American Journal of Bioinformatics Research 2016, 6(2): 56-97
DOI: 10.5923/j.bioinformatics.20160602.03

Computational Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) in Human MutS Homolog6 (*MSH6*) Gene

Nahla E. Abdelraheem[1,*], Marwa Mohamed Osman[2], Osama Muhieldin Elgemaabi[3], Afra Abdelhamid Fadl Alla[2], Mosab Mohamed Ismail[2], Soada Ahmed Osman[2], Aisha Ismail Ibrahim[2], Nihad Elsadig Babekir[4], Salwa Osman Mekki[5], Mohamed A. Hassan[2,6]

[1]Department of Histopathology and Cytology, Faculty of Medical Laboratory Sciences, National University, Khartoum, Sudan
[2]Department of Biotechnology, Africa City of Technology, Khartoum, Sudan
[3]Department of General Surgery, Omdurman Military Hospital, Omdurman, Sudan
[4]National Center for Neurological Sciences (NCNS), Khartoum, Sudan
[5]Deparment of Pathology, Faculty of Medicine, Alneelin University, Khartoum, Sudan
[6]Division of Molecular Genetics, Institute of Human Genetics, University of Tubingen, Germany

Abstract **Background:** Point mutations in MSH6 gene had been related to group of cancers called lynch syndrome which accounts for 3% to 5% of all colorectal cancers. Despite the excessively studied MSH6 mutations, the mechanism by which these mutations promote carcinogenesis remains controversial. **Methods:** MSH6 was investigated in dbSNP/NCBI in December 2015, 3666 SNPs were found in human; 388 were coding synonymous, 937 non synonymous, 201 in frame shift, 63 in 3' un-translated region and 347 in 5' un-translated region. Non synonymous and 3'UTR SNPs were selected for insilico analysis; SIFT, Polyphen2, Imutant3.0, MUpro, PhD-SNP, SNPs & GO, MutPred, ELASPIC, Mutation 3D, UCSF Chimera 1.8, PolymiRTs, and GENEMAIA softwares and servers were used to investigate the effect of SNPs on MSH6 protein's structure and function. **Results:** 21 SNPs were found to be highly damaging for the protein by SIFT and Polyphen, and were further analyzed by I-Mutant, MUpro, PHD-SNP, SNPs & GO, ELASPIS, Mutation 3D and Chimera. 2 SNPs were predicted by PolymiRTs to induce disruption or creation of miR binding sites; rs200412142 contained 2 disrupting and 1 creating functional classes in 3 miRSite, while rs184571821 SNP contained 3 creating functional classes in 3 miRSites. GENEMANIA revealed five genes similar in their expression level with MSH6 and seven genes share the same protein domain with it. **Conclusions:** 14 nsSNPs (R1217K, H1248D, E1214A, T1219I, K1140R, A1303T, M1137T, A1303G, R915K, H946D, T917I, A1001T, L899F, A1001G) were located at the interface of the MSH6 protein interfering with its relation with MSH2ISO2, MSH3, MSH2 and E9PHA6. Interactions of MSH6 with these proteins are critical for its MMR function and any structural alterations that interfere or harm these networks interactions would probably increase susceptibility to tumors formation and progression. 2 SNPs at the 3UTR; rs200412142 and rs184571821 introduced a change in the micro RNA binding site at the 3UT which might result in deregulation of the gene function.

Keywords Lynch syndrome, MSH6, SNPs, In Silico, SIFT, Polyphen2, MutPred, Genemania, Polymirts and UCSFChimera

1. Introduction

Single nucleotide polymorphism (SNP) is the variation in a genetic sequence that affects only one of the basic building blocks—adenine (A), guanine (G), thymine (T), or cytosine (C)—in a segment of a DNA molecule and that occurs in more than 1% of a population. The DNA of humans may contain many SNPs, since these variations occur at a rate of one in every 100–300 nucleotides in the human genome. In fact, roughly 90 percent of the genetic variation that exists between humans is the result of SNPs. Although the majority of variations do not alter cellular function and thus have no effect, some SNPs have been discovered to contribute to the development of diseases such as cancer and to influence physiological responses to drugs. [1] An understanding of the relationship between these genetic variations and their phenotypic effects could therefore be a step toward exploring the causes of various disorders or diseases. SNPs can fall within the coding regions (coding SNPs) or noncoding

* Corresponding author:
nahlaelrayah@gmail.com (Nahla E. Abdelraheem)
Published online at http://journal.sapub.org/bioinformatics

regions of genes (non coding SNPs), or in the intergenic region between two genes [2, 3]. While the two others are quite natural in the human genome and phenotypically neutral [4], nonsynonymous coding SNPs (nsSNPs) are thought to have the principal impact on phenotype by changing the protein sequence. As they cause amino acid alteration in the corresponding protein product, it may exert deleterious effects on the structure, function, solubility, or stability of proteins [5]. Also the nsSNPs perturb gene regulation by modifying DNA and transcriptional binding factors [6] and the maintenance of the formational integrity of cells and tissues [7]. Thus, it is likely that nsSNPs play a major role in the functional diversity of coded proteins in human populations and often associated with human diseases. Indeed, studies have revealed that more than 50% of the mutations associated with inherited genetic disorders are resulted from nsSNPs [8]. Disease-causing mutations are frequently observed at either core or interface residues mediating protein interactions. Mutations at core residues frequently destabilize protein structure while mutations at interface residues can specifically affect the binding energies of protein-protein interactions. Missense mutations at protein–protein interaction sites, called interfaces, are important contributors to human disease. Interfaces are non-uniform surface areas characterized by two main regions, "core" and "rim", which differ in terms of evolutionary conservation and physicochemical properties. Moreover, within interfaces, only a small subset of residues ("hot spots") is crucial for the binding free energy of the protein–protein complex [9]. Protein-protein interactions are critical for nearly every process in the cell and deleterious mutations hindering these interactions can have severe consequences on the associated cellular function. A variety of efforts from personalized medicine to understand viral evolution require knowing how specific mutations effect the protein-protein interactions. Conversely, designing proteins with improved binding or altered specificity requires that the impact of mutations on the native interface be understood. Currently this information is not available experimentally on the proteome-wide scale necessary for these tasks. Considerable effort has been devoted towards developing methods to predict the impact of mutations on binding affinity. Most of these approaches rely on physics based methods that attempt to faithfully model on the atomic level the interactions determining protein-protein binding affinity. However, a major obstacle of such approaches is the need for the reconstruction of the full-atomic model for every mutant complex, which limits the accuracy of the approach (since the position of the side-chains is difficult to model) and reduces the computational speed and the range of applications (since rebuilding the full-atomic model is generally the most time-consuming step) [10]. MSH6 gene is located on chromosome 2 and consists of 13 exons. It encodes a member of the mismatch repair family (MMR). In E.coli; the MutS protein helps in the recognition of mismatched nucleotides prior to their repair. A highly conserved region of approximately 150 amino acids, called the walker-A adenine nucleotide binding motif exists in MutS homologs. The encoded protein heterodimerizes with MSH2 to form a mismatch recognition complex that functions as a bidirectional molecular switch that exchange ADP and ATP as DNA mismatches are bound and dissociated [11]. MSH6 gene mutations had been found to be involved in Lynch syndrome, leading to the production of an abnormally short, nonfunctional MSH6 protein or a partially active form of the protein. When the MSH6 protein is absent or nonfunctional, the number of mistakes that are left unrepaired during cell division increases substantially. The errors accumulate as the cells continue to divide, which may cause the cells to function abnormally, increasing the risk of tumors formation in the body [12]. In a MMR mechanism, the mismatch recognition function is fulfilled by one of the heterodimeric protein complexes, MSH2–MSH6 (MutSα) or MSH2–MSH3 (MutSβ), dependent on the type of mutation. The MutSα complex recognizes base–base mismatches and small insertion–deletion loops (IDL), whereas the MutSβ complex recognizes IDLs basically larger than one extra helical nucleotide [13-16]. The MSH6 and MSH3 proteins are shown to be functionally redundant, so that MutSα can partially compensate the function of MutSβ while MutSβ appears to only recognize insertions and deletions [17-19]. Because of this redundancy, mutations in MSH6 cause accumulation of base substitutions but less frequently frameshift mutations in microsatellite sequences [17, 20]. This explains the low microsatellite instability (MSI) in MSH6-deficient tumors. Lack of frameshift mutations, which can easily target repetitive sequences also in tumor suppressor genes and result in their inactivation during tumorigenesis, may further explain the late onset in many MSH6 mutation carriers. Lynch syndrome, also known as hereditary non polyposis colorectal cancer syndrome (HNPCC), accounts for 3% to 5% of all colorectal cancers and is an autosomal dominant inheritance, the susceptibility disorder is caused by germline mutations in mismatch repair (MMR) genes, 10% in MSH6 and PMS2 [2]. Carriers of MMR gene mutations are at high risk of early-onset colorectal and endometrial cancer. The Lynch syndrome includes tumors of the ovaries, small bowel, urothelium, biliary tract, and stomach [21, 22], and is generally suspected if there is familial aggregation of Lynch syndrome – associated cancers using criteria such as Amsterdam II or Bethesda [23, 24] or a tumor phenotype showing high DNA microsatellite instability [25]. MSH6 mutations are associated with many cancers; sporadic and hereditary colorectal cancer, endometrial cancer, prostate cancer, gastrointestinal cancers, childhood hematologic malignancies, glioblastoma, anaplastic oligodendroglioma and melanoma [26-34]. The number of mutations reported in the MSH6 gene is continually rising [35]. Families associated with MSH6 mutations unusually often display carcinomas of the endometrium [36-38]. The significance of MSH6 in endometrial carcinomas development is emphasized by the observation that lack of MSH6 protein characterizes endometrial but not colon carcinomas in

