely that there are unreported nsSNPs that increase disease predisposition by altering protein function or structure. The findings of this study may help in the early diagnosis of the detrimental SNPs that have the probability of increasing the risk of different types of cancers. Individuals diagnosed with the above nsSNPs can take precautions to avoid other risk factors associated with cancer development as they are susceptible to cancer due to these nsSNPs in p53, a significant tumor suppressor gene. However, population-based studies and wet-laboratory experiments are beyond our scope for verifying the current study's findings. Therefore, extensive clinical studies are required to characterize the vastly available SNP data.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors report no competing interests. The authors alone are responsible for the content and writing of this article.

Author details
[1]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh. [2]Department of Mathematics and Natural Sciences, BRAC University, Dhaka 1212, Bangladesh. [3]Laboratorio de Diseño y Desarrollo de Nuevos Fármacos E Innovación Biotecnológica de La Escuela Superior de Medicina, Instituto Politécnico Nacional, Mexico City, Mexico. [4]Department of Mathematics and Natural Sciences, Biotechnology Program, School of Data and Sciences, BRAC University, Dhaka, Bangladesh. [5]Department of Biological Sciences, St John's University, Queens, NY 11439, USA. [6]Department of Biochemistry and Molecular Biology, Tejgaon College, National University of Bangladesh, Gazipur 1704, Bangladesh. [7]Department of Public Health, North South University, Bashundhara, Dhaka 1229, Bangladesh.

References

Ahuja H, Bar-Eli M, Advani SH et al (1989) Alterations in the p53 gene and the clonal evolution of the blast crisis of chronic myelocytic leukemia. Proc Natl Acad Sci USA 86:6783–6787

Ashkenazy H, Erez E, Martz E et al (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529–W533. https://doi.org/10.1093/nar/gkq399

Babicki S, Arndt D, Marcu A et al (2016) Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res 44:W147–W153. https://doi.org/10.1093/NAR/GKW419

Baker SJ, Markowitz S, Fearon ER et al (1990) Suppression of human colorectal carcinoma cell growth by wild-type p53. Science 249:912–915

Baker T (2017) Molecular computer simulations of graphene oxide intercalated with methanol: swelling properties and interlayer structure.

Barroso I, Gurnell M, Crowley VEF et al (1999) Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 402:880–883. https://doi.org/10.1038/47254

Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS (2018) Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ 25(1):154–160. https://doi.org/10.1038/cdd.2017.180

Bendl J, Stourac J, Salanda O et al (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 10:e1003440. https://doi.org/10.1371/journal.pcbi.1003440

Bhagwat M (2010) Searching NCBI's dbSNP database. Curr Protoc Bioinform Chap. https://doi.org/10.1002/0471250953.BI0119S32

Bressac B, Galvin KM, Liang TJ et al (1990) Abnormal structure and expression of p53 gene in human hepatocellular carcinoma. Proc Natl Acad Sci USA 87:1973–1977

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306. https://doi.org/10.1093/NAR/GKI375

Carninci P, Kasukawa T, Katayama S et al (2005) The transcriptional landscape of the mammalian genome. Science (80-) 309:1559–1563. https://doi.org/10.1126/SCIENCE.1112014

Case DA, Cheatham TE, Darden T et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688. https://doi.org/10.1002/JCC.20290

Chakravarti A (2001) Single nucleotide polymorphisms to a future of genetic medicine. Nature 409:822–823. https://doi.org/10.1038/35057281

Chasman D, Adams RM (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation11Edited by F. Cohen J Mol Biol 307:683–706. https://doi.org/10.1006/jmbi.2001.4510

Chng WJ, Price-Troska T, Gonzalez-Paz N, Van Wier S, Jacobus S, Blood E, Henderson K, Oken M, Van Ness B, Greipp P, Rajkumar SV (2007) Clinical significance of TP53 mutation in myeloma. Leukemia 21(3):582–584. https://doi.org/10.1038/sj.leu.2404524

Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Sci 265:346–355. https://doi.org/10.1126/SCIENCE.8023157

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092. https://doi.org/10.1063/1.464397

Datta A, Mazumder M, Hasan H, Chowdhury AS, Hasan M (2015) Functional and structural consequences of damaging single nucleotide polymorphisms in human prostate cancer predisposition gene RNASEL. Biomed Res Int 8:2015. https://doi.org/10.1155/2015/271458

