

Applications of computational algorithm tools to identify functional SNPs

C. George Priya Doss · C. Sudandiradoss ·
R. Rajasekaran · Parikshit Choudhury ·
Priyanka Sinha · Pragnya Hota · Udit Prakash Batra ·
Sethumadhavan Rao

Received: 27 March 2008 / Revised: 8 April 2008 / Accepted: 19 April 2008 / Published online: 19 June 2008
© Springer-Verlag 2008

Abstract Single nucleotide polymorphisms (SNPs) are the most common type of genetic variations in humans. Understanding the functions of SNPs can greatly help to understand the genetics of the human phenotype variation and especially the genetic basis of human complex diseases. The method to identify functional SNPs from a pool, containing both functional and neutral SNPs is challenging by experimental protocols. To explore possible relationships between genetic mutation and phenotypic variation, different computational algorithm tools like Sorting Intolerant from Tolerant, Polymorphism Phenotyping, UTRscan, FASTSNP, and PupaSuite were used for prioritization of high-risk SNPs in coding region (exonic nonsynonymous SNPs) and noncoding regions (intronic and exonic 5' and 3'-untranslated region (UTR) SNPs). In this work, we have analyzed the SNPs that can alter the expression and function of transcriptional factor TP53 as a pipeline and for providing a guide to experimental work. We identified the possible mutations and proposed modeled structure for the mutant proteins and compared them with the native protein. These nsSNPs play a critical role in cancer association studies aiming to explain the disparity in cancer treatment responses as well as to improve the effectiveness of the cancer treatments. Our results endorse the study with in vivo experimental protocols.

Keywords TP53 gene · SNPs · Modeled structure

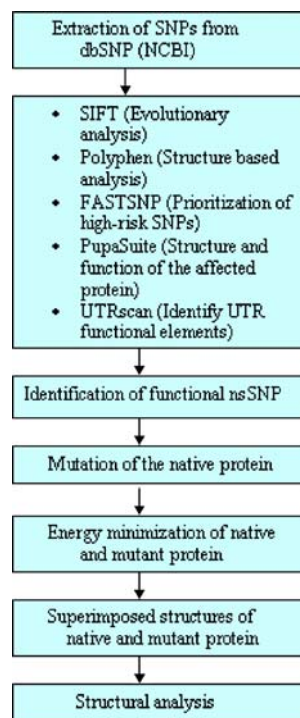
Introduction

Single nucleotide polymorphisms (SNPs) are the simplest and most frequent type of DNA sequence variation among individuals. Human genome is composed of more than three billion base pairs and ten million SNPs (Cargill et al. 1999). SNPs are not only important as markers for constructing genetic maps but also have potential as direct functional polymorphic variants involved within common and genetically complex diseases as well as drug response because (1) SNPs within the coding regions (cSNPs) of functional genes introduce biological variations directly into the gene products through the creation of missense substitutions or premature termination codon, (2) SNPs present in noncoding regions have effects on gene expression by affecting regulatory elements and (3) some intronic SNPs activate cryptic splice sites, leading to alternative splicing (Goto et al. 2001). SNPs in coding and regulatory regions may be implicated in disease themselves. Nonsynonymous SNPs that lead to an amino acid change in the protein product are of major interest because amino acid substitutions currently account for approximately half of the known gene lesions responsible for human inherited disease (Krawczak et al. 2000).

Several different SNP assays have been applied for SNP prioritization experimentally, most involving target amplification following allelic discrimination and a detection step (Chen and Sullivan 2003; Kwock 2003). Though experimental-based approach will provide the strongest evidence for the functional role of a genetic variant, these studies are difficult for characterizing all human genetic variants and also their results might not always reflect in vivo genotype function in humans. On the other hand, computational algorithms are a high priority for characterizing variants because they have the ability to be employed on a scale that

C. George Priya Doss · C. Sudandiradoss · R. Rajasekaran ·
P. Choudhury · P. Sinha · P. Hota · U. P. Batra · S. Rao (✉)
School of Biotechnology, Chemical and Biomedical Engineering,
Bioinformatics Division, Vellore Institute of Technology,
Vellore 632014 Tamil Nadu, India
e-mail: rsethumadhavan@vit.ac.in, raosethu@bsnl.in

Fig. 1 Flow chart for computational analysis of functional SNPs



is consistent with the large number of variants being identified in systematic screening of representatives of human populations for variation. Computational analysis undertaken for an *in silico* investigation of nsSNPs in *TP53* gene is scarce. In this work, we have analyzed the SNPs that can alter the expression and function of transcriptional factor TP53 as a pipeline and for providing a guide to experimental work. We have developed a computational strategy to identify potential and functionally significant SNPs (Fig. 1). We applied different computational algorithms namely Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping (PolyPhen), PupaSuite, UTRscan, and FASTSNP to identify the candidate SNPs that are likely to affect the protein and subsequently the cellular functions. We identified the possible mutations, proposed modeled structures for the mutant proteins, and compared them with the native protein. Our computational study demonstrates the presence of deleterious mutations in *TP53* gene that affect the expression and function of proteins with possible roles in human cancers.

