Biomedical Letters ISSN 2410-955X DOI: https://doi.org/10.47262/BL/7.2.20210705

Research article
2021 | Volume 7 | issue 2 | Pages 141-154

ARTICLE INFO

Received
August 05, 2021
Revised
October 09, 2021
Accepted
November 21, 2021

***Corresponding Author**
Farkhanda Yasmin

E-mail
farkhanda.yasmin@kfueit.edu.pk

Keywords
Damaging
Insilico
Mutant
Evolutionary relationship
Polymorphisms

How to Cite
Baksh M, Iqbal MS, Yasmin F, Suleman S, Rasheed M, Farooq W, Javed T. Computational detection, analysis and interpretations of genomic variants in human diseases-associated GENEMDM 2. Biomedical Letters 2021; 7(2):141-154.

Scan QR code to see this publication on your mobile device.

Open Access

Computational detection, analysis and interpretations of genomic variants in human diseases-associated GENEMDM 2

Muqadas Baksh[1], Muhammad Sarfaraz Iqbal[1,2], Farkhanda Yasmin[3]*, Saiema Suleman[4], Majeeda Rasheed[5], Waqas Farooq[6], Talha Javed[7]

[1]Department of Bioinformatics &Computational biology, Virtual University of Pakistan
[2]Department of Bioinformatics & Biotechnology, GC University Faisalabad, Pakistan
[3]Department of Biosciences, Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan
[4]Institute of Molecular Biology & Biotechnology, University of Lahore, Lahore
[5]Department of Life Sciences, Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan
[6]Department of Biochemistry, University of Agriculture Faisalabad, Faisalabad, Pakistan
[7]Department of Biotechnology & Bioinformatics, Virtual University of Pakistan

Abstract

Most of the mutations described in human MDM2 are tolerated without significantly disrupting the corresponding structural or molecular function. However, some of them are associated with a variety of human diseases, including cancer. Numerous computational methods have been developed to predict the effects of missense single nucleotide variants (SNVs). The non-synonymous single nucleotide polymorphisms affect the function of XRCC1, which impairs the ability to repair DNA and therefore increases the risk of diseases such as cancer. In this study, sequence and structure-based computational tools were used to screen the total listed coding SNPs of the MDM2 gene in order to recognize and describe them. The potential 6 ns SNP of MDM2 were identified from 29 ns SNP by consistent analysis using computational tools PolyPhen 2, SIFT, PANTHER and cSNP. The computational methods were used to systematically classify functional mutations in the regulatory and coding regions that modify the expression and function of the MDM2 enzyme. The HOPE project also made it possible to elaborate the structural effects of the substitutions of amino acids. *In silico* analysis predicted that rs759244097 is harmful. This study concluded that identifying this SNP will help to determine an individual's cancer susceptibility, prognosis and further treatment. Furthermore, current high-throughput sequencing efforts and the need for extensive interpretation of protein sequence variants requires more efficient and accurate computational methods in the coming years.

Introduction

Single nucleotide polymorphisms (SNP) are source variants in a genome. SNP is an only basic mutation of DNA. SNPs are the least complex structure and most common wellspring of hereditary polymorphism in the human. There are several publicly accessible databases for SNPs, such as dbSNP, GWAS Central and Swiss Var. It is particularly important that only non-synonymous SNPs (nsSNPs), also known as missense variants, cause the translated amino acid residue sequences to be altered. NsSNP plays an essential role in the functional diversity of proteins encoded in the human population and can be associated with many diseases. nsSNPs can affect protein function by decreasing protein solubility or destabilizing protein structure, and by altering transcription and translation can affect gene regulation [1, 2].

The most notable cell-binding partner of p53 is the product of the MDM2 gene, which can lead to p53 accumulation and which controls p53 activity in a self-regulating feedback loop[3]. In numerous tumours, overexpression of MDM2 (mainly as a result of gene amplification) has been shown to simultaneously inactivate and stabilize p53[4, 5]. In normal cells, p53 protein is constitutively stated at very low, generally imperceptible levels. In response to stimulations such as DNA damage, levels rise sharply and p53 ultimately mediates its function as a growth inhibitor or determine cell apoptosis[6].

In MDM2, single-nucleotide polymorphisms have been found that control MDM2 levels and effect p53 action, adding to tumour threat [7]. These changes affect basal stages of MDM2 and may modify p53 mono and poly ubiquitination. Human MDM2 is found on chromosome number 12 with an exact position in 12q15 and it has 491-amino acid long chain that binds p53 with hydrophobic pocket (aa 25-100), a significant region with nuclear localization (aa 179-185) and nuclear export signals (aa 179-185), acidic domain (aa 243-301) and in the C-terminus RING domain (aa 432-491)[8].

MDM2 targets six major lysine amino acids on p53 for ubiquitination which are K370, K372, K373, K381, K382, and K386. In vitro investigation demonstrates that alternation of these lysine residues reduces p53 ubiquitination, expanding its movement. In any case, knock-in mice with lysine to-arginine changes that forestall ubiquitination yet don't alter charge at these six deposits (p53-6KR) don't show essentially expanded p53 levels[9]. Thyroid nodules were routinely assessed by fine needle aspiration (FNA) biopsy and ultrasonography. Several thyroid nodules were not inhibited by thyroxin (T4) and only selected nodules were surgically treated. FNA biopsy is appropriate, simple and reliable. FNA biopsy is safe without technical hitches and is not associated with sowing tumour cells in the needle tube, but FNA biopsy is sensitive and delicate. Hormonal factors also have a major role in the disease expression, thyroid stimulating hormone (TSH) large expression can lead to disease progression and in that situation thyroid cancer patients receive TSH inhibitory L-T4 treatment[10].Apart from ionizing radiations, there are many other risk factors responsible for thyroid cancer include genetic factors, environmental influences and access to a medical car are the major factors of variability in the thyroid cancer incidences in different areas. Overexpression of MDM2 was observed in a variety of human tumours at a frequency of 7% including sarcomas, gliomas, harmful Schwannomas, and leukaemia. MDM2 is fundamentally an E3 ubiquitin ligase target p53 for proteasome degradation after ubiquitination[11]. Mutations associated with p53 often occur in human cancers and MDM2 amplification usually preserves wild-type[12, 13].

In this study, the comprehensive analysis of genetic variation was predicted to result in altered protein function ("rs759244097") in 806 drug-related genes including 628 drug targets (163 targeted by cancer therapeutics). further describe how this may affect the likelihood of 1236 FDA-approved drugs being affected by functional variants in their targets and how this likelihood varies between different populations. Even though variants in non-coding regions, copy number alterations, and chromosomal structural changes as well as epigenetics may further contribute to drug PK and PD variability [21], such alterations were not part of this study.

Materials and Methods

Data mining

Dataset of nsSNPs was Retrieved from dbSNP. Associated information about the SNP was recovered from the Swiss-Prot and PubMed data sets. Protein Data Bank (PDB) was used for acquiring structural information of MDM2. The SNP information just like accession number of protein and SNP of the gene MDM2 was getting from the NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp/) and Swiss-Prot datasets (https://expasy.org/).

Prediction of functional consequences of non-synonymous coding SNPs

Various computational tools that can recognize nsSNP are damaging or neutral, changes in protein structure and stability are based on a single point of alteration. There are utilized SIFT and PANTHER, which work evolutionary sequences relationship to find the intolerance or tolerance and another hand PolyPhen and I-Mutant used structural and functional parts of the protein (**Fig. 1**).