DeLano WL (2002) The PyMOL molecular graphics system. https://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=1958992. Accessed 20 Sep 2020

Diller L, Kassel J, Nelson CE et al (1990) p53 functions as a cell cycle control protein in osteosarcomas. Mol Cell Biol 10:5772–5781

Discovery Studio 4.0 - Updates. http://accelrys.com/resource-center/downloads/updates/discovery-studio/dstudio40/latest.html. Accessed 21 May 2018

Doniger SW, Kim HS, Swain D et al (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4:e1000183. https://doi.org/10.1371/journal.pgen.1000183

Dryja TP, McGee TL, Hahn LB et al (1990) Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. N Engl J Med 323:1302–1307. https://doi.org/10.1056/NEJM199011083231903

Duan Y, Wu C, Chowdhury S et al (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012. https://doi.org/10.1002/jcc.10349

Eastman P, Pande VS (2010) Constant constraint matrix approximation: a robust, parallelizable constraint method for molecular simulations. J Chem Theory Comput 6(2):434–437. https://doi.org/10.1021/ct900463w

Eliyahu D, Michalovitz D, Eliyahu S et al (1989) Wild-type p53 can inhibit oncogene-mediated focus formation. Proc Natl Acad Sci U S A 86:8763–8767

fathmm - Analyze Cancer-Associated Variants. http://fathmm.biocompute.org.uk/cancer.html. Accessed 21 May 2018

Ferreira JC, Patino CM (2016) What is survival analysis, and when should I use it? J Bras Pneumol 42:77. https://doi.org/10.1590/S1806-37562016000000013

Finlay CA, Hinds PW, Levine AJ (1989) The p53 proto-oncogene can act as a suppressor of transformation. Cell 57:1083–1093

Fridman JS, Lowe SW (2003) Control of apoptosis by p53. Oncogene 22:9030–9040. https://doi.org/10.1038/sj.onc.1207116

Greenblatt MS, Bennett WP, Hollstein M, Harris CC (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis - PubMed. Cancer Res 54:4855–4878

Grillo G, Turi A, Licciulli F et al (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 38:D75–D80. https://doi.org/10.1093/nar/gkp902

Hainaut P, Soussi T, Shomer B et al (1997) Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. Nucleic Acids Res 25:151

Alam *et al. Bulletin of the National Research Centre*   (2022) 46:191

Page 17 of 17

Hamosh A, Scott AF, Amberger JS et al (2004) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33:D514–D517. https://doi.org/10.1093/nar/gki033

Hamzehloie T, Mojarrad M, Hasanzadeh Nazarabadi M, Shekouhi S (2012) The role of tumor protein 53 mutations in common human cancers and targeting the murine double minute 2–p53 interaction for cancer therapy. Iran J Med Sci 37(1):3

Hess B, Bekker H, Berendsen HJC, GEM. Fraaije J (1998) LINCS: a linear constraint solver for molecular simulations - Hess – 1997. J Comput Chem 18

HOPE (2018) http://www.cmbi.ru.nl/hope/input/. Accessed 21 May 2018

IARC TP53 Database. http://p53.iarc.fr/. Accessed 4 Oct 2018

Kastenhuber ER, Lowe SW (2017) Putting p53 in context. Cell 170(6):1062–1078. https://doi.org/10.1016/j.cell.2017.08.028

Kelly JN, Barr SD (2014) In silico analysis of functional single nucleotide polymorphisms in the human TRIM22 gene. PLoS ONE 9(7):e101436. https://doi.org/10.1371/journal.pone.0101436

Krieger E, Vriend G (2015) New ways to boost molecular dynamics simulations. J Comput Chem 36(13):996–1007. https://doi.org/10.1002/jcc.23899

Krieger E, Joo K, Lee J et al (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins Struct Funct Bioinforma 77:114–122. https://doi.org/10.1002/prot.22570

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Liu J, Gough J, Rost B (2006) Distinguishing protein-coding from non-coding RNAs through support vector machines. PLoS Genet 2:e29. https://doi.org/10.1371/journal.pgen.0020029