Suites of software for assessing SNPs

Functional analysis of coding nsSNPs by SIFT

SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. SIFT is based on the premise

that protein evolution is correlated with protein function. We used the program SIFT (Pauline and Henikoff 2003) available at <http://blocks.fhcrc.org/sift/SIFT.html> to detect the deleterious coding nonsynonymous SNPs and submitted the query in the form of either SNPids or as protein sequences. The underlying principle of this program is that SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure in the sense that, given a protein sequence, (1) searches for similar sequences, (2) chooses closely related sequences that may share similar function, (3) obtains the multiple alignment of these chosen sequences, and (4) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and those greater than or equal to the cutoff are predicted to be tolerated (Pauline and Henikoff 2001). The cutoff value in SIFT program is tolerance index of ≥ 0.05 . The higher a tolerance index, the less functional impact a particular amino acid substitution is likely to have.

Simulation for functional change in coding nsSNPs

PolyPhen is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a human protein available at <http://coot.embl.de/PolyPhen/>. This prediction is based on straightforward empirical rules which are applied to the sequence, phylogenetic, and structural information characterizing the substitution. Input options for PolyPhen server (Ramensky et al. 2002) is protein sequence or SWALL database ID or accession number together with sequence position with two amino acid variants. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. Basically, PolyPhen searches for 3D protein structures, multiple alignments of homologous sequences, and amino acid contact information in several protein structure databases, calculates position-specific independent counts (PSIC) scores for each of two variants, and then computes the PSIC score difference of two variants. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. A PSIC score difference of 1.5 and above is considered to be damaging.

Analyzing the molecular phenotypic effects of human nonsynonymous SNPs

The SNPeffect (Reumers et al. 2006) and PupaSuite (Conde et al. 2006) are now synchronized to deliver annotations for both noncoding and coding SNP, as well as annotations for

the SwissProt set of human disease mutations. In this approach, the input consists of a list of genes (genes belonging to a given pathway, involved in a particular biological function, etc). The user must specify the type of gene identifiers by selecting either Ensembl or an external database (which include GenBank, Swissprot/TrEMBL, and other gene ids supported by Ensembl). PupaSuite is a unique and more integrated interface of PupaSNP (Conde et al. 2004) and PupasView (Conde et al. 2005) accessible at <http://pupasuite.bioinfo.cipf.es> and through <http://www.pupasnp.org>. PupasView retrieves SNPs that could affect conserved regions that the cellular machinery uses for the correct processing of genes (intron–exon boundaries or exonic splicing enhancers), predicted transcription factor binding sites, and changes in amino acids in the proteins for which a putative pathological effect. PupaSuite finds all the SNPs mapping in locations that might cause a loss of functionality in the genes.

Scanning of noncoding SNPs

Functional significance of each SNP in untranslated region (UTR) was determined by UTRscan (Pesole and Liuni 1999) available at (<http://www.ba.itb.cnr.it/BIG/UTRScan>). UTRresource, which is an internet resource of sequence analysis of 5' and 3' UTR of eukaryotic mRNAs which are involved in many posttranscriptional regulatory pathways that control mRNA localization, stability, and translation efficiency (Sonenberg 1994; Nowak 1994). Briefly, two or three sequences of each UTR SNP that have a different nucleotide at an SNP position were analyzed by UTRscan, which looks for UTR functional elements by searching through user-submitted sequence data for the patterns defined in the UTR site and UTR databases. If different sequences for each UTR SNP are found to have different functional patterns, this UTR SNP is predicted to have functional significance. The internet resources for UTR analysis were UTRdb and UTRsite. UTRdb contains experimentally proven biological activity of functional patterns of UTR sequence from eukaryotic mRNAs (Pesole et al. 2002). The UTRsite has the data collected from UTRdb and also is continuously enriched with new functional patterns.