SIFT

SIFT is expected to manage future analyses and not proposed for direct use in a clinical setting, on the grounds that in silico forecasts are not a substitute for research facility tests. SIFT is helpful for analysts who are keen on researching the impacts of transformations on protein functions[14].

SIFT is a sequence homology-based tool that categories intolerant and tolerant amino acid substitutions and predicts an amino acid substitution in a protein that will have a phenotypic impact. SIFT depend on the premise that protein development is associated with protein function. Location significant for a function should be preserved in an arrangement of the protein family, though insignificant positions ought to seem differing in an arrangement.

SIFT takes a query sequence and used various arrangement information to predict tolerated and damaging substitutions for each position of the query sequence. SIFT has a multistep procedure that (1) scan for parallel sequences, (2) picks firmly related arrangements that may have the comparable capacity to the question succession, 3) gets the arrangement of these picked arrangements, and (4) ascertains standardized probabilities for every single possible substitution from the arrangement. Positions with standardized probabilities under 0.05 are anticipated to be damaging, those more prominent than or equivalent to 0.05 are anticipated to be tolerated [15].

PolyPhen-2

PolyPhen-2 is an automatic tool that can predict the possible effect on the structure and capacity of a human protein of an amino acid substitution. This type of automated prediction is necessary to interpret large data sets of occasional genetic variants, which have numerous applications of the latest technology in human genetics research. Used in the latest research consist of detecting uncommon alleles which caused Mendelian disease [16], checking for possibly therapeutically noteworthy alleles in an individual's genome [17], and the profile of the spectrum of sparse mutations by deep sequencing of large populations Ring [18]. This prediction is based on the characterization alternatives of many sequences, phylogeny, and structural features.PolyPhen-2 extracts numerous sequence and structure-based features of the substitution site and provides them to a probabilistic classifier, forgiven amino acid substitution in a protein [19].

PANTHER

Protein analysis through evolutionary relationships of the PANTHER arrangement system (http://www.pantherdb.org/tools/csnpScore), with which the functional nsSNPs in the MDM2 gene were described using statistical modelling of relationships based on HMM and evolutionary of the protein family [20]. PANTHER can predict the functional sizes of the protein variants. He agrees with the order of the protein with a family of evolutionary proteins and found the possibility that the change in nsSNP in the protein's function is harmful or not. For the functioning of the PANTHER cSNP tool, the entry was the replacement of amino acids and the protein sequence.

Position-specific evolutionary preservations which measure the distance of time at which the position of current protein has been protected by following back to its reproduce direct ancestors. This is the interpretation of the PSEP score. Probably damaging, in which substitution is very likely to interrupt protein function and the preservation time is greater than 450my. Possibly damaging, in which substitution may interrupt protein function and the time of preservation is between 200my and 400my. Probably time, in which the substitution is more likely not to interrupt protein function and the preservation is smaller than 200my.

I-Mutant

I-Mutant2.0 is a Support Vector Machine, and it is based web server used for automatically predicting protein stability, which can change single point mutations. This tool prepared on an informational index which got from ProThem which is currently the very inclusive database of protein mutations experiments. This predictor can assess the stability alteration on single point mutation opening with the structure of protein and sequence of a protein. I-

Mutant2.0 be able to predict the direction changing of the free energy and also its value. It can evaluate the free energy changes value just like difference amid wild type of protein and mutant protein (DG mutant-DG wild type). Its demonstrations the amino acid of Wild type Protein (WT), New residue after Mutation (NAW), Temperature (T) and the PH. It also can predict the potential in free energy change rate, laterally through a reliability index (RI) , zero as the lowest and 10 as highest RI [20].

Project HOPE

HOPE project is used to identify the structural effect of definite point mutations in the protein sequence of MDM2 (http://www.cmbi.ru.nl/hope/). To get the tertiary structure of this protein and construct a homology model, firstly a BLAST was done through algorithm against UniProt and PDB. Then used the Distributed Annotation System server to predict protein function[21].

Clustal Omega

Tool Clustal Omega used to create evolutionary relations and biologically evocative multiple sequence alignments for different portions. The input of this tool was the FASTA sequence of Homo sapiens, Musmus culus, Canis lupus familiarise and cricetulus griseus proteins with UniProtKB/SwissProt format to recover Phylograms or Cladogram. The preservation output has "*" or asterisk for point to the position which means the
residue is fully conserved and a single residue ":" Or indicate conservation between groups with very similar characteristics and "." () to specify Conservation between weakly similar attribute groups [22, 23].

STRING

Protein to protein inter-linkages are important to assess the functional interactions of all proteins present in the cells. These protein-to-protein inter-linkages are crucial for the evaluation of all these interactions which are present in cells. STRING is a tool used for prediction interactions of the protein with different protein in the cell. By using its 5,214,234 protein databases of 1,113 species, STRING produces protein-protein interactions through direct or indirect relations amid recognized protein and other proteins [24].

Chimera

UCSF Chimera is a highly scalable program for the analysis and interactive visualization of molecular structures and related data, conformational sets, supramolecular sets, fitting results, trajectories for sequence alignments and including density maps. It can be generated. High quality animations and images [25]. Chimera software produced by the University of California; San Francisco used this tool to create the mutant protein model. The result of this tool is a mutation for the display of a graphic model (https://www.cgl.ucsf.edu/chimera/).

Results

SNP data set

Total recorded SNPs in the human MDM2 gene sequence were 6978 genes. All stated SNPs of the MDM2 gene have been taken from thdbSNP in NCBI and Ensemble Browser. Arrangement of SNPs is dependent on their functional classes which are described in (**Table 1**).

Table 1: Functional significance kinds of SNPs present in MDM2 gene. Data derived from Ensembl.

Consequence type	Count
Coding sequence variant	2863
Frameshift variant	72
Inframe deletion	69
Inframe insertion	9
Missense variant	2056
Missense variant _ splice region variant	802
Start retained variants	4
Stop retained variants	4
Start lost	15
Stop gained	7
Synonymous variant	1077
Grand Total	**6978**

On the other hand, estimation of missense affect MDM2 functions is essential. The summary of these results is given in following (**Table 2**).

Table 2: The detailed Summary of estimation of missense effect of MDM2 protein functions after nsSNPs. Data derived from SIFT, Polyphen-2 and PANTHER.