HNPCC [39].

Recently bioinformatics has became increasingly important in biology and bioinformatic tools aid in the processing and extraction of useful results from large amounts of raw data; textual mining of biological literature, analysis of gene and protein expression, simulation and modeling of DNA, RNA, and protein structures, comparison of genetic and genomic data and helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. [40-43]. The International Collaborative Group on HNPCC (ICG-HNPCC) has over 30 potentially pathogenic MSH6 mutations in the database (http://www.nfdht.nl). A significant proportion (35%) of them results in a single amino acid substitution, which is difficult to interpret. The pathogenicity of HNPCC mutations is linked to malfunction of MMR. despite the association of MSH6 mutations with many types of cancers; the exact role of many reported mutations in tumorgenesis and cancer progression remain unknown. In this study we adopted an insilico approach to analyze human MSH6 reported mutations using different bioinformatics softwares to investigate the effect of single nucleotide polymorphisms on protein's structure and function and whether these variations can contribute in disease or not.

2. Material and Methods

Different soft-wares; SIFT, polyphen-2, Imutant3.0, MUpro, PhD- SNP, SNPs & GO MutPred, ELASPIC, mutation 3D, GENEMANIA, PolymiRTs and chimera were used to investigate the effect of SNPs mutations on MSH6 protein structure and function. Prediction of deleterious effect of non synonymous SNPs was done by SIFT and Polyphen-2 soft-wares. Prediction of stability changes was investigated in I mutant-3 and MUpro. The association of nsSNPs with disease was done by PhD-SNP and SNPs & GO software. The structural changes in 3D structure including hydrogen bonding, clashes and contacts for each residue were analyzed using Chimera software. SNPs at the 3'UTR were analyzed to detect the effect on microRNA binding sites using PolymiRTs soft-ware. GENEMANIA was used to investigate MSH gene interactions. In this study we selected nsSNPs and those at the 3'UTR regions for analysis.

2.1. Investigation of MSH6 Gene's Interactions and Appearance in Networks in GENEMANIA Database GENEMANIA: is an online database that helps in the prediction of gene function; it also finds other genes that are related to a set of input genes, using a very large set of functional association data that include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. It can also be used to find new members of a pathway or complex, find additional genes that may have been missed in screening or find new genes with a specific function, such as protein kinases. The question is defined by the set of genes in the input [44]. Available at: http://www.genemania.org/.

2.2. Prediction of Structural Impact of nsSNPs on Protein by SIFT software (v5.1)

SIFT (Separating Intolerant from Tolerant): 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. Positions important for function should be conserved in an alignment of the protein family, whereas unimportant positions should appear diverse in an alignment. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. It is a multistep procedure that searches for similar sequences, then chooses closely related sequences that may share similar function to the query sequence, followed by obtaining the alignment of these chosen sequences, and finally calculates normalized probabilities for all possible substitutions from the alignment. [45] The input SNPs' rs-IDs were submitted to the server for analysis, positions with normalized probabilities less than 0.05 were predicted to be deleterious; those greater than or equal to 0.05 were predicted to be tolerated. Available at: http://sift.bii.a-star.edu.sg/.

2.3. Prediction of Deleterious nsSNPs by PolyPhen-2

Polyphen (Polymorphism Phenotyping v2) is available as software and via a Web server. It predicts the possible impact of amino acid substitutions on the stability and function of human proteins using structural and comparative evolutionary considerations. It performs functional annotation of single-nucleotide polymorphisms (SNPs), maps coding SNPs to gene transcripts, extracts protein sequence annotations and structural attributes, and builds conservation profiles, then estimates the probability of the missense mutation being damaging based on a combination of all these properties. PolyPhen-2 features include a high-quality multiple protein sequence alignment pipeline and a prediction method employing machine-learning classification. The software also integrates the UCSC Genome Browser's human genome annotations and MultiZ multiple alignments of vertebrate genomes with the human genome [46]. The input FASTA sequence of protein with the position of interest and the new residue were submitted to Polyphen to predict functional impact of mutations. Available at: http://genetics.bwh.harvard.edu/pph2/.

2.4. Analysis of nsSNPs' Impact on Protein Stability

2.4.1. I Mutant 3.0 Server: I-Mutant is a Support Vector Machine-based web server for the automatic prediction of protein stability changes upon single-site mutations starting from the protein structure or sequence. In both cases, it can predict the protein stability change corresponding to all possible mutations of a particular residue, or ask only for a

specific mutation. In either case, I-Mutant3.0 can predict the direction of the free energy change and its value [47]. The input FASTA sequence of protein along with the residues changes were submitted to I mutant server for the analysis of DDG value (kcal/mol) and the RI value (reliabilityindex) was also computed. Available at http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/IMutant3.0.cgi.

2.4.2. Mupro: is a set of machine learning programs to predict how single-site amino acid mutation affects protein stability. It uses two machine learning methods: Support Vector Machines and Neural Networks. Both of them were trained on a large mutation dataset and show accuracy above 84% via 20 fold cross validation, which is better than other methods. One advantage of the method is that it do not require tertiary structures to predict protein stability changes [48, 49]. The value of the energy change is predicted, and a confidence score between -1 and 1 for measuring the confidence of the prediction is calculated. A score <0 means the variant decreases the protein stability; conversely, a score >0 means the variant increases the protein stability. Available at: http://www.ics.uci.edu/~baldig/mutation.html.