We used the FASTSNP (Yuan et al. 2006) to identify the polymorphism involving the intron which may lead to defects in RNA and mRNA processing. The FASTSNP server (<http://fastsnp.ibms.sinica.edu.tw>) follows the decision tree principle with external Web service access to TFSearch, which predicts whether a noncoding SNP alters the transcription factor binding site of a gene. The score will be given on the basis of levels of risk with a ranking of 0, 1, 2, 3, 4, or 5. This signifies the levels of no, very low, low, medium, high, and very high effect, respectively.

Modeling nsSNP locations on protein structure and their RMSD difference

Structure analysis was performed for evaluating the structural stability of native and mutant protein. We used the web resource SAAPdb (Cavallo et al. 2005) and dbSNP to identify the protein coded by *TP53* gene (PDB id 1TSR). We also confirmed the mutation positions and the mutation residues from this server. These mutation positions and residues were in complete agreement with the results obtained with SIFT and PolyPhen programs. The mutation was performed by using SWISSPDB viewer and energy minimization for 3D structures was performed by NOMAD-Ref server (Lindahl et al. 2006). This server use Gromacs as default force field for energy minimization based on the methods of steepest descent, conjugate gradient, and L-BFGS methods (Delarue and Dumas 2004). We used conjugate gradient method for optimizing the 3D structures. Deviation between the two structures was evaluated by their RMSD values.

Using the human transcription factor *TP53* as a test case

The human gene *TP53* (tumor suppressor gene) is mutated in more than 50% of human cancers and p53 dysfunction is caused through a direct mutation within the DNA binding domain of the gene (Vogelstein et al. 2000). The p53 tumor protein is essential for regulating cell division; it has been nicknamed the “guardian of the genome.” The gene is located on the short (p) arm human chromosome 17 (17p13.1; Strachan et al. 1999). The human p53 protein consists of 393 amino acids with five evolutionarily conserved domains (I to V). Domains II to V correspond to the DNA binding domain which is the target for p53 mutations. Ninety percent of p53 mutations occur in the central region (101–306) which harbors four of the five highly conserved evolutionary domains and this region is essential for the p53–DNA interaction. The p53 tumor suppressor is a transcription factor that coordinates cellular responses to DNA damage and stress, initiating cell cycle arrest or triggering apoptosis. There are a variety of stresses that have been shown to activate p53 including DNA damage, cell cycle aberrations, hypoxia, and aberrant growth signals resulting from expression of oncogenes (Monica et al. 2007). The spectrum of TP53 mutations in various types of cancers is mainly due to GC→AT transition at the CpG nucleotide [colon, ovary brain, or leukemia] and GC→TA transversion [lung, head and neck, or HCC] (Hollstein et al. 1991). Incidence of p53 mutations is highest in ovarian cancer (48.3%), followed by colorectal cancer (43.6%), esophageal cancer (42.6%), head-and-neck cancer (41.5%), and lung cancer (38.4%; Petitjean et al. 2007). Our survey

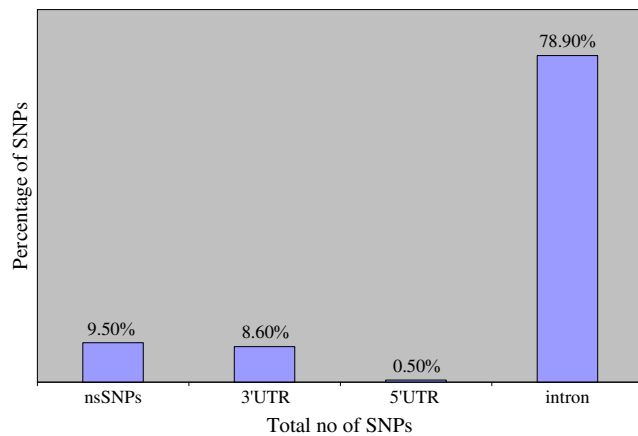


Fig. 2 Distribution of SNPs in nsSNPs, 3' UTR, 5' UTR and intron regions

shows that there is a wide choice of literature in mutations of the *TP53* gene associated with various forms of human cancers.