Calculation	Numerals of nsSNPs (%)	
Poly-Phen2	SIFT	PANTHER
Benign	15 (52%)	
Certainly damaging	6 (21%)	
Undoubtedly damaging	8 (27%)	
DAMAGING *Warning! Low confidence	9 (31%)	
DAMGING	14 (48%)	
TOLERATED	6 (21%)	
Possibly damaging		8(28%)
Probably damaging		11(38%)
Probably benign		10(34%)

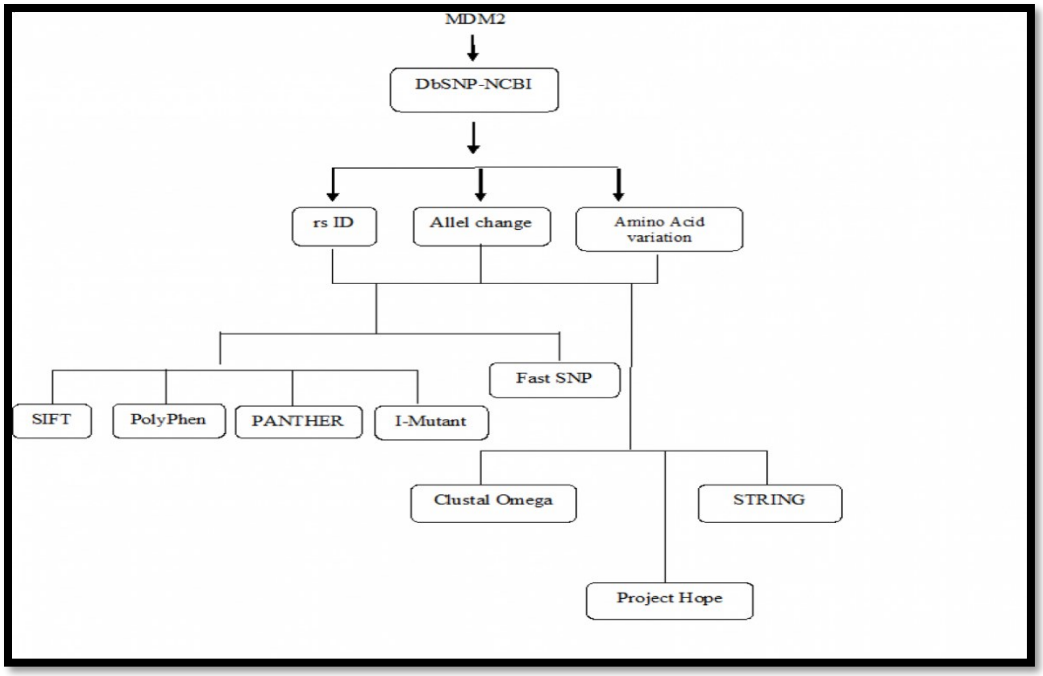

Fig. 1: Predicting the functional consequences of non-synonym encoding SNPs.

Analysis of Molecular Phenotypic Effects by SIFT

SIFT is used to describes the impact of amino acid substitution on protein function. SIFT performed chemical assessment and determined whether the protein is tolerable or not. From the selected 29 SNPs, the software evidence that 9 SNPs results were DAMAGING *Warning! Low confidence, 14 were DAMAGING and 6 were TOLERATED (**Table 3**).

Table 3: SIFT prediction for 29 SNPs of MDM2 with RefSeq Protein ID isNP_002383.

Gene ID	rs IDs	Coordinates	Codons	Substitution	Score	Median Info	Prediction
ENSG00000135679	rs1411053475	12,69202268,1,G/C	AGC-AcC	S4T	0	3.52	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs1319092231	12,69202992,1,A/G	ATG-gTG	M1V	0	4.32	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs1347430167	12,69203002,1,C/T	ACC-AtC	T10I	0.05	3.36	DAMAGING *Warning! Low confidence.
12,69203002,1,C/T	rs1135874	12,69203006,1,C/A	AAC-AAa	N5K	0	4.32	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs1450828029	12,69203019,1,A/G	ACT-gCT	T16A	0.09	3.14	TOLERATED
ENSG00000135679	rs1347430167	12,69203002,1,C/T	ACC-AtC	T4I	0.04	3.34	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs781195441	12,69203011,1,C/T	TCT-TtT	S7F	0	4.32	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs577736495	12,69203016,1,C/T	CCT-tCT	P9S	0	4.32	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs201821879	12,69203017,1,C/T	CCT-CtT	P9L	0.01	3.04	DAMAGING
ENSG00000135679	rs1450828029	12,69203019,1,A/G	ACT-gCT	T10A	0	3.03	DAMAGING
ENSG00000135679	rs746284240	12,69203023,1,A/G	GAT-GgT	D11G	0.05	2.91	DAMAGING
ENSG00000135679	rs1183704933	12,69203032,1,T/G	GTA-GgA	V14G	0.05	2.88	DAMAGING
ENSG00000135679	rs771046322	12,69203047,1,T/C	ATT-AcT	I19T	0	3.24	DAMAGING
ENSG00000135679	rs1361311287	12,69203058,1,G/A	GAA-aAA	E23K	0.01	3.19	DAMAGING
ENSG00000135679	rs1361311287	12,69203058,1,G/C	GAA-cAA	E23Q	0.01	3.19	DAMAGING
ENSG00000135679	rs1246173207	12,69203064,1,G/C	GAG-cAG	E25Q	0.05	2.78	DAMAGING
ENSG00000135679	rs746439458	12,69203065,1,A/G	GAG-GgG	E25G	0.01	3.24	DAMAGING
ENSG00000135679	rs772554005	12,69203068,1,C/T	ACC-AtC	T26I	0	4.32	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs1470334779	12,69207334,1,G/A	GTT-aTT	V28I	0.05	2.77	DAMAGING
ENSG00000135679	rs1226240451	12,69207356,1,T/C	TTG-TcG	L35S	0	3.67	DAMAGING *Warning! Low confidence.
ENSG00000135679	rs1321827832	12,69207373,1,G/A	GTT-aTT	V41I	0.02	2.78	DAMAGING
ENSG00000135679	rs777599995	12,69207374,1,T/C	GTT-GcT	V41A	0.69	3.18	TOLERATED
ENSG00000135679	rs1487823798	12,69207391,1,A/T	ACT-tCT	T47S	0.02	2.78	DAMAGING
ENSG00000135679	rs1385829631	12,69210629,1,G/A	CGA-CaA	R65Q	1	2.78	TOLERATED
ENSG00000135679	rs1385829631	12,69210629,1,G/A	CGA-CaA	R71Q	1	2.77	TOLERATED
ENSG00000135679	rs774100788	12,69210643,1,A/C	AAG-cAG	K45Q	0.29	2.79	TOLERATED
ENSG00000135679	rs774100788	12,69210643,1,A/C	AAG-cAG	K70Q	0.3	2.77	TOLERATED
ENSG00000135679	rs759244097	12,69210644,1,A/T	AAG-AtG	K45M	0	2.79	DAMAGING
ENSG00000135679	rs759244097	12,69210644,1,A/T	AAG-AtG	K70M	0	2.85	DAMAGING

Simulation of functional significances of nsSNPs usingPolyPhen-2

The batch mode was used to analyze SNPs that contain basic protocols for obtaining PolyPhen-2 through its web interface. After analyzing the SNPs, this software revealed that 8 out of 29 SNPs are likely to be harmful (**Table 4**). Thus, 6 SNPs were considered to be potentially harmful (**Table 5**) and 15 SNPs were benign (**Table 6**). The mutation is shown in (**Fig. 2**).