Matsuda T, Tomita M, Uchihara J-N et al (2005) Human T cell leukemia virus Type I-infected patients with Hashimoto's thyroiditis and Graves' disease. J Clin Endocrinol Metab 90:5704–5710. https://doi.org/10.1210/jc.2005-0679

Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962. https://doi.org/10.1002/JCC.540130805

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814

Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61–80. https://doi.org/10.1146/annurev.genom.7.080505.115630

Perri F, Pisconti S, Vittoria Scarpati GD (2016) P53 mutations and cancer: a tight linkage. Ann Transl Med. https://doi.org/10.21037/ATM.2016.12.40

Pesole G, Liuni S (1999) Internet resources for the functional analysis of 5′ and 3′ untranslated regions of eukaryotic mRNAs. Trends Genet 15:378. https://doi.org/10.1016/S0168-9525(99)01795-3

Pesole G, Mignone F, Gissi C et al (2001) Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276:73–81

Pesole G, Liuni S, Grillo G et al (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5′ and 3′ untranslated regions of eukaryotic mRNAs. Update 2002. Nucleic Acids Res 30:335–340

Pronk S, Páll S, Schulz R et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854. https://doi.org/10.1093/BIOINFORMATICS/BTT055

Radivojac P, Vacic V, Haynes C et al (2010) Identification, analysis, and prediction of protein ubiquitination sites. Proteins Struct Funct Bioinform 78:365–380. https://doi.org/10.1002/prot.22555

Riley T, Sontag E, Chen P, Levine A (2008) Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol 9:402–412. https://doi.org/10.1038/nrm2395

Sequence search using InterProScan < InterPro < EMBL-EBI. http://www.ebi.ac.uk/interpro/search/sequence-search. Accessed 21 May 2018

Sherry ST, Ward MH, Kholodov M et al (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311

Singh A, Thakur M, Singh SK, Sharma LK, Chandra K (2020) Exploring the effect of nsSNPs in human YPEL3 gene in cellular senescence. Sci Rep 10(1):1. https://doi.org/10.1038/s41598-020-72333-8

Smith EP, Boyd J, Frank GR et al (1994) Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 331:1056–1061. https://doi.org/10.1056/NEJM199410203311604

Susmi TF, Rahman A, Khan MMR et al (2021) Prognostic and clinicopathological insights of phosphodiesterase 9A gene as novel biomarker in human colorectal cancer. BMC Cancer 21:1–18. https://doi.org/10.1186/S12885-021-08332-3/FIGURES/12

Takahashi T, Nau MM, Chiba I et al (1989) p53: a frequent target for genetic abnormalities in lung cancer. Science 246:491–494

Tang Z, Kang B, Li C et al (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 47:W556. https://doi.org/10.1093/NAR/GKZ430

UniProt Consortium (2007) The Universal Protein Resource (UniProt). Nucleic Acids Res 35:D193–D197. https://doi.org/10.1093/nar/gkl929

Venselaar H, te Beek TA, Kuipers RK et al (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinform 11:548. https://doi.org/10.1186/1471-2105-11-548

Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. Nature 408:307–310. https://doi.org/10.1038/35042675

Walker KK, Levine AJ (1996) Identification of a novel p53 functional domain that is necessary for efficient growth suppression. Proc Natl Acad Sci USA 93:15335–15340. https://doi.org/10.1073/PNAS.93.26.15335

Wang X, Sun Q (2017) TP53 mutations, expression and interaction networks in human cancers. Oncotarget 8:624. https://doi.org/10.18632/ONCOTARGET.13483

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309. https://doi.org/10.1093/nar/gki524

Zhang D, Chen C-F, Zhao B-B et al (2013) A novel antibody humanization method based on epitopes scanning and molecular dynamics simulation. PLoS ONE 8:e80636. https://doi.org/10.1371/journal.pone.0080636

Zhang C, Liu J, Xu D, Zhang T, Hu W, Feng Z (2020) Gain-of-function mutant p53 in cancer progression and therapy. J Mol Cell Biol 12(9):674–687. https://doi.org/10.1093/jmcb/mjaa040

Zhu G, Pan C, Bei JX et al (2020) Mutant p53 in cancer progression and targeted therapies. Front Oncol 10:2418. https://doi.org/10.3389/FONC.2020.595187/BIBTEX

Publisher's Note