SNP dataset from dbSNP

SNP dataset for *TP53* gene investigated in this work was retrieved from dbSNP (Sherry et al. 2001) <http://www.ncbi.nlm.nih.gov/SNP/> for our computational analysis. We selected (1) nonsynonymous coding SNPs, (2) 5' and 3' UTR SNPs, and (3) introns for our investigation and distribution of SNPs are shown in (Fig. 2). Out of 209 SNPs, 20 were nonsynonymous SNPs (nsSNPs) and five SNPs in coding synonymous. Noncoding region comprises of one SNP in 5' UTR, 18 SNPs in 3' UTR, and 165 SNPs were in the intron region. Since the numbers of SNPs in 5' UTR are much less as compared to the SNPs in the 3' UTR and in the coding regions, it may be presumed that *TP53* gene shows no functional significance in the 5' UTR.

Deleterious nsSNP by SIFT program

SIFT, a sequence homology-based tool, was used to identify the conservation level of a particular position in a protein. Protein sequences of 20 nsSNPs were submitted independently to SIFT program and 14 nsSNPs (70%) were identified to be deleterious having the tolerance index score of ≤ 0.05 . Thirteen nsSNPs showed a highly deleterious tolerance index score of 0.00 and one nsSNP had a tolerance index score of 0.03, respectively.

Damaged nsSNP by PolyPhen server

The structural levels of alteration were determined by applying PolyPhen program. All the 20 protein sequences of nsSNPs submitted to SIFT were also submitted as input to the PolyPhen server. Fourteen nsSNPs (70%) were considered to be damaging (Table 1) and exhibited a range of PSIC score difference between 1.74 and 3.14. Interestingly, we found significant correlation between SIFT and PolyPhen tools. So, we could infer from the results obtained using the SIFT and PolyPhen tool that these nsSNPs may disrupt both the protein function and structure and mutations occurring in these nsSNPs could be of significant importance in causing of various forms of cancer.

Predicting the effect of coding nonsynonymous SNPs

The PupaSuite server aims to provide a platform for predicting the effect of coding nonsynonymous SNPs on the structure and function of the affected protein. We submitted *TP53* gene in Pupa Suite and specified the type of gene identifiers by selecting either Ensembl or an external database. Out of 20 nsSNPs in *TP53* gene, 11 nsSNPs disrupted the exonic splicing enhancers, three nsSNPs disrupted the exonic splicing

Table 1 List of SNPs that were predicted to be of functional significance by SIFT and PolyPhen

SNPs ID	Nucleotide change	AA change	Tolerance index	PSIC
rs11540654	C/G/T	110(R→P)	0.03	1.740
rs28934873	C/T	133(M→T)	0.00	1.760
rs28934875	C/G	138(A→P)	0.00	2.185
rs28934874	A/C/T	151(P→T)	0.00	2.797
rs28934578	A/G	175(R→H)	0.00	2.468
rs28934573	C/T	241(S→F)	0.00	2.622
rs28934572	A/G	244(G→D)	0.00	2.396
rs28934575	A/G	245(G→S)	0.00	2.171
rs11540652	A/G	248(R→Q)	0.00	2.243
rs28934571	G/T	249(R→S)	0.00	2.693
rs28934577	A/T	257(L→Q)	0.00	2.447
rs28934576	A/G	273(R→H)	0.00	2.468
rs17849781	C/G	278(P→A)	0.00	2.797
rs28934574	C/T	282(R→W)	0.00	3.143

Table 2 List of SNPs that were predicted to be of functional significance by PupaSuite

SNPs ID	Nucleotide change	Region	Functional significance
rs1800371	C/T	Coding nonsynonymous	Exon splicing enhancer
rs1042522	C/G	Coding nonsynonymous	Exon splicing enhancer
rs11540654	C/G/T	Coding nonsynonymous	Cellular processing Functional sites Exon splicing enhancer
rs35163653	A/G	Coding nonsynonymous	Exon splicing silencer
rs28934573	C/T	Coding nonsynonymous	Exon splicing enhancer
rs28934572	A/G	Coding nonsynonymous	Cellular processing
rs28934575	A/G	Coding nonsynonymous	Exon splicing enhancer Cellular processing
rs11540652	A/G	Coding nonsynonymous	Exon splicing enhancer
rs28934571	G/T	Coding nonsynonymous	Exon splicing enhancer
rs28934577	A/T	Coding nonsynonymous	Exon splicing enhancer
rs28934576	A/G	Coding nonsynonymous	Exon splicing enhancer Selective constraints prediction
rs17849781	C/G	Coding nonsynonymous	Exon splicing silencer
rs28934574	C/T	Coding nonsynonymous	Exon splicing enhancer Selective constraints prediction
rs17882252	A/G	Coding nonsynonymous	Functional sites
rs35993958	C/G	Coding nonsynonymous	Exon splicing silencer
rs17881470	G/T	Coding nonsynonymous	Cellular processing Exon splicing enhancer
rs17881366	A/G	mRNA	Exon splicing enhancer
rs1794291	A/G	mRNA	Exon splicing enhancer
rs1794292	C/G	mRNA	Exon splicing enhancer
rs1794293	C/G	mRNA	Exon splicing enhancer
rs17884410	A/G	mRNA	Exon splicing silencer
rs16956880	C/T	mRNA	Exon splicing silencer
rs34486624	A/G	mRNA	Exon splicing silencer
rs916132	C/T	mRNA	Exon splicing silencer
rs11575997	G/T	Intron	Intron–exon junction