Table 4. Analysis of MDM2 probably damaging variants through polyphen

#o_acc	Opos	o_aa1	o_aa2	Rsid	acc	Pos	aa1	aa2	Prediction	pph2_prob	pph2_FPR	pph2_TPR
Q00987	10	T	I	rs1347430167	Q00987	10	T	I	probably damaging	0.995	0.0277	0.681
Q00987	25	E	Q	rs1246173207	Q00987	25	E	Q	probably damaging	0.999	0.00574	0.136
Q00987	25	E	G	rs746439458	Q00987	25	E	G	probably damaging	0.999	0.00574	0.136
Q00987	28	V	I	rs1470334779	Q00987	28	V	I	probably damaging	0.982	0.0393	0.748
Q00987	35	L	S	rs1226240451	Q00987	35	L	S	probably damaging	1	0.00026	0.00018
Q00987	45	K	M	rs759244097	Q00987	45	K	M	probably damaging	0.995	0.0277	0.681
Q00987	70	K	M	rs759244097	Q00987	70	K	M	probably damaging	0.979	0.0411	0.755
Q00987	71	R	Q	rs1385829631	Q00987	71	Q	R	probably damaging	0.966	0.0465	0.775

Table 5: Analysis of MDM2 possibly damaging variants through polyphen

#o_acc	o_pos	o_aa1	o_aa2	Rsid	acc	pos	aa1	aa2	Prediction	pph2_prob	pph2_FPR	pph2_TPR
Q00987	7	S	F	rs781195441	Q00987	7	S	F	possibly damaging	0.939	0.0562	0.8
Q00987	23	E	K	rs1361311287	Q00987	23	E	K	possibly damaging	0.808	0.0715	0.841
Q00987	23	E	Q	rs1361311287	Q00987	23	E	Q	possibly damaging	0.859	0.0667	0.831
Q00987	26	T	I	rs772554005	Q00987	26	T	I	possibly damaging	0.944	0.0544	0.797
Q00987	45	K	Q	rs774100788	Q00987	45	K	Q	possibly damaging	0.863	0.0664	0.83
Q00987	70	K	Q	rs774100788	Q00987	70	K	Q	possibly damaging	0.669	0.0853	0.863

Table 6: Analysis of MDM2 damaging variants through polyphen

#o_acc	o_pos	o_aa1	o_aa2	Rsid	acc	pos	aa1	aa2	Prediction	pph2_prob	pph2_FPR	pph2_TPR
Q00987	4	S	T	rs1411053475	Q00987	4	T	S	Benign	0.058	0.162	0.939
Q00987	1	M	V	rs1319092231	Q00987	1	M	V	Benign	0.004	0.408	0.975
Q00987	4	T	I	rs1347430167	Q00987	4	T	I	Benign	0.001	0.852	0.994
Q00987	5	N	K	rs1135874	Q00987	5	N	K	Benign	0	1	1
Q00987	9	P	S	rs577736495	Q00987	9	P	S	Benign	0	1	1
Q00987	9	P	L	rs201821879	Q00987	9	P	L	Benign	0.006	0.253	0.967
Q00987	10	T	A	rs1450828029	Q00987	10	T	A	Benign	0.044	0.17	0.942
Q00987	16	T	A	rs1450828029	Q00987	16	T	A	Benign	0.004	0.408	0.975
Q00987	11	D	G	rs746284240	Q00987	11	D	G	Benign	0	1	1
Q00987	14	V	G	rs1183704933	Q00987	14	V	G	Benign	0.019	0.2	0.953
Q00987	19	I	T	rs771046322	Q00987	19	I	T	Benign	0.368	0.108	0.898
Q00987	41	V	I	rs1321827832	Q00987	41	V	I	Benign	0.097	0.146	0.93
Q00987	41	V	A	rs777599995	Q00987	41	V	A	Benign	0.002	0.704	0.987
Q00987	47	T	S	rs1487823798	Q00987	47	T	S	Benign	0.025	0.189	0.949
Q00987	65	R	Q	rs1385829631	Q00987	65	R	Q	Benign	0.034	0.178	0.946

(acc original protein identifier., pos original substitution position in the protein sequence, aa1 original wild type (reference) amino acid residue, aa2 original mutant (substitution) amino acid residue, rsiddbSNP reference SNP identifier (rsID) if available, accUniProtKB accession if known protein, otherwise same as, pos substitution position in UniProtKB protein sequence, Otherwise same as o

pos, aa1 wild-type amino acid residue in relation to UniProtKB sequence, aa2 mutant amino acid residue in relation to UniProtKB sequence, Prediction qualitative ternary classification appraised at 5%/10% (HumDiv) or 10%/20% (HumVar) FPR thresholds (benign, possibly damaging, probably damaging)16 pph2 prob classifier probability of the variation being damaging, Pph2 prob classifier probability of the variation being damaging, Pph2 FPR classifier model False Positive Rate (1−specificity) at the above probability, pph2 TPR classifier model True Positive Rate (sensitivity) at the above probability.

Characterization of Functional nsSNPs by PANTHER

We use PANTHER to analyze MDM2-nsSNPs to include an additional level of refinement in the characterization of SNPs. A total of 29 SNPs are used and are therefore classified as 8 potentially harmful.

Table 7: Possible functional effects of amino acid variation through statistical modelling and special effects of these mutation on the stability of the described protein

PANTHER HMM: null (PTHR10360)			I-Mutant results:				
Substitution	**preservation time**	**Message**	**Stability**	**RI**	**pH**	**T**	**rs IDSs**
T4S	361	possibly damaging	Decrease	6	7.0	25	rs1411053475
M1V	361	possibly damaging	Decrease	8	7.0	25	rs1319092231
T10I	455	probably damaging	Decrease	3	7.0	25	rs1347430167
N5K	176	probably begign	Decrease	3	7.0	25	rs1347430167
T16A	324	possibly damaging	Decrease	6	7.0	25	rs1135874
T4I	361	possibly damaging	Decrease	3	7.0	25	rs781195441
S7F	176	probably begign	Increase	5	7.0	25	rs577736495
P9S	91	probably begign	Decrease	8	7.0	25	rs201821879
P9L	91	probably begign	Decrease	0	7.0	25	rs1450828029
T10A	455	probably damaging	Decrease	8	7.0	25	rs1450828029
D11G	324	possibly damaging	Decrease	6	7.0	25	rs746284240
V14G	176	probably begign	Decrease	9	7.0	25	rs1183704933
I19T	176	probably begign	Decrease	7	7.0	25	rs771046322
E23K	220	possibly damaging	Decrease	8	7.0	25	rs1361311287
E23Q	220	possibly damaging	Decrease	6	7.0	25	rs1361311287
E25Q	455	probably damaging	Decrease	5	7.0	25	rs1246173207
E25G	455	probably damaging	Decrease	6	7.0	25	rs746439458
T26I	176	probably begign	Increase	2	7.0	25	rs772554005
V28I	750	probably damaging	Decrease	1	7.0	25	rs1470334779
L35S	455	probably damaging	Decrease	5	7.0	25	rs1226240451
V41I	176	probably begign	Decrease	2	7.0	25	rs1321827832
V41A	176	probably begign	Decrease	7	7.0	25	rs777599995
T47S	361	possibly damaging	Decrease	7	7.0	25	rs1487823798
R65Q	176	probably begign	Decrease	5	7.0	25	rs1385829631
Q71R	456	probably damaging	Decrease	3	7.0	25	rs1385829631
K45Q	455	probably damaging	Increase	6	7.0	25	rs774100788
K70Q	455	probably damaging	Decrease	3	7.0	25	rs774100788
K45M	455	probably damaging	Increase	8	7.0	25	rs759244097
K70M	455	probably damaging	Increase	3	7.0	25	rs759244097

Note: (RI: Reliability Index, DDG: DG(NewProtein)-DG(WildType) in Kcal/mol, DDG<0: Decrease Stability, DDG>0: Increase Stability, T: Temperature in Celsius degrees, pH: -log[H+])

11 probably harmful and 10 probably benign (**Table 7**).

The analysis of stability of the human MDM2 protein

I-Mutant decides on adjusting the free energy and gives the direction for the change. The single point mutation showed a decrease in the stability of all SNPs with the exception of rs781195441 (S7F), rs772554005 (T36I), rs774100788 (K45Q), rs759244097 (K45M) and rs759244297 (K70M).