2.5. Prediction of Disease Associated Variations

2.5.1. PhD-SNP Software: PhD-SNP is Support Vector Machine based classifier that is optimized to predict if a given single point protein mutation can be classified as disease-related or as neutral polymorphism [50-53]. The input FASTA sequences of protein along with the residues change were submitted to PhD-SNP server for the analysis. Available at: http://snps.biofold.org/phd-snp/phd-snp.html.

2.5.2. SNPs & GO: is a support vector machine (SVM) based on the method to accurately predict the mutation related to disease from protein sequence. The input is the FASTA sequence of the whole protein, the output is based on the difference among the neutral and disease related variations of the protein sequence. The RI (reliability index) with value of greater than 5 depicts the disease related effect caused by mutation on the function of parent protein [54]. Available at: http://snps.biofold.org/snps-and-go/snps-and-go.html.

2.6. Prediction of Harmful Mutations by Mutpred

The Mutpred server was employed to classify an amino acid substitution (AAS) as disease-associated or neutral. In addition, it predicts molecular cause of disease/deleterious AAS. Mutpred is based upon SIFT and a gain/loss of 14 different structural and functional properties. The output of Mutpred contains a general score (g), i.e., the probability that the amino acid substitution is deleterious/disease-associated, and top 5 property scores (p), where p is the P-value that certain structural and functional properties are impacted [55]. Available at: http://mutpred.mutdb.org/.

2.7. Prediction of the Stability Effects of Mutation in Domain Cores and Domain-Domain Interfaces

ELASPIC is a novel ensemble machine learning approach that predicts the effects of mutations on protein folding and protein-protein interactions. Here we present the ELASPIC web server, which makes the ELASPIC pipeline available through a fast and intuitive interface. The web server can be used to evaluate the effect of mutations on any protein in the Uniprot database, and allows all predicted results, including modeled wild-type and mutated structures, to be managed and viewed online and downloaded if needed. It is backed by a database which contains improved structural domain definitions, and a list of curated domain-domain interactions for all known proteins, as well as homology models of domains and domain-domain interactions for the human proteome. Homology models for proteins of other organisms are calculated on the fly, and mutations are evaluated within minutes once the homology model is available [10]. Available at: http://elaspic.kimlab.org/many/.

2.8. Distribution of nsSNPs in MutS domains by Mutations 3D

Mutation3D is a functional prediction and visualization tool for studying the spatial arrangement of amino acid substitutions on protein models and structures. It is intended to be used to identify clusters of amino acid substitutions arising from somatic cancer mutations across many patients in order to identify functional hotspots and fuel downstream hypotheses. It is also useful for clustering other kinds of mutational data, or simply as a tool to quickly assess relative locations of amino acids in proteins [56]. Available at: http://mutation3d.org/index.shtml.

2.9. Protein 3d Modeling and Detection of Hydrogen Bonding and Clashes by UCSF Chimera

CPH models 3.2: is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions [57]. Available at: http://www.cbs.dtu.dk/services/CPHmodels/.

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles [58]. Chimera (version 1.8) software was used to scan the 3D (three-dimensional) structure of specific protein, and hence modifies the original amino acid with the mutated one to see the impact that can be produced. The outcome is then a graphic model depicting the mutation. Chimera (version 1.8) is available within the Chimera package from the Chimera web site http://www.cgl.ucsf.edu/chimera/.

2.11. Prediction of the Impact of SNPs at the 3Un Translated Region (3UTR) by PolymiRTS Database

PolymiRTS database: is designed specifically for the analysis of non-coding SNPs at 3'UTR. It identifies single-nucleotide polymorphisms (SNPs) that affect miRNA (micro RNA) targets in human and mouse [59]. We used this

computational server to analyze 3'UTR SNPs in MSH6 gene that may alter miRNA binding on target sites resulting in diverse functional consequences. All SNPs located in that region were selected and submitted to PolymiRTS (v3.0), available at: http://compbio.uthsc.edu/miRSNP.

3. Results

3.1. Retrieving SNPs and Protein's Sequence from the Database

A total of 7654 MSH6 SNPs were investigated in dbSNP/NCBI database (http://www.ncbi.nlm.nih.gov/snp) in December 2015; 3666 SNPs were in Homo sapiens; of which 388 were coding synonymous and 937 non synonymous SNPs, 201 were frame shift, 63 were in 3' un-translated region and 347 were in the 5' un-translated region. The FASTA formats of the protein (its isoforms and fragments) were obtained from Uniprot at Expassy database (http://expasy.org); the Uniprot accession numbers: (P52701, C9JH55 and C9J7Y7).

3.2. MSH6 Gene Interactions and Appearance in Networks

Genemania revealed that MSH6 has many vital functions; a substantial role in DNA recombination, mismatch repair complex binding, negative regulation of DNA metabolic process, ATP metabolic process, DNA-dependent ATPase activity, reciprocal meiotic recombination, purine nucleoside monophosphate catabolic process, structure-specific DNA binding, purine, ATP catabolic process, nucleoside monophosphate catabolic process, coupled meiotic nuclear division, meiotic cell cycle, nucleoside monophosphate metabolic process, meiosis I, nuclear chromosome, double-stranded DNA binding, regulation of DNA metabolic process, DNA secondary structure binding, cellular process involved in reproduction, nuclear division, organelle fission, somatic diversification of immunoglobulins, somatic diversification of immune receptors, imunoglobulin production, production of molecular mediator of immune response, magnesium ion binding, isotype switching and response to radiation. The genes co-expressed with, share similar protein domain, or participate to achieve similar function are listed in table (7, 8 in the appendix) and Figure (1).