silencers, and four nsSNPs (pathological SNPs) were involved in cellular processing and two nsSNPs in functional sites and two nsSNPs showed selective constraints prediction. In non-coding region of *TP53* gene, out of eight SNPs in mRNA regions, four SNPs disrupted the exon splicing enhancer, four SNPs disrupted the exon splicing silencer. The results of nsSNPs obtained by PupaSuite (Table 2) correlates with the nsSNPs affected by SIFT and PolyPhen tools.

Functional SNPs in noncoding SNPs

Polymorphism in the 3' UTR affects the gene expression by affecting the ribosomal translation of mRNA or by influenc-

ing the RNA half-life (Deventer 2000). Among 18 SNPs in 3' UTR, four SNPs were related to the functional pattern change of IRES by UTRscan (Table 3). Internal ribosome entry site (IRES) is bound by internal mRNA ribosome. It is an alternative mechanism of translation initiation compared to the conventional 50-cap dependent ribosome scanning mechanism (Becky and Anne 2005). Out of 165 SNPs in intron region of *TP53* gene (Table 4), 73 SNPs were predicted to be functionally significance, with a risk ranking of 1–2 (71 SNPs) and 3–4 (2 SNPs), respectively, by FASTSNP. SNP with an id (rs11575997) present in intron region showed functionally significance in both FASTSNP and PupaSuite. Similarly, SNPs, namely with id rs17881366

Table 3 List of SNPs (mRNA) that were predicted to be of functional significance by UTR scan

SNPs ID	Nucleotide change	UTR Position	Functional element change
rs35919705	C/T	3'UTR	IRES→ no pattern
rs17884306	A/G	3'UTR	IRES→ no pattern
rs17881366	A/G	3'UTR	IRES→ no pattern
rs16956880	C/T	3'UTR	IRES→ no pattern

Table 4 List of SNPs (intron) that were predicted to be of functional significance by FASTSNP

SNPs ID	Nucleotide change	Level of risk	Possible functional effect
rs34361146	C/T	3–4	Splicing site
rs35117667	C/T	3–4	Intronic enhancer
rs17886820	A/G	1–2	Intronic enhancer
rs2856754	A/T	1–2	Intronic enhancer
rs2543550	A/T	1–2	Intronic enhancer
rs17883852	C/G/T	1–2	Intronic enhancer
rs12938753	A/C	1–2	Intronic enhancer
rs36118870	A/G	1–2	Intronic enhancer
rs17879377	A/G	1–2	Intronic enhancer
rs1641549	A/G	1–2	Intronic enhancer
rs1642794	G/T	1–2	Intronic enhancer
rs12949655	A/G	1–2	Intronic enhancer
rs1642792	C/T	1–2	Intronic enhancer
rs1642791	C/T	1–2	Intronic enhancer
rs1642790	C/T	1–2	Intronic enhancer
rs1642789	A/T	1–2	Intronic enhancer
rs11575997	G/T	1–2	Intronic enhancer
rs11653656	A/G	1–2	Intronic enhancer
rs1794288	C/T	1–2	Intronic enhancer
rs1642788	C/T	1–2	Intronic enhancer
rs1642787	C/T	1–2	Intronic enhancer
rs12951053	A/C	1–2	Intronic enhancer
rs8069054	A/C	1–2	Intronic enhancer
rs1625895	A/G	1–2	Intronic enhancer
rs34949160	A/G	1–2	Intronic enhancer
rs2856920	C/G	1–2	Intronic enhancer
rs1794287	C/T	1–2	Intronic enhancer
rs34686922	C/T	1–2	Intronic enhancer
rs17886250	A/G	1–2	Intronic enhancer
rs9905505	A/G	1–2	Intronic enhancer
rs17881524	A/C	1–2	Intronic enhancer
rs1642775	C/T	1–2	Intronic enhancer
rs1794296	C/G	1–2	Intronic enhancer
rs1642776	A/C	1–2	Intronic enhancer
rs1642777	G/T	1–2	Intronic enhancer
rs17882227	C/T	1–2	Intronic enhancer
rs35898718	A/T	1–2	Intronic enhancer
rs35750430	A/T	1–2	Intronic enhancer
rs35162515	A/T	1–2	Intronic enhancer
rs9903378	A/C	1–2	Intronic enhancer
rs2909421	A/T	1–2	Intronic enhancer
rs17881035	A/G	1–2	Intronic enhancer
rs11656607	C/T	1–2	Intronic enhancer
rs17880340	C/T	1–2	Intronic enhancer
rs1794298	A/G	1–2	Intronic enhancer
rs17886213	A/T	1–2	Intronic enhancer
rs17883925	C/G	1–2	Intronic enhancer
rs17884159	C/T	1–2	Intronic enhancer
rs8068551	C/T	1–2	Intronic enhancer
rs35175270	A/G	1–2	Intronic enhancer
rs2909425	A/C	1–2	Intronic enhancer
rs11652704	C/T	1–2	Intronic enhancer
rs722494	C/G	1–2	Intronic enhancer
rs34766441	C/T	1–2	Intronic enhancer