The structural Impacts of Functional MDM2 Mutations

The HOPE project showed that all mutations change in size depending on the wild type and the mutated amino acid. The altered residues are smaller than the wild type residues T10A, L35S, K45M and K70M. T10A leads to a possible loss of external interactions. The modified V28I residue is larger than the wild type residue.

The load on these wilds was POSITIVA, K45M and K70M, but now their load on the mutated wastes is neutral. The charge of the E25Q wild-type residue is lost as a result of this mutation. This can lead to the

loss of interactions with other molecules. The wild type residue V28I was buried in the nucleus of the protein. The rest of the mutation is larger and probably doesn't fit. L35S leads to a possible loss of external interactions. The hydrophobicity of wild-type and mutant residues is different. The mutation can lead to the loss of hydrophobic interactions with other molecules on the protein's surface. It is known that protein charge and mass influence the dynamic space-time forces of protein-protein interactions [27, 28]. Thus, these fluctuations can alter the ability of MDM2 to relate to other proteins. All results are shown in **Table 8 & fig. 6I, II, III, IV, V**.CLUSTAL O (1.2.4) multiple sequence alignment Clustal Omega creates precise, high-quality alignments by adapting various protein sequences. The result of the multiple sequence alignment is that there are 4 residues (T10A, E25Q, E25G, V28I, L35S and K70M) of MDM2 with "*", which means that they are in the much-conserved region. However, 1 residue (K45M) was marked with ":", which means that this was demonstrated in the case of preservation between groups with very similar properties (**Fig. 3a & b**).