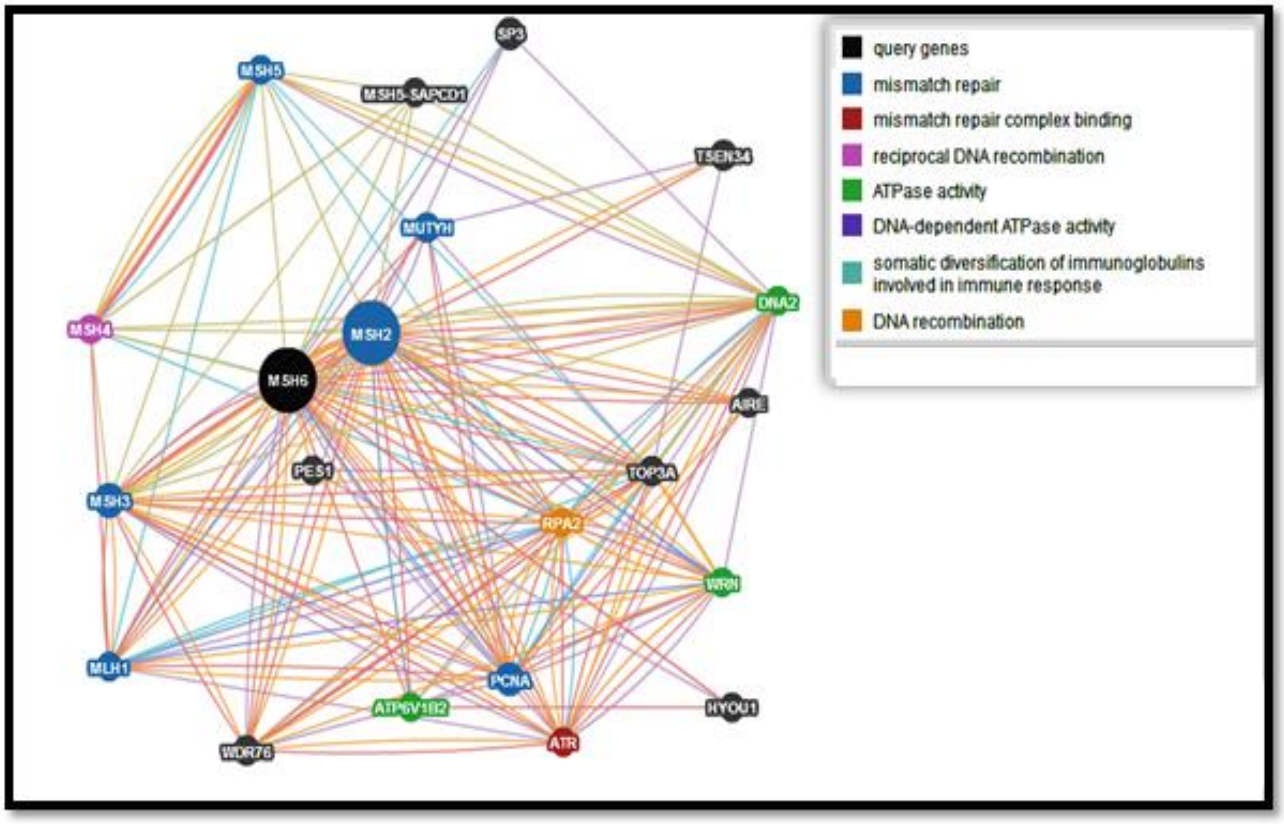

Figure 1. Functional interactions between MSH6 and its related genes

3.3. Prediction of Deleterious nsSNPs by SIFT and Polyphen

Coding SNPs were analyzed using SIFT and Polyphen soft-wares. Batch nsSNPs (rs-IDs) were submitted to SIFT server; 108 SNPs (288 mutations) were predicted to be deleterious out of 937 SNPs, Table (appendix). Deleterious SNPs were submitted to Polyphen-2 as query sequences in FASTA Format, 79 SNPs (202 mutations) were predicted to be probably damaging, the other 30 SNPs (53 mutations) were scored as possibly damaging, 101 SNPs (255 mutations) were predicted to be damaging by both servers (table 9 in the appendix). 21 SNPs (59 mutations) achieved high scores (TI= 0 I sift server and PSIC SD=1 by polyphen-2 software) and had been chosen for further analysis. Table (1)

3.4. Prediction of Harmful nsSNPs by Mutpred

MutPred was used to determine the tolerance degree for each amino acid substitution on the basis of physio-chemical properties. The results obtained from MutPred server are shown in table (2). These results suggest that some nsSNPs may account for potential structural and functional changes in MSH6 protein.

3.5. Identification of Disease Related nsSNPs by PhD-SNP and SNPs & GO

PhD-SNP and SNPs & GO softwares were used to predict the association of SNPs with disease. According to PhD-SNP software, 19 SNPs (53 mutations) (R→K, R→W, Y→C, R→Q, H→D, E→A, T→I, K→R, C→W, G→S, A→T, E→K, M→T, L→F, A→G, Y→H, T→M and R→C) were disease related while 3 SNPs (6 mutations) (S→L, H→N, A→P) were predicted to be neutral polymorphisms. SNPs & GO predicted 7 SNPs (20 mutations) as neutral and 14 SNPs (39 mutations) (S→L, H→N, A→P, R→C, Y→H, Y→C and R→K) (R→W, R→Q, H→D, E→A, T→I, K→R, C→W, G→S, A→T, E→K, M→T, L→F, A→G, T→M and) were disease related. Table (3)

3.6. Prediction of nsSNPs Impact on the Protein Stability by I-Mutant and MUpro

In I-Mutant 3.0, the protein stability changed due to a single point mutation in 20 SNPs (56 mutations) in *MSH6* gene (S→L, R→K, R→W, R→Q, H→D, E→A, T→I, K→R, C→W, G→S, A→T, E→K, M→T, L→F, A→G, H→N, A→P, T→M, R→C, Y→H) decrease effective stability of the protein and one SNP (3 mutations) (Y→C) was predicted to increase protein stability. In MUpro 6 SNPs (17mutations) (S→L, R→W, R→Q, H→D, K→R, T→M) were found to increase the protein stability and 15 SNPs (42mutations) were found to decrease protein stability (R→K, E→A, T→I, C→W, G→S, A→T, E→K, M→T, L→F, A→G, H→N, A→P, R→C, Y→H, Y→C). Table (4)

3.7. Prediction of nsSNPs at the Protein Core and Interface

ELASPIC server predicted 27 mutations in the core of MSH6 protein and 14 mutations were at the interface while 18 mutations were not found by the server. Table (5)

3.8. Distribution of Mutations in MutS Domain by Mutation3D Server

Five domain were detected in MSH6 protein; MutS I (PF01624), MutS II (PF05188), MutS III (PF05192), MutS V (PF00488), PWWP (PF00855) 54 mutations were located in a protein domain structure, 4 mutations were uncovered (H65N, H237N, A137P, R338C) and E1063 was located in the inter domain region. Those mutations in the domains regions are considered as higher risk mutations for MSH6 protein. The result is shown in figure 2.

3.9. Protein 3d Modeling and Detection of Hydrogen Bonding and Clashes by UCSF Chimera

UCSF Chimera v 1.8 was used to model the 3d structure of both wild and mutant residue and also to show hydrogen bonds, clashes and contacts of the mutant residue. Results are shown in figures 3-22.

Table 1. Shows highly damaging non synonymous SNPs predicted by SIFT and Polyphen2

SNP ID	Protein ID	Nucleotide Change	Amino Acid Change	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs41295270	ENSP00000234420	C/T	S580L	Deleterious	0	Probably Damaging	1
rs41295270	ENSP00000446475	C/T	S450L	Deleterious	0	Probably Damaging	1
rs63749898	ENSP00000234420	G/A	R1217K	Deleterious	0	Probably Damaging	1
rs63749898	ENSP00000438580	G/A	R915K	Deleterious	0	Probably Damaging	1
rs63749898	ENSP00000446475	G/A	R1087K	Deleterious	0	Probably Damaging	1
rs63750138	ENSP00000234420	C/T	R772W	Deleterious	0	Probably Damaging	1
rs63750138	ENSP00000438580	C/T	R470W	Deleterious	0	Probably Damaging	1
rs63750138	ENSP00000446475	C/T	R642W	Deleterious	0	Probably Damaging	1
rs63750389	ENSP00000234420	A/G	Y850C	Deleterious	0	Probably Damaging	1
rs63750389	ENSP00000438580	A/G	Y548C	Deleterious	0	Probably Damaging	1
rs63750389	ENSP00000446475	A/G	Y720C	Deleterious	0	Probably Damaging	1
rs63750725	ENSP00000234420	G/A	R772Q	Deleterious	0	Probably Damaging	1