Table 4 (continued)

SNPs ID	Nucleotide change	Level of risk	Possible functional effect
rs2909427	C/T	1–2	Intronic enhancer
rs3021071	G/T	1–2	Intronic enhancer
rs17880541	A/G	1–2	Intronic enhancer
rs12603869	A/G	1–2	Intronic enhancer
rs2909429	C/T	1–2	Intronic enhancer
rs17886508	C/T	1–2	Intronic enhancer
rs4407108	G/T	1–2	Intronic enhancer
rs4411556	A/G	1–2	Intronic enhancer
rs2543544	A/T	1–2	Intronic enhancer
rs2543545	A/T	1–2	Intronic enhancer
rs8078476	A/G	1–2	Intronic enhancer
rs1642781	A/G	1–2	Intronic enhancer
rs9897559	C/T	1–2	Intronic enhancer
rs34130474	C/G	1–2	Intronic enhancer
rs4380105	C/T	1–2	Intronic enhancer
rs9894227	A/G	1–2	Intronic enhancer
rs2856751	A/T	1–2	Intronic enhancer
rs2543548	C/T	1–2	Intronic enhancer

and rs16956880, in mRNA region showed functional significance in both UTRscan and PupaSuite. SNPs with id rs1625895 and 12951053 present in intronic region showed risk of 1–2 by FASTSNP, which is in good correlation with experimental studies (Brian et al. 2007).

Modeling of mutant structure

Mapping the deleterious nsSNPs into protein structure information was obtained from dbSNP and SAAPdb. The available structure for *TP53* gene is reported to be having a PDB id (1TSR).

According to this resource, the mutations mainly occurred for 1TSR with 14 SNP rs ids, namely, rs11540654, rs28934873, rs28934875, rs28934874, rs28934578, rs28934573, rs28934572, rs28934575, rs11540652, rs28934571, rs28934577, rs28934576, rs17849781, and rs28934574. The mutation for 1TSR at the corresponding positions were performed by SWISSPDB viewer independently to get modeled structures. Then, energy minimizations were performed by NOMAD-Ref server for the native-type protein (1TSR) and the mutant-type structures.

From the modeled structures, RMSD values between the native and mutant amino acid residues were calculated (Table 5) and exhibited a range from 0.14 to 2.14 Å, respectively. It can be seen from the Table 5 that total energy for the native-type structure (1TSR) after energy minimization is found to be 13,161.882 kcal/mol and the mutant-type structures (1TSR) exhibited a minimization energy range of –12,792.573 to –13,157.354 kcal/mol,

Table 5 RMSD of native-type protein (1TSR) and mutant modeled proteins

SNPs ID	AA change	Total energy after energy minimization (kcal/mol)	RMSD between native and mutant AA (Å)
rs11540654	110(R→P)	-12,859.432	0.27
rs28934873	133(M→T)	-13,078.386	0.91
rs28934875	138(A→P)	-13,066.606	1.07
rs28934874	151(P→T)	-13,117.384	1.97
rs28934578	175(R→H)	-12872.659	1.17
rs28934573	241(S→F)	-13,039.299	0.14
rs28934572	244(G→D)	-13,128.101	0.51
rs28934575	245(G→S)	-13,157.354	2.14
rs11540652	248(R→Q)	-13,106.054	0.24
rs28934571	249(R→S)	-12,792.573	0.97
rs28934577	257(L→Q)	-13,102.755	0.39
rs28934576	273(R→H)	-12,907.404	0.43
rs17849781	278(P→A)	-13,102.737	0.65
rs28934574	282(R→W)	-12,888.564	0.66