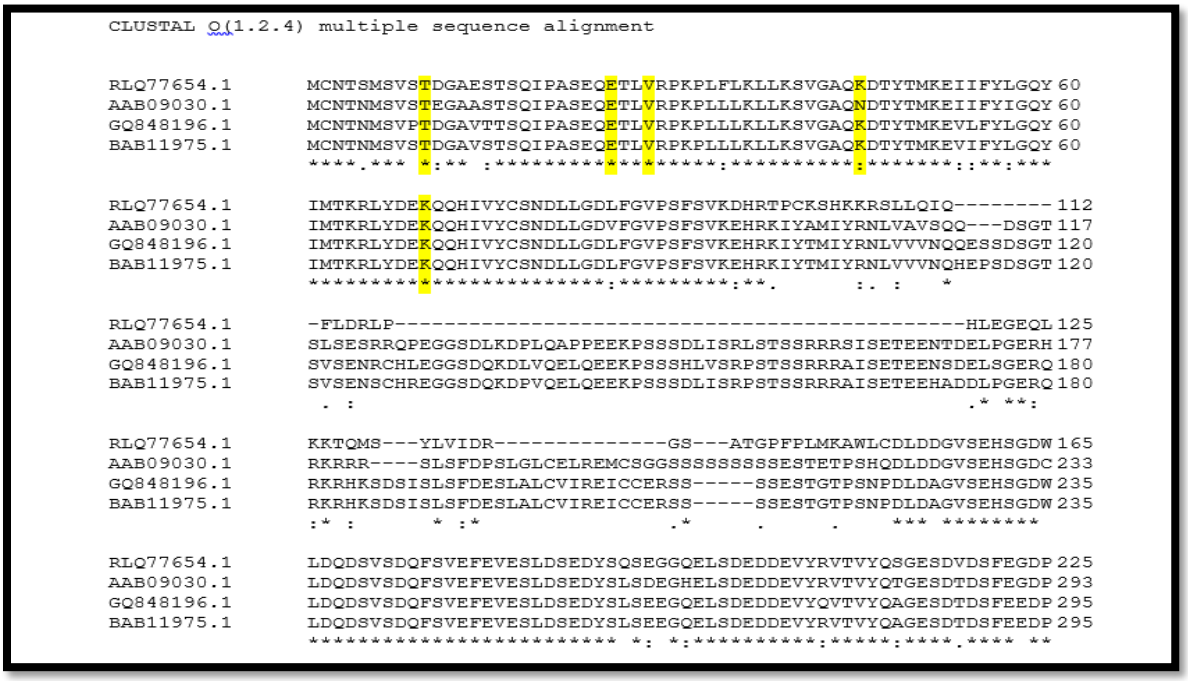

Fig. 3a: CLUSTAL O (1.2.4) multiple sequence alignment

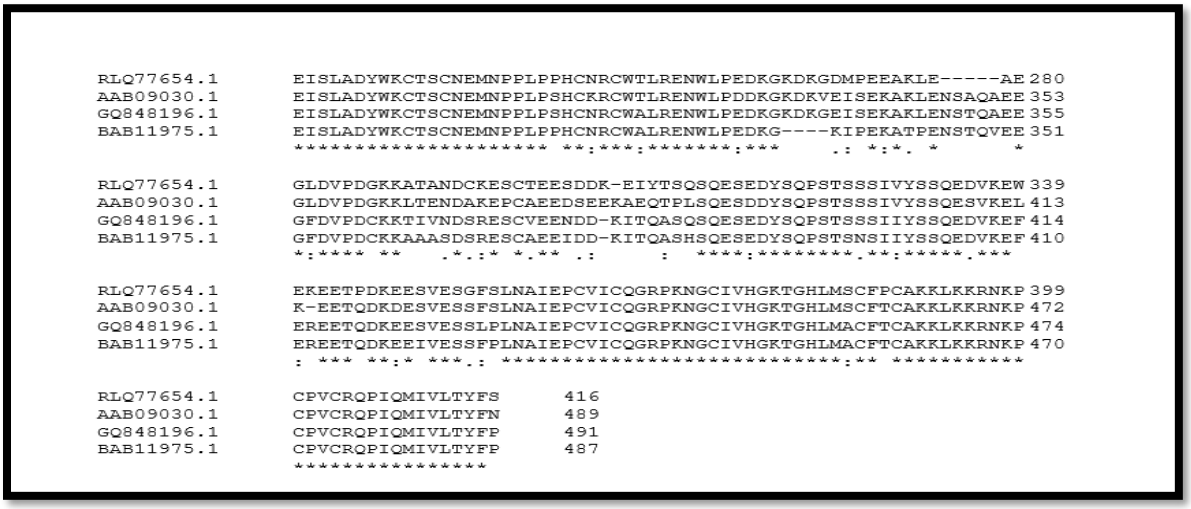

Fig. 3b: CLUSTAL O (1.2.4) multiple sequence alignment

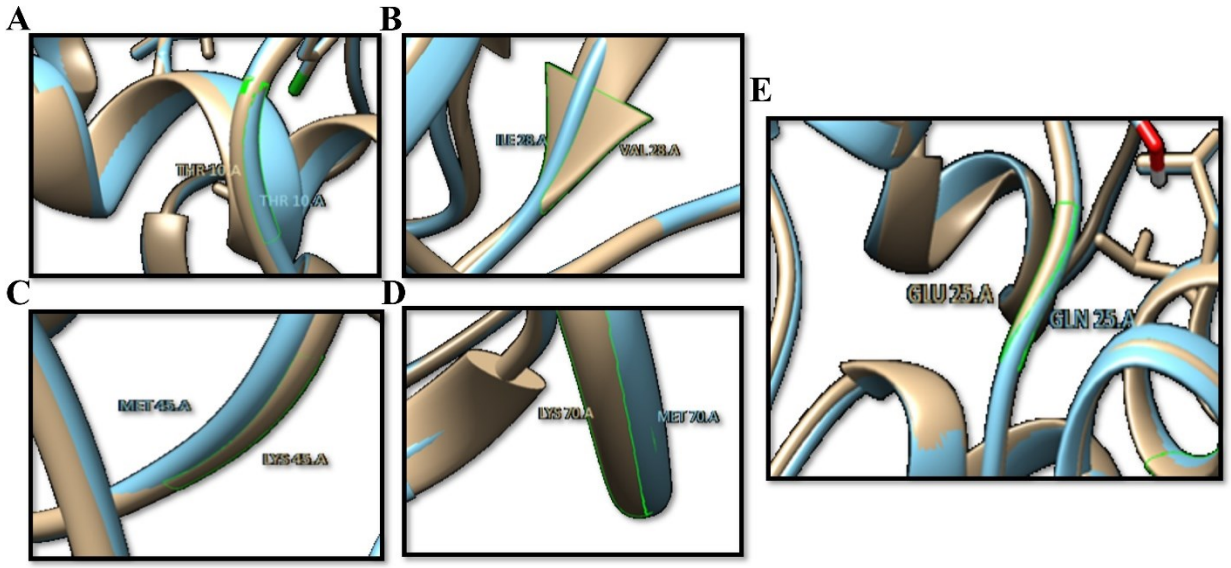

Fig. 4: Superimposed structures of MDM2 (camel color) and mutant (blue color) models to visualize the stereo chemical conformation of wild type and mutant residues at 10, 25, 28, 45 and 70 positions.

Prediction of the interactions of MDM2 protein with others

By using its 5,214,234 protein database of 1,113 species, the STRING for creates protein-protein interaction through, either direct or indirect, associations amid known proteins and different proteins by applying its database. The input of this tool was the MDM2 gene name and Homo sapiens. This network has 11 nodes, 40 edges, average node degree 7.27 and average local clustering coefficient 0.851.

The network's PPI enrichment p-value is 0.00012, which implies that your proteins have a large number of interactions among themselves than what might be normal for an irregular arrangement of proteins of relative size, extracted from the genome. This enrichment indicates that the protein is at least partially biologically linked as a group.

Biological processes in which these genes of the network have significant enrichment parameter of G1/S transition of the mitotic cell cycle, the parameter of signal transduction by the p53 class mediator and mitotic G1 DNA damage checkpoint.

Molecular functions enriched in this network are MDM2/MDM4 family protein binding, disorder domain specific binding, enzyme binding, damaged DNA binding, ubiquitin ligase inhibitor activity, and protein N-terminus binding and pathways enriched are p53 signaling pathway, microRNAs in cancer, Platinum drug resistance, Cellular senescence Glioma (**Fig. 5**).

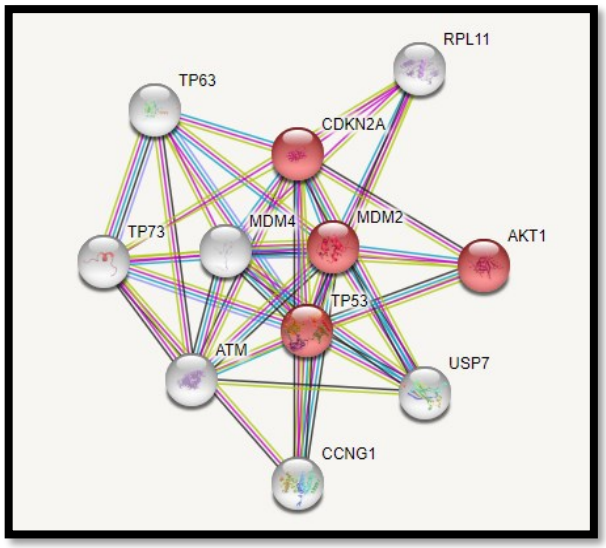

Fig. 5: Strings showing Protein-Protein interactions.

The position of the mutant and replace it with the new amino acid

Chimera was used to visualize the structural characteristics of amino acids in native and mutated protein chains. During structural visualization for all 5 mutations, only the mutated residue (isoleucine) at position 28 showed a network of collisions with valine 28 (**Fig. 4**).

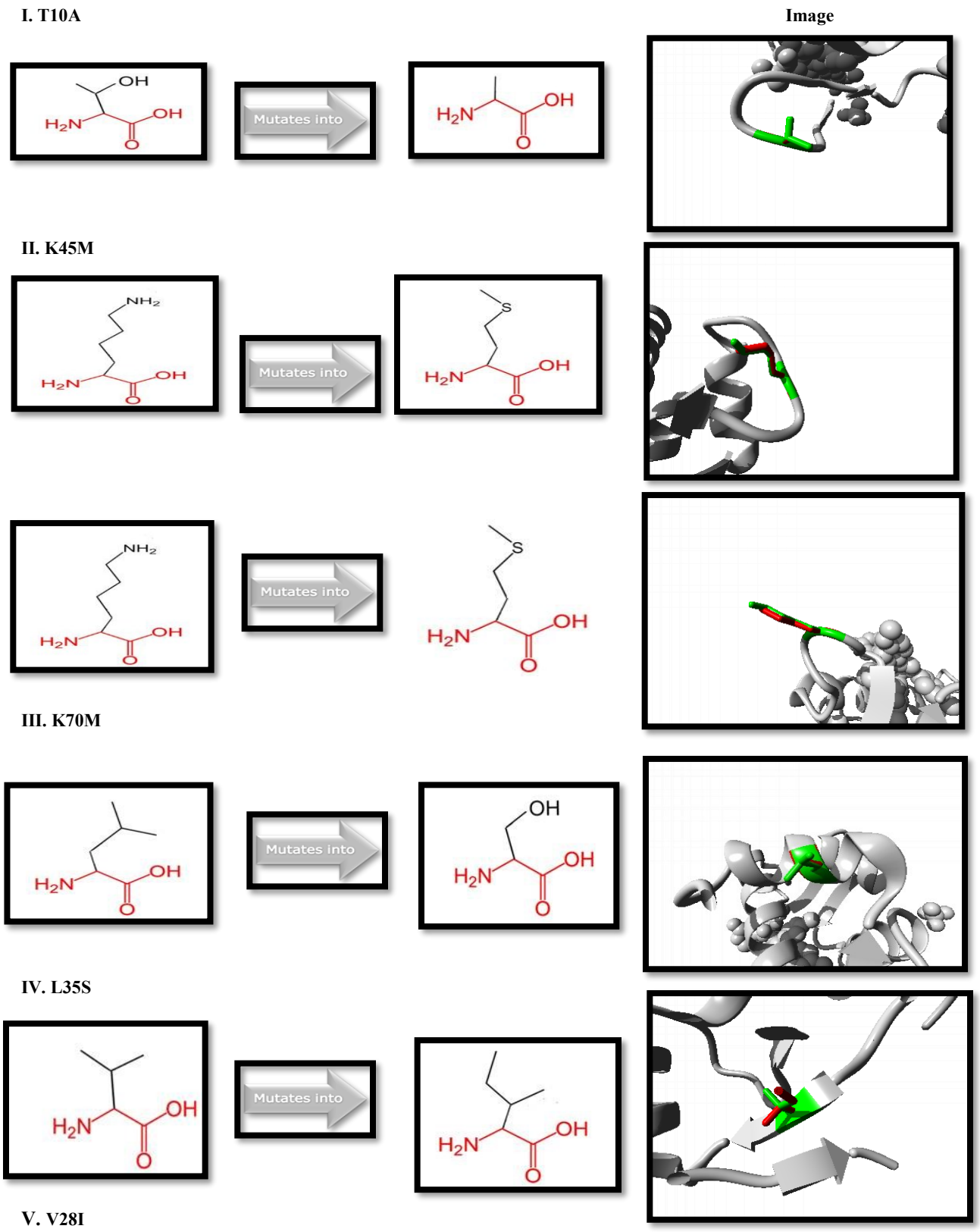

Fig. 6I, II, III, IV, V: Schematic structures of the original (left) and mutated (right) amino acids (HOPE project). The backbone, which is the same for each amino acid, is colored red. The unique side chain of each amino acid is black. In close-up of the mutation image. The protein is gray in color, the side chains of the wild type and the mutated residue are shown and the colors are green and red, respectively.