SNP ID	Protein ID	Nucleotide Change	Amino Acid Change	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs63750725	ENSP00000438580	G/A	R470Q	Deleterious	0	Probably Damaging	1
rs63750725	ENSP00000446475	G/A	R642Q	Deleterious	0	Probably Damaging	1
rs63750882	ENSP00000234420	C/G	H1248D	Deleterious	0	Probably Damaging	1
rs63750882	ENSP00000438580	C/G	H946D	Deleterious	0	Probably Damaging	1
rs63750882	ENSP00000446475	C/G	H1118D	Deleterious	0	Probably Damaging	1
rs63750914	ENSP00000234420	A/C	E1214A	Deleterious	0	Probably Damaging	1
rs63750914	ENSP00000438580	A/C	E912A	Deleterious	0	Probably Damaging	1
rs63750914	ENSP00000446475	A/C	E1084A	Deleterious	0	Probably Damaging	1
rs63750949	ENSP00000234420	C/T	T1219I	Deleterious	0	Probably Damaging	1
rs63750949	ENSP00000438580	C/T	T917I	Deleterious	0	Probably Damaging	1
rs63750949	ENSP00000446475	C/T	T1089I	Deleterious	0	Probably Damaging	1
rs63750969	ENSP00000234420	A/G	K1140R	Deleterious	0	Probably Damaging	1
rs63750969	ENSP00000438580	A/G	K838R	Deleterious	0	Probably Damaging	1
rs63750969	ENSP00000446475	A/G	K1010R	Deleterious	0	Probably Damaging	1
rs63750985	ENSP00000438580	C/G	C463W	Deleterious	0	Probably Damaging	1
rs63751063	ENSP00000234420	G/A	G1139S	Deleterious	0	Probably Damaging	1
rs63751063	ENSP00000438580	G/A	G837S	Deleterious	0	Probably Damaging	1
rs63751063	ENSP00000446475	G/A	G1009S	Deleterious	0	Probably Damaging	1
rs63751064	ENSP00000234420	G/A	A1303T	Deleterious	0	Probably Damaging	1
rs63751064	ENSP00000438580	G/A	A1001T	Deleterious	0	Probably Damaging	1
rs63751064	ENSP00000446475	G/A	A1173T	Deleterious	0	Probably Damaging	1
rs63751328	ENSP00000234420	G/A	E1193K	Deleterious	0	Probably Damaging	1
rs63751328	ENSP00000438580	G/A	E891K	Deleterious	0	Probably Damaging	1
rs63751328	ENSP00000446475	G/A	E1063K	Deleterious	0	Probably Damaging	1
rs148445930	ENSP00000234420	T/C	M1267T	Deleterious	0	Probably Damaging	1
rs148445930	ENSP00000438580	T/C	M965T	Deleterious	0	Probably Damaging	1
rs148445930	ENSP00000446475	T/C	M1137T	Deleterious	0	Probably Damaging	1
rs182024561	ENSP00000438580	C/T	L899F	Deleterious	0	Probably Damaging	1
rs182024561	ENSP00000446475	T/C	L1071F	Deleterious	0	Probably Damaging	1
rs201060668	ENSP00000234420	C/G	A1303G	Deleterious	0	Probably Damaging	1
rs201060668	ENSP00000438580	C/G	A1001G	Deleterious	0	Probably Damaging	1
rs201060668	ENSP00000446475	C/G	A1173G	Deleterious	0	Probably Damaging	1
rs201193496	ENSP00000234420	C/A	H367N	Deleterious	0	Probably Damaging	1
rs201193496	ENSP00000438580	C/A	H65N	Deleterious	0	Probably Damaging	1
rs201193496	ENSP00000446475	C/A	H237N	Deleterious	0	Probably Damaging	1
rs267608052	ENSP00000234420	G/C	A457P	Deleterious	0	Probably Damaging	1
rs267608052	ENSP00000438580	G/C	A155P	Deleterious	0	Probably Damaging	1
rs267608052	ENSP00000446475	G/C	A327P	Deleterious	0	Probably Damaging	1
rs267608089	ENSP00000234420	C/T	T1142M	Deleterious	0	Probably Damaging	1
rs267608089	ENSP00000438580	C/T	T840M	Deleterious	0	Probably Damaging	1
rs267608089	ENSP00000446475	C/T	T1012M	Deleterious	0	Probably Damaging	1
rs369456858	ENSP00000234420	C/T	R468C	Deleterious	0	Probably Damaging	1
rs369456858	ENSP00000438580	C/T	R166C	Deleterious	0	Probably Damaging	1
rs369456858	ENSP00000446475	C/T	R338C	Deleterious	0	Probably Damaging	1
rs373622047	ENSP00000234420	T/C	Y994H	Deleterious	0	Probably Damaging	1
rs373622047	ENSP00000438580	T/C	Y692H	Deleterious	0	Probably Damaging	1
rs373622047	ENSP00000446475	T/C	Y864H	Deleterious	0	Probably Damaging	1

PolyPhen-2 result: POROBABLY DAMAGING (more confident prediction) / POSSIBLY DAMAGING (less confident prediction), **PSIC SD:** Position-Specific Independent Counts software, **Tolerance Index**: Ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is ≤ 0.05, and tolerated if the score is > 0.05**.**

Table 2. Prediction of the functional impact of nsSNPS on MSH6 protein by MutPred

Amino Acid Change	Probability of deleterious mutation	Top 5 features	Hypotheses of molecular mechanism disrupted
S580L	0.237 (Not harmful)	Loss of phosphorylation at S580 (P = 0.0675) Loss of disorder (P = 0.0825) Gain of catalytic residue at S580 (P = 0.1945) Gain of methylation at R583 (P = 0.2312) Gain of helix (P = 0.2684	
S450L	0.216 (Not harmful)	Loss of phosphorylation at S450 (P = 0.0675) Loss of disorder (P = 0.0825) Gain of helix (P = 0.1736) Gain of catalytic residue at S450 (P = 0.1945) Gain of methylation at R453 (P = 0.2634)	
R1217K	0.925 (highly harmful)	Gain of ubiquitination at R1217 (P = 0.0258) Gain of methylation at R1217 (P = 0.0377) Gain of glycosylation at R1217 (P = 0.0597) Gain of disorder (P = 0.1598) Loss of catalytic residue at T1219 (P = 0.1873)	Gain of ubiquitination at R1217 (P = 0.0258) Gain of methylation at R1217 (P = 0.0377)
R915K	0.936 (highly harmful)	Gain of ubiquitination at R915 (P = 0.0289) Gain of methylation at R915 (P = 0.0326) Gain of glycosylation at R915 (P = 0.0597) Gain of disorder (P = 0.1598) Loss of catalytic residue at T917 (P = 0.1873)	Gain of ubiquitination at R915 (P = 0.0289) Gain of methylation at R915 (P = 0.0326)
R1087K	0.934 (highly harmful)	Gain of ubiquitination at R1087 (P = 0.0276) Gain of methylation at R1087 (P = 0.0338) Gain of glycosylation at R1087 (P = 0.0597) Gain of disorder (P = 0.1598) Loss of catalytic residue at T1089 (P = 0.1873)	Gain of ubiquitination at R1087 (P = 0.0276) Gain of methylation at R1087 (P = 0.0338)
R772W	0.955 (highly harmful)	Gain of catalytic residue at R772 (P = 0.0596) Gain of ubiquitination at K775 (P = 0.1097) Gain of loop (P = 0.2045) Loss of MoRF binding (P = 0.2336) Loss of methylation at K771 (P = 0.274	
R470W	0.932 (highly harmful)	Gain of ubiquitination at K473 (P = 0.0582) Gain of catalytic residue at R470 (P = 0.0596) Gain of loop (P = 0.2045) Loss of MoRF binding (P = 0.2336) Loss of methylation at K473 (P = 0.2499)	
R642W	0.949 (highly harmful)	Gain of catalytic residue at R642 (P = 0.0596) Gain of ubiquitination at K645 (P = 0.0599) Gain of loop (P = 0.2045) Loss of MoRF binding (P = 0.2336) Loss of methylation at K641 (P = 0.2562)	
Y850C	0.806 (highly harmful)	Loss of solvent accessibility (P = 0.0053) Gain of helix (P = 0.062) Loss of relative solvent accessibility (P = 0.0676) Loss of ubiquitination at K852 (P = 0.0699) Loss of loop (P = 0.0804)	Loss of solvent accessibility (P = 0.0053)
Y548C	0.672 (harmful)	Loss of solvent accessibility (P = 0.0053) Loss of ubiquitination at K550 (P = 0.0609) Gain of helix (P = 0.062) Loss of relative solvent accessibility (P = 0.0676) Loss of loop (P = 0.0804)	Loss of solvent accessibility (P = 0.0053)
Y720C	0.765 (highly harmful)	Loss of solvent accessibility (P = 0.0053) Loss of ubiquitination at K722 (P = 0.0613) Gain of helix (P = 0.062) Loss of relative solvent accessibility (P = 0.0676) Loss of loop (P = 0.0804)	Loss of solvent accessibility (P = 0.0053)
R772Q	0.934 (highly harmful)	Loss of MoRF binding (P = 0.123) Gain of catalytic residue at R772 (P = 0.1388) Gain of ubiquitination at K775 (P = 0.1397) Gain of loop (P = 0.2045) Gain of glycosylation at K775 (P = 0.2543)	