Total energy after energy minimization of native protein (1TSR)
-13,161.882 kcal/mol

respectively. The total energy after energy minimization is almost similar or high for the mutant-type structures as compared to the native-type structure (1TSR). The highest and least RMSD values of superimposed structures of the native protein (1TSR) with mutant-type proteins 245(G→S) and 241(S→F) are 2.14 and 0.14 Å, respectively, and shown in (Figs. 3 and 4). All the deleterious nsSNPs which were mapped by SAAPdb and dbSNP in the native protein (1TSR) of *TP53* gene should be considered functionally significant based on both sequence homology (SIFT), structural homology (PolyPhen), and PupaSuite for causing various forms of cancer. Our results obtained through

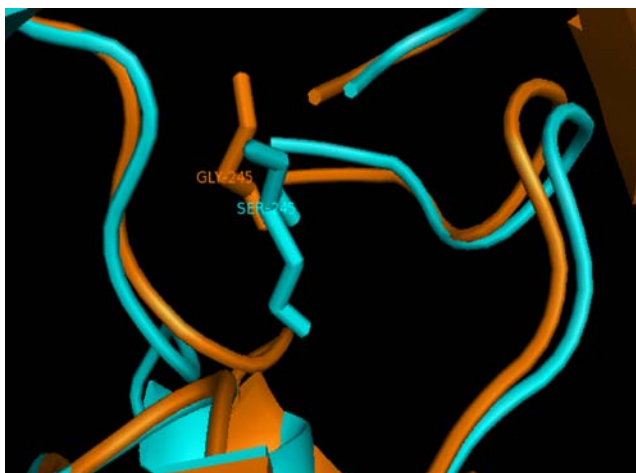


Fig. 3 Superimposed structure of native protein 1TSR (orange) with mutant protein (cyan) 245(G→S)

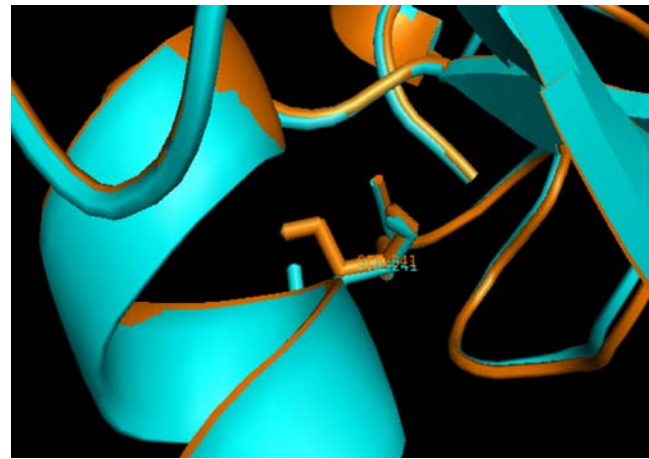


Fig. 4 Superimposed structure of native protein 1TSR (orange) with mutant protein (cyan) 241(S→F)

computational algorithms are in good correlation with the experimental results (Soussi et al. 2006).

Conclusions

Given that hundreds of thousands of SNPs are estimated to exist in the human population. However, only a small subset of variants that affect the phenotype will confer to disease risk. The effect of many nsSNPs will probably be neutral as natural selection will have removed mutations on essential positions. Assessment of nonneutral SNPs is mainly based on phylogenetic information (i.e., correlation with residue conservation) extended to a certain degree with structural approaches (PolyPhen). Much attention has been focused on modeling by different methods the possible phenotypic effect of SNPs that cause amino acid changes and only recently has interest focused on functional SNPs affecting regulatory regions or the splicing process. However, there is increasing evidence that many human disease genes are the result of exonic or noncoding mutations affecting regulatory regions. Study of the molecular basis of diseases by experimental methods is laborious and time-consuming, and at the structural level often nearly impossible, especially in cases where there are several missense mutations causing the disease. By contrast, precise and useful information about the effects of mutations on protein structure and function can be readily obtained by theoretical methods.