Table 8: Amino Acid & their properties can fluctuate the ability of MDM2 to relate to other proteins.

Amino Acid Name	Amino Acid Properties
T10A	The wild-type and mutant amino acids differ in size. The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant residue differs.
K45M	There is a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type residue is lost by this mutation. This can cause loss of interactions with other molecules. The mutant residue is smaller than the wild-type residue. The wild-type and mutant amino acids differ in size. This will cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant residue differs. The wild-type residue charge was POSITIVE; the mutant residue charge is NEUTRAL.
K70M	There is a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type residue is lost by this mutation. This can cause loss of interactions with other molecules. The wild-type and mutant amino acids differ in size. The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant residue differs. The wild-type residue charge was POSITIVE; the mutant residue charge is NEUTRAL.
L35S	The wild-type and mutant amino acids differ in size. The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant residue differs. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein.
V28I	The wild-type and mutant amino acids differ in size. The mutant residue is bigger than the wild-type residue. The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit.

Discussion

Genome sequencing and whole exome, the list of genetic mutations is growing over time with the introduction of high-throughput technology. After sequencing the missense mutations associated with the disease, the next stage is to find functional and structural consequences of the protein levels in order to better understand the biological mechanisms of the disease association that can help to identify the main pharmaceutical components using in silico methods. The human MDM2 gene is located with 13 exons in the region of chromosome 12q15. A 2746 base pair linear mRNA encodes the functional protein MDM2 with 491 amino acids. To date, approximately 12,951 variants have been found in the coding, non-coding and regulatory regions of the human MDM2 gene in the dbSNP database.

In the current study, algorithms based on sequence and structure (for example, SIFT, PolyPhen-2 and PANTHER) were used to track functional genetic variations of MDM2 in the coding region. A functional study by SIFT predicted that 48% of all nsSNPs are harmful, while PolyPhen-2 predicted that 52% are benign and 27% of all nsSNPs are likely to be harmful. nsSNP, which encodes variants T10A, E25Q, V28I, L35S, K45M and K70M, is functionally significant when compared to the results of three different silica tools (SIFT, PolyPhen-2 and PANTHERcSNP). The difference in predictability is due to the fact that each method used a different sequence and alignment group. In comparison, predictive sequence-based analysis has many advantages over structure-based analysis because it takes into account a wide range of effects at the protein level and is suitable for proteins familiar with relatives [26]. Amino acid substitutions were later provided in I-Mutant 2.0 to confirm their deleterious effects after the functional consequences of the candidate alternatives were analysed by SIFT, PolyPhen and PANTHER.

In the mutant results, 24 (83%) of the 29 mutations were scored as a decrease in stability and considered to be very deleterious mutations of the MDM2 proteins (Figure 3 (I, II)). This decrease in stability

was an impediment with DDG values from -1.96 to 0.71 and IR values from 1 to 9 by I-Mutant Suit. We predict that K45M and K70M are associated with the disease (**Fig. 5.1 & 5.2**). A highly conserved region was found in the output file of MDM2 residues from Clustal Omega 4, which have the asterisk "*" (T10A, E25Q, V28I, L35S and K70M) (Figure 4). Although conservation between groups of very similar properties (K45M) has been observed, this is indicated by ";" displayed. To understand the global organization of proteomics in a functional network environment, the study of protein-protein interaction is an integrative approach. The interactive network shows biomolecules as nodes and interactions that connect two nodes as edges. Functional network visualizations of individual genomes are widely used today to develop the statistical power of human molecular genetics, to provide drug discovery benefits, to understand metabolic pathways well, and to derive genotype-phenotype associations [27]. STRING shows that MDM2 interacts with MDM4, TP53, CDKN2A, ATM, CCNG1, USP7, AKT1, RPL11, TP63 and TP73 (Figure 5). However, sequence-based predictions cannot describe the basic mechanisms between protein genotype and phenotype relationships. In contrast, structure-based methods have limitations, as they cannot be used for proteins in the case of unknown 3D structures. Insilico study tools that combine sequence and structure-based methods provide additional benefits, provided reliable prediction results are obtained, with different aspects of SNP analysis covering a wider area[28]. However, the variability in the predicted output of these algorithms reflects the pros and cons.

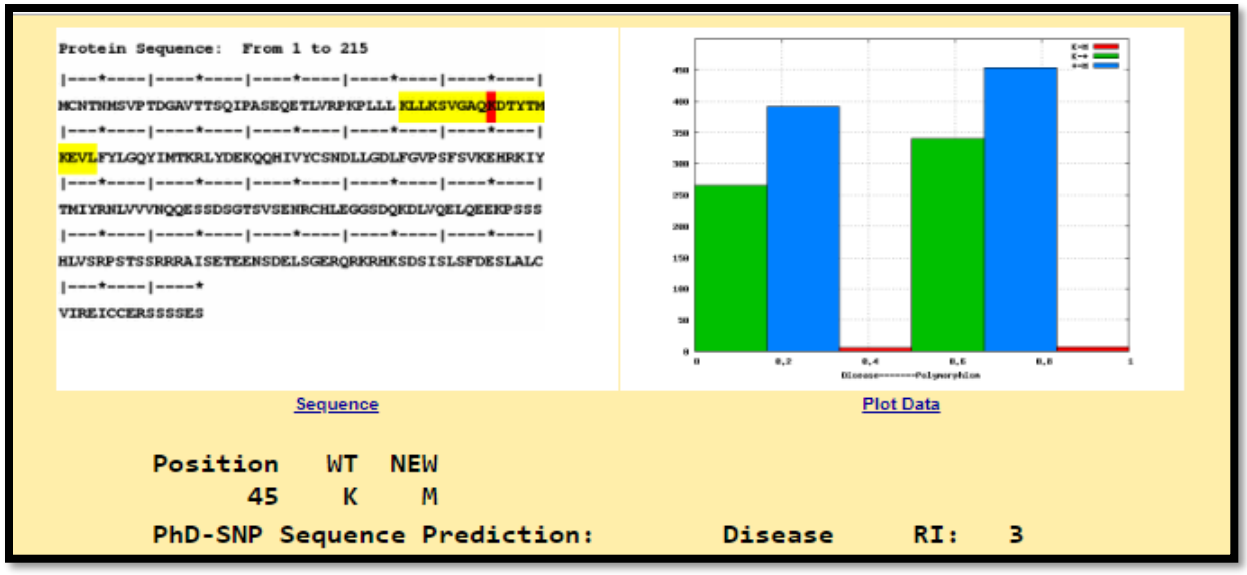

Fig. 7: Prediction of disease associate single point mutation K45M from protein sequence