Amino Acid Change	Probability of deleterious mutation	Top 5 features	Hypotheses of molecular mechanism disrupted
R470Q	0.913 (highly harmful)	Gain of ubiquitination at K473 (P = 0.0853) Loss of MoRF binding (P = 0.123) Gain of catalytic residue at R470 (P = 0.1388) Gain of loop (P = 0.2045) Gain of glycosylation at K473 (P = 0.2543)	
R642Q	0.935 (highly harmful)	Gain of ubiquitination at K645 (P = 0.115) Loss of MoRF binding (P = 0.123) Gain of catalytic residue at R642 (P = 0.1388) Gain of loop (P = 0.2045) Gain of glycosylation at K645 (P = 0.2543)	
H1248D	0.893 (highly harmful)	Gain of phosphorylation at Y1249 (P = 0.1236) Loss of helix (P = 0.1706) Loss of disorder (P = 0.1902) Gain of loop (P = 0.2045) Loss of sheet (P = 0.302	
H946D	0.962 (highly harmful)	Gain of phosphorylation at Y947 (P = 0.1236) Loss of helix (P = 0.1706) Loss of disorder (P = 0.1902) Gain of loop (P = 0.2045) Loss of sheet (P = 0.302)	
H1118D	0.961 (highly harmful)	Gain of phosphorylation at Y1119 (P = 0.1236) Loss of helix (P = 0.1706) Loss of disorder (P = 0.1902) Gain of loop (P = 0.2045) Loss of sheet (P = 0.302)	
E1214A	0.915 (highly harmful)	Gain of methylation at R1217 (P = 0.0553) Loss of catalytic residue at E1214 (P = 0.1089) Loss of stability (P = 0.2868) Loss of sheet (P = 0.302) Loss of disorder (P = 0.3528)	
E912A	0.929 (highly harmful)	Gain of methylation at R915 (P = 0.0534) Loss of catalytic residue at E912 (P = 0.1089) Loss of stability (P = 0.2868) Loss of sheet (P = 0.302) Loss of disorder (P = 0.3528)	
E1084A	0.925 (highly harmful)	Gain of methylation at R1087 (P = 0.0557) Loss of catalytic residue at E1084 (P = 0.1089) Loss of stability (P = 0.2868) Loss of sheet (P = 0.302) Loss of disorder (P = 0.3528)	
T1219I	0.974 (highly harmful)	Loss of methylation at R1217 (P = 0.091) Loss of disorder (P = 0.1169) Loss of catalytic residue at F1222 (P = 0.1306) Loss of phosphorylation at T1221 (P = 0.1795) Loss of helix (P = 0.3949)	
T917I	0.972 (highly harmful)	Loss of methylation at R915 (P = 0.0814) Loss of disorder (P = 0.1169) Loss of catalytic residue at F920 (P = 0.1306) Loss of phosphorylation at T919 (P = 0.1795) Loss of helix (P = 0.3949)	
T1089I	0.976 (highly harmful)	Loss of methylation at R1087 (P = 0.0919) Loss of disorder (P = 0.1169) Loss of catalytic residue at F1092 (P = 0.1306) Loss of phosphorylation at T1091 (P = 0.1795) Loss of helix (P = 0.3949)	
K1140R	0.913 (highly harmful)	Loss of ubiquitination at K1140 (P = 0.0174) Gain of methylation at K1140 (P = 0.0485) Gain of MoRF binding (P = 0.0955) Loss of glycosylation at K1140 (P = 0.1169) Gain of phosphorylation at T1142 (P = 0.1274)	Loss of ubiquitination at K1140 (P = 0.0174) Gain of methylation at K1140 (P = 0.0485)

Amino Acid Change	Probability of deleterious mutation	Top 5 features	Hypotheses of molecular mechanism disrupted
K838R	0.855 (highly harmful)	Loss of ubiquitination at K838 (P = 0.0174) Gain of methylation at K838 (P = 0.0426) Gain of MoRF binding (P = 0.0955) Loss of glycosylation at K838 (P = 0.1169) Gain of phosphorylation at T840 (P = 0.1274)	Loss of ubiquitination at K838 (P = 0.0174) Gain of methylation at K838 (P = 0.0426)
K1010R	0.886 (highly harmful)	Loss of ubiquitination at K1010 (P = 0.0174) Gain of methylation at K1010 (P = 0.05) Gain of MoRF binding (P = 0.0955) Loss of glycosylation at K1010 (P = 0.1169) Gain of phosphorylation at T1012 (P = 0.1274)	Loss of ubiquitination at K1010 (P = 0.0174) Gain of methylation at K1010 (P = 0.05)
C463W	0.883 (highly harmful)	Gain of catalytic residue at P466 (P = 0.0237) Gain of MoRF binding (P = 0.0768) Loss of stability (P = 0.1456) Loss of phosphorylation at T465 (P = 0.3476) Loss of disorder (P = 0.3669)	Gain of catalytic residue at P466 (P = 0.0237)
G1139S	0.975 (highly harmful)	Loss of ubiquitination at K1140 (P = 0.0699) Loss of catalytic residue at P1135 (P = 0.0914) Gain of disorder (P = 0.1064) Loss of methylation at K1140 (P = 0.1297) Loss of phosphorylation at T1142 (P = 0.193)	
G837S	0.950 (highly harmful)	Loss of ubiquitination at K838 (P = 0.0699) Loss of catalytic residue at P833 (P = 0.0914) Gain of disorder (P = 0.1064) Loss of methylation at K838 (P = 0.1511) Loss of phosphorylation at T840 (P = 0.193	
G1009S	0.970 (highly harmful)	Loss of ubiquitination at K1010 (P = 0.0699) Loss of catalytic residue at P1005 (P = 0.0914) Gain of disorder (P = 0.1064) Loss of methylation at K1010 (P = 0.1285) Loss of phosphorylation at T1012 (P = 0.193)	
A1303T	0.926 (highly harmful)	Loss of helix (P = 0.1299) Gain of phosphorylation at A1303 (P = 0.1653) Loss of stability (P = 0.1957) Loss of methylation at R1304 (P = 0.2097) Loss of MoRF binding (P = 0.2535)	
A1001T	0.937 (highly harmful)	Loss of helix (P = 0.1299) Gain of phosphorylation at A1001 (P = 0.1653) Loss of methylation at R1002 (P = 0.1934) Loss of stability (P = 0.1957) Loss of MoRF binding (P = 0.2535)	
A1173T	0.934 (highly harmful)	Loss of helix (P = 0.1299) Gain of phosphorylation at A1173 (P = 0.1653) Loss of stability (P = 0.1957) Loss of methylation at R1174 (P = 0.2061) Loss of MoRF binding (P = 0.2535)	
E1193K	0.978 (highly harmful)	Gain of ubiquitination at E1193 (P = 0.0146) Loss of helix (P = 0.1299) Loss of phosphorylation at T1189 (P = 0.2262) Loss of disorder (P = 0.2501) Loss of stability (P = 0.2938)	Gain of ubiquitination at E1193 (P = 0.0146)
E891K	0.973 (highly harmful)	Gain of ubiquitination at E891 (P = 0.0128) Loss of helix (P = 0.1299) Loss of phosphorylation at T887 (P = 0.2262) Loss of disorder (P = 0.2501) Loss of stability (P = 0.2938)	Gain of ubiquitination at E891 (P = 0.0128)
E1063K	0.973 (highly harmful)	Gain of ubiquitination at E1063 (P = 0.0131) Loss of helix (P = 0.1299) Loss of phosphorylation at T1059 (P = 0.2262) Loss of disorder (P = 0.2501) Loss of stability (P = 0.2938)	Gain of ubiquitination at E1063 (P = 0.0131)