Out of 20 nsSNPs in *TP53* gene, 14 of them were found to be deleterious (SIFT) and damaging (PolyPhen); 16 nsSNPs and eight SNPs in mRNA showed molecular phenotypic variations by PupaSuite. Four SNPs in the 3' UTR were found to be functionally significant by UTRScan and 73 SNPs in the intronic region were found to be functionally significant by FASTSNP. Our results from this

study suggests that the application of computational algorithms namely SIFT, PolyPhen, PupaSuite, UTRresource, and FASTSNP might provide an alternative approach to select target SNPs by understanding the effect of SNPs on the functional attributes or “molecular phenotype” of a protein. Some detailed laboratory work has been published already, but our analysis reveals that further structural and functional information can be derived using computer-assisted methods. The models built in this work would be applicable for predicting the deleterious SNPs and their functions in gene regulation which would be helpful for further genotype–phenotype researches as well as the pharmacogenetics study.

Acknowledgments The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this work. All the authors thank the Editor-in-Chief Prof. Rudi Appels for his invaluable suggestions during the preparation of this compact review.

References

- Becky MP, Anne EW (2005) The implications of structured 5' untranslated regions on translation and disease. *Cell Dev Biol* 16:39–47
- Brian Sprague L, Dietz AT, Closas MG, Newcomb AP, Ernstoff LT, Hampton JM, Chanock JS, Haines LJ, Egan MK (2007) Genetic variation in TP53 and risk of breast cancer in a population-based case-control study. *Carcinogenesis* 28(8): 680–1686
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238
- Cavallo A, Martin AC (2005) Mapping SNPs to protein sequence and structure data. *Bioinformatics* 8:1443–1450
- Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* 3:77–96
- Conde L, Vaquerizas MJ, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, Robledo M, Dopazo J (2004) PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res* 32:W242–W248
- Conde L, Vaquerizas JM, Ferrer-Costa C, de la Cruz X, Orozco M, Dopazo J (2005) PupasView: a visual tool for selecting suitable SNPs, with putative pathological effect in genes, for genotyping purposes. *Nucleic Acids Res* 33:W501–W505
- Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J (2006) PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 34:W621–W625
- Delarue M, Dumas P (2004) On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci* 101:6957–6962
- Deventer VS (2000) Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med* 26:S98–S102
- Goto Y, Yue L, Yokoi A, Nishimura R, Uehara T, Koizumi S, Saikawa Y (2001) A novel single-nucleotide polymorphism in the 3'-untranslated region of the human dihydrofolate reductase gene with enhanced expression. *Clin Cancer Res* 7:1952–1956
- Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991) p53 mutations in human cancers. *Science* 253:49–53
- Horvath MM, Wang X, Resnick MA, Bel DA (2007) Divergent evolution of human p53 binding sites: cell cycle versus apoptosis. *PLoS Genet* 3(7):e127
- Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, Strachan T, Read AP, Cooper DN (2000) Human gene mutation database—a biomedical information and research resource. *Hum Mutat* 15:45–51
- Kwock PY (2003) Single nucleotide polymorphisms methods and protocols (Methods in molecular biology), vol. 212. Humana Press, Totowa
- Lindahl E, Azuara C, Koehl P, Delarue M (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res* 34:W52–W56
- Nowak R (1994) Mining treasures from ‘junk DNA’. *Science* 263:608–610
- Pauline Ng C, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874
- Pauline Ng C, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. *Nucl Acids Res* 31:3812–3814
- Pesole G, Liuni S (1999) Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNA. *TIG* 15:378
- Pesole G, Liuni S, Grillo G, Licciulli F, Mignone F, Gissi C, Saccone C (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 30:335–340
- Petitjean E, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 28(6):622–629
- Ramensky V, Pork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30(17):3894–3900
- Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F (2006) SNPeff v2.0: a new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. *Bioinformatics* 22(17):2183–2185
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. *Nucl Acids Res* 29:308–311
- Sonenberg N (1994) mRNA translation: influence of the 5' and 3' untranslated regions. *Curr Opin Genet Dev* 4:310–315
- Soussi T, Asselain B, Hamroun D, Kato S, Ishioka C, Claustres M, Bérout C (2006) Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. *Clin Cancer Res* 12:62–69
- Strachan T, Read AP (1999) Human molecular genetics, 2nd edn. Bios Scientific, Oxford Ch. 18 Cancer Genet
- Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. *Nature* 408:307–310
- Yuan H, Chiou J, Tseng W, Liu C, Liu C, Lin Y, Wang H, Yao A, Chen Y, Hsu C (2006) FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 34:W635–W641