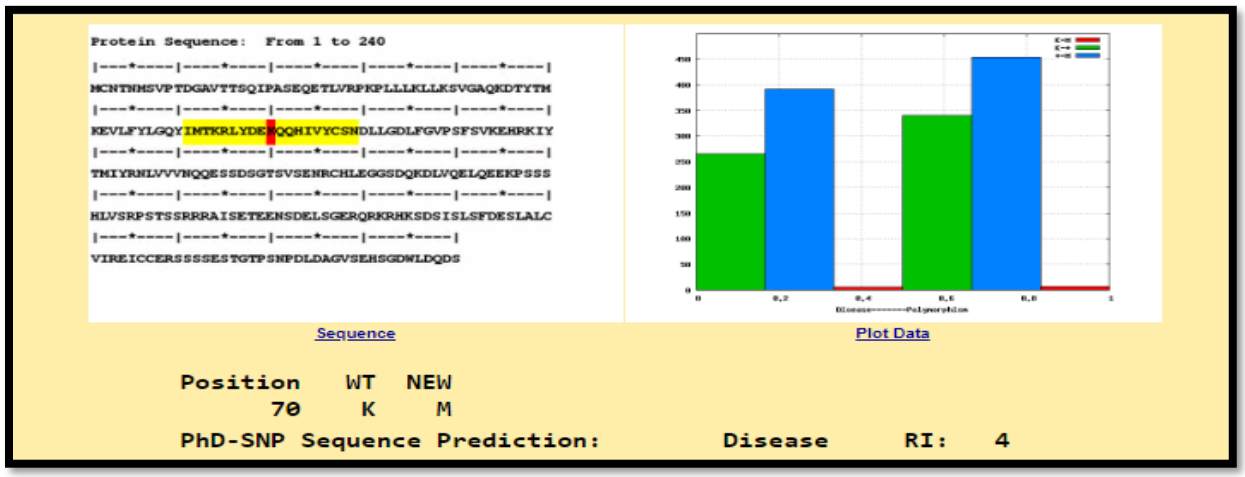

Fig. 8: Prediction of disease associate single point mutation K70M from protein sequence

When mapping the SIFT, PolyPhen-2, I-Mutant and PANTHER consensus study, we performed the analysis of the protein structure to better match the image. Calculations of solvent accessibility and power field energy provide an understanding of the structural and functional effects of charged amino acids. 3D models are designed to visualize the differences between wild-type and mutated protein models. To study the structural effects of the selected amino acid changes used the HOPE project [29]. The T10A residue is a mutated residue that is more hydrophobic compared to the wild type residue. After the move, the E25Q lost its charge of wild-type waste. This can lead to the loss of interactions with different molecules. The K70M mutated residue is more hydrophobic compared to the wild type residue. The mutant residues T10A, K45M and K70M are smaller compared to the wild type residue. The mutated residue E25K is larger than the wild type residue.

Conclusion

Compared to experimental methods, computational methods are simpler and more economical. It is difficult to assess a large number of SNPs using experimental methods. In this study, nsSNPs in the MDM2 gene were found to be affected in structure and functions, which may lead to susceptibility to various diseases (such as cancer) that are linked to DNA stability. Nine software programs were used to identify the target SNPs that have the greatest impact on the protein's function. Six of the 29 nsSNPs were selected and one was predicted to be the most destructive of all software programs and to analyze structural changes. All of these nsSNPs are being extensively studied for the association of cancer in many populations. No one evaluated the association of K45M (rs759244097) with diseases that were considered the most harmful in our study. This can be an important SNP target for cancer susceptibility and should be verified experimentally. It is also related to the risk of breast, lung and colon cancer. Therefore, this can also play a role in neuropathological diseases. Damage to the MDM2 function reduces the repair efficiency of DNA repair systems and therefore increases the risk of developing cancer.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Barroso I, Gurnell M, Crowley V, Agostini M, Schwabe J, Soos M, et al. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. Nature. 1999;402:880.

[2] Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation1. Journal of molecular biology. 2001;307:683-706.

[3] Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993;362:59.

[4] Oliner J, Kinzler KW, Meltzer P, George D, Vogelstein B. Amplification of a gene encoding a p53-associated protein in human sarcomas. Nature. 1992;358:80.

[5] Keleti J, Quezado MM, Abaza MM, Raffeld M, Tsokos M. The MDM2 oncoprotein is overexpressed in rhabdomyosarcoma cell lines and stabilizes wild-type p53 protein. The American journal of pathology. 1996;149:143.

[6] Bates S, Vousden KH. p53 in signaling checkpoint arrest or apoptosis. Current opinion in genetics & development. 1996;6:12-8.

[7] !!! INVALID CITATION !!!

[8] Fåhraeus R, Olivares-Illana V. MDM2's social network. Oncogene. 2014;33:4365.

[9] Rodriguez MS, Desterro JM, Lain S, Lane DP, Hay RT. Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation. Molecular and cellular biology. 2000;20:8458-67.

[10] Mazzaferri E. Papillary and follicular thyroid cancer: selective therapy. Comprehensive therapy. 1981;7:6-14.

[11] Pant V, Lozano G. Limiting the power of p53 through the ubiquitin proteasome pathway. Genes & development. 2014;28:1739-51.

[12] Momand J, Jung D, Wilczynski S, Niland J. The MDM2 gene amplification database. Nucleic acids research. 1998;26:3453-9.

[13] Oliner JD, Pietenpol JA, Thiagalingam S, Gyuris J, Kinzler KW, Vogelstein B. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. Nature. 1993;362:857.

[14] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009;4:1073.

[15] (https://sift.bii.a-star.edu.sg/www/SIFT_help.html#SIFT).

[16] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics. 2011;12:745.

[17] Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, et al. Clinical assessment incorporating a personal genome. The Lancet. 2010;375:1525-35.

[18] Ja T. Bigham aW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and Functional Impact of Rare

Coding Variation from Deep Sequencing of Human Exomes. Science. 2012;337:64-9.

[19] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics. 2013;76:7.20. 1-7.. 41.

[20] Mavroconstanti T, Johansson S, Winge I, Knappskog PM, Haavik J. Functional properties of rare missense variants of human CDH13 found in adult attention deficit/hyperactivity disorder (ADHD) patients. PLoS One. 2013;8:e71445.

[21] Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC bioinformatics. 2010;11:548.

[22] Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic acids research. 2010;38:W695-W9.

[23] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7:539.

[24] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research. 2010;39:D561-D8.

[25] http://www.cgl.ucsf.edu/chimera/.

[26] Hussain MRM, Shaik NA, Al-Aama JY, Asfour HZ, Khan FS, Masoodi TA, et al. In silico analysis of Single Nucleotide Polymorphisms (SNPs) in human BRAF gene. Gene. 2012;508:188-96.

[27] Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterology and Hepatology from bed to bench. 2014;7:17.

[28] Yazar M, Özbek P. In Silico Tools and Approaches for the Prediction of Functional and Structural Effects of Single-Nucleotide Polymorphisms on Proteins: An Expert Review. OMICS: A Journal of Integrative Biology. 2021;25:23-37.

[29] Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS genetics. 2008;4:e1000083.