Amino Acid Change	Probability of deleterious mutation	Top 5 features	Hypotheses of molecular mechanism disrupted
M1267T	0.588 (Harmful)	Loss of catalytic residue at M1267 (P = 0.027) Loss of sheet (P = 0.0817) Loss of stability (P = 0.251) Gain of disorder (P = 0.2718) Gain of loop (P = 0.2754)	Loss of catalytic residue at M1267 (P = 0.027)
M965T	0.682 (Harmful)	Loss of catalytic residue at M965 (P = 0.027) Loss of sheet (P = 0.0817) Loss of stability (P = 0.251) Gain of disorder (P = 0.2718) Gain of loop (P = 0.2754)	Loss of catalytic residue at M965 (P = 0.027)
M1137T	0.896 (highly harmful)	Gain of phosphorylation at M1137 (P = 0.0564) Gain of disorder (P = 0.0686) Gain of ubiquitination at K1140 (P = 0.0797) Loss of MoRF binding (P = 0.1385) Loss of sheet (P = 0.1501	
L899F	0.827 (highly harmful)	Gain of catalytic residue at L899 (P = 0.0971) Loss of stability (P = 0.2847) Loss of disorder (P = 0.3187) Gain of glycosylation at S897 (P = 0.3745) Loss of helix (P = 0.3949)	
L1071F	0.807 (highly harmful)	Gain of catalytic residue at L1071 (P = 0.0971) Loss of stability (P = 0.2847) Loss of disorder (P = 0.3187) Gain of glycosylation at S1069 (P = 0.3745) Loss of helix (P = 0.3949)	
A1303G	-	-	-
A1001G	0.935 (highly harmful)	Loss of stability (P = 0.0472) Gain of methylation at R1002 (P = 0.0896) Loss of helix (P = 0.1299) Gain of catalytic residue at A1000 (P = 0.1682) Loss of MoRF binding (P = 0.2546)	Loss of stability (P = 0.0472)
A1173G	0.936 (highly harmful) (highly harmful)	Loss of stability (P = 0.0472) Gain of methylation at R1174 (P = 0.0953) Loss of helix (P = 0.1299) Gain of catalytic residue at A1172 (P = 0.1682) Loss of MoRF binding (P = 0.2546)	Loss of stability (P = 0.0472)
H367N	0.164 (Not harmful)	Gain of helix (P = 0.132) Loss of loop (P = 0.2237) Loss of catalytic residue at H367 (P = 0.2403) Gain of glycosylation at T363 (P = 0.244) Gain of relative solvent accessibility (P = 0.2629)	
H65N	0.201 (Not harmful)	Gain of relative solvent accessibility (P = 0.0249) Gain of solvent accessibility (P = 0.0488) Gain of loop (P = 0.0851) Loss of helix (P = 0.1299) Loss of catalytic residue at H65 (P = 0.2403)	
H237N	0.243 (Not harmful)	Loss of helix (P = 0.079) Gain of relative solvent accessibility (P = 0.1012) Gain of solvent accessibility (P = 0.199) Gain of loop (P = 0.2045) Loss of catalytic residue at H237 (P = 0.2403)	
A457P	-	-	-
A155P	0.672 (Harmful)	Loss of catalytic residue at A155 (P = 0.0401) Loss of stability (P = 0.1203) Gain of glycosylation at S157 (P = 0.1237) Loss of ubiquitination at K151 (P = 0.2391) Gain of disorder (P = 0.288)	Loss of catalytic residue at A155 (P = 0.0401)
A327P	0.689 (Harmful)	Loss of catalytic residue at A327 (P = 0.0401) Loss of stability (P = 0.1203) Gain of glycosylation at S329 (P = 0.1237) Gain of sheet (P = 0.1945)	

Amino Acid Change	Probability of deleterious mutation	Top 5 features	Hypotheses of molecular mechanism disrupted
		Loss of ubiquitination at K323 (P = 0.226)	
T1142M	-	-	-
T840M	0.815 (highly harmful)	Loss of phosphorylation at T840 (P = 0.0437) Loss of ubiquitination at K838 (P = 0.0863) Gain of MoRF binding (P = 0.0972) Loss of catalytic residue at T840 (P = 0.1574) Gain of methylation at K838 (P = 0.1612)	Loss of phosphorylation at T840 (P = 0.0437)
T1012M	0.839 (highly harmful)	Loss of phosphorylation at T1012 (P = 0.0437) Loss of ubiquitination at K1010 (P = 0.0863) Gain of MoRF binding (P = 0.0972) Gain of methylation at K1010 (P = 0.1428) Loss of catalytic residue at T1012 (P = 0.1574)	Loss of phosphorylation at T1012 (P = 0.0437)
R468C	0.725 (Harmful)	Loss of MoRF binding (P = 0.0099) Gain of catalytic residue at S472 (P = 0.0714) Loss of methylation at R468 (P = 0.0798) Loss of phosphorylation at S470 (P = 0.0953) Loss of disorder (P = 0.1277)	Loss of MoRF binding (P = 0.0099)
R166C	0.717 (Harmful)	Loss of MoRF binding (P = 0.0099) Gain of catalytic residue at S170 (P = 0.0714) Loss of phosphorylation at S168 (P = 0.0953) Loss of methylation at R166 (P = 0.1086) Loss of disorder (P = 0.1277)	Loss of MoRF binding (P = 0.0099)
R338C	0.721 (Harmful)	Loss of MoRF binding (P = 0.0099) Gain of catalytic residue at S342 (P = 0.0714) Loss of methylation at R338 (P = 0.0875) Loss of phosphorylation at S340 (P = 0.0953) Loss of disorder (P = 0.1277)	Loss of MoRF binding (P = 0.0099)
Y994H	0.579 (Harmful)	Gain of disorder (P = 0.0379) Gain of catalytic residue at L996 (P = 0.0863) Loss of MoRF binding (P = 0.1319) Loss of phosphorylation at Y994 (P = 0.1525) Loss of helix (P = 0.1706)	Gain of disorder (P = 0.0379)
Y692H	0.542 (Harmful)	Gain of disorder (P = 0.0379) Gain of catalytic residue at L694 (P = 0.0863) Loss of MoRF binding (P = 0.1319) Loss of phosphorylation at Y692 (P = 0.1525) Loss of helix (P = 0.1706)	Gain of disorder (P = 0.0379)
Y864H	0.681 (Harmful)	Gain of disorder (P = 0.0379) Gain of catalytic residue at L866 (P = 0.0863) Loss of MoRF binding (P = 0.1319) Loss of phosphorylation at Y864 (P = 0.1525) Loss of helix (P = 0.1706)	Gain of disorder (P = 0.0379)

Table 3. Shows prediction of disease related non synonymous SNPs by PhD-SNP and SNPs & GO

Amino Acid Change	PhD-SNP Prediction	RI	probability	SNPs & GO Prediction	RI	Probability
S580L	NEUTRAL	1	0.464	NEUTRAL	7	0.142
S450L	NEUTRAL	1	0453	NEUTRAL	7	0.139
R1217K	DISEASE	5	0.573	NEUTRAL	1	0.454
R915K	DISEASE	5	0.754	NEUTRAL	1	0.454
R1087K	DISEASE	5	0.753	NEUTRAL	1	0.454
R772W	DISEASE	7	0.842	DISEASE	4	0.700
R470W	DISEASE	7	0.843	DISEASE	4	0.701
R642W	DISEASE	7	0.842	DISEASE	4	0.701
Y850C	DISEASE	4	0.704	NEUTRAL	0	0.488
Y548C	DISEASE	4	0.688	NEUTRAL	1	0.474
Y720C	DISEASE	4	0.703	NEUTRAL	0	0.487
R772Q	DISEASE	7	0.827	DISEASE	3	0.635

Amino Acid Change	PhD-SNP Prediction	RI	probability	SNPs & GO Prediction	RI	Probability
R470Q	DISEASE	7	0.827	DISEASE	3	0.636
R642Q	DISEASE	7	0.828	DISEASE	3	0.365
H1248D	DISEASE	7	0.781	DISEASE	5	0.738
H946D	DISEASE	7	0.870	DISEASE	5	0.735
H1118D	DISEASE	7	0.871	DISEASE	6	0.791
E1214A	DISEASE	5	0.738	DISEASE	5	0.744
E912A	DISEASE	5	0.741	DISEASE	5	0.745
E1084A	DISEASE	5	0.739	DISEASE	5	0.744
T1219I	DISEASE	8	0.884	DISEASE	5	0.764
T917I	DISEASE	8	0.882	DISEASE	5	0.759
T1089I	DISEASE	8	0.883	DISEASE	5	0.791
K1140R	DISEASE	4	0.717	DISEASE	6	0.782
K838R	DISEASE	4	0.722	DISEASE	6	0.784
K1010R	DISEASE	4	0.718	DISEASE	6	0.782
C463W	DISEASE	8	0.918	DISEASE	4	0.724
G1139S	DISEASE	8	0.884	DISEASE	6	0.825
G837S	DISEASE		0.885	DISEASE	7	0.825
G1009S	DISEASE	8	0.885	DISEASE	7	0.825
A1303T	DISEASE	7	0.861	DISEASE	6	0.776
A1001T	DISEASE	7	0.861	DISEASE	6	0.776
A1173T	DISEASE	7	0.861	DISEASE	6	0.777
E1193K	DISEASE	6	0.819	DISEASE	4	0.693
E891K	DISEASE	6	0.821	DISEASE	4	0.694
E1063K	DISEASE	6	0.819	DISEASE	4	0.693
M1267T	DISEASE	3	0.637	DISEASE	5	0.771
M965T	DISEASE	3	0.631	DISEASE	5	0.767
M1137T	DISEASE	8	0.895	DISEASE	8	0.885
L899F	DISEASE	5	0.728	DISEASE	1	0.556
L1071F	DISEASE	5	0.736	DISEASE	1	0.564
A1303G	DISEASE	5	0.728	DISEASE	1	0.535
A1001G	DISEASE	5	0.727	DISEASE	1	0.537
A1173G	DISEASE	5	0.729	DISEASE	1	0.538
H367N	DISEASE	0	0.500	NEUTRAL	7	0.142
H65N	DISEASE	0	0.500	NEUTRAL	7	0.144
H237N	NEUTRAL	0	0.498	NEUTRAL	7	0.141
A457P	NEUTRAL	3	0.359	NEUTRAL	7	0.130
A155P	NEUTRAL	3	0.363	NEUTRAL	7	0.132
A327P	NEUTRAL	3	0.347	NEUTRAL	8	0.125
T1142M	DISEASE	6	0.787	DISEASE	4	0.667
T840M	DISEASE	6	0.786	DISEASE	4	0.676
T1012M	DISEASE	6	0.786	DISEASE	4	0.676
R468C	DISEASE	2	0623	NEUTRAL	3	0358
R166C	DISEASE	3	0.629	NEUTRAL	3	0.364
R338C	DISEASE	3	0.625	NEUTRAL	3	0.360
Y994H	DISEASE	2	0.685	NEUTRAL	3	0.347
Y692H	DISEASE	4	0.685	NEUTRAL	3	0.347
Y864H	DISEASE	4	0.686	NEUTRAL	3	0.348

RI: Reliability Index, **Probability:** Disease probability (if >0.5 mutation is predicted Disease), NEUTRAL: neutral variation, DISEASE: disease associated variation

Table 4. Shows the impact of nsSNPs on protein stability by I-mutant and MUpro

Amino Acid Change	I-Mutant Prediction	RI	DDG	MUpro Prediction	Confidence Score
S580L	Decrease	2	-0.19	Increase	0.52749269
S450L	Decrease	2	-0.19	Increase	0.52749269
R1217K	Decrease	8	-0.66	Decrease	-0.94371603
R915K	Decrease	8	-0.66	Decrease	-0.94371603
R1087K	Decrease	8	-0.66	Decrease	-0.94371603
R772W	Decrease	4	-0.00	Increase	0.31427857
R470W	Decrease	7	0	Increase	0.31427857
R642W	Decrease	4	0	Increase	0.31427857
Y850C	Increase	2	-0.91	Decrease	-0.57556848
Y548C	Increase	2	-0.91	Decrease	-0.57556848
Y720C	Increase	2	-0.91	Decrease	-0.57556848
R772Q	Decrease	7	-0.53	Increase	0.43532108
R470Q	Decrease	7	-0.53	Increase	0.43532108
R642Q	Decrease	7	-0.53	Increase	0.43532108
H1248D	Decrease	2	-0.48	Increase	0.057844857
H946D	Decrease	2	-0.48	Increase	0.057844857
H1118D	Decrease	2	-0.48	Increase	0.057844857
E1214A	Decrease	9	-0.92	Decrease	-0.77337457
E912A	Decrease	9	-0.92	Decrease	-0.77337457
E1084A	Decrease	9	-0.92	Decrease	-0.77337457
T1219I	Decrease	4	-0.2	Decrease	-0.50547757
T917I	Decrease	4	-0.2	Decrease	-0.50547757
T1089I	Decrease	4	-0.2	Decrease	-0.50547757
K1140R	Decrease	2	-0.07	Increase	0.28512217
K838R	Decrease	2	-0.07	Increase	0.28512217
K1010R	Decrease	2	-0.07	Increase	0.28512217
C463W	Decrease	1	0.04	Decrease	-1
G1139S	Decrease	4	-1.15	Decrease	-0.99017034
G837S	Decrease	4	-1.15	Decrease	-0.99017034
G1009S	Decrease	4	-1.15	Decrease	-0.99017034
A1303T	Decrease	8	-0.86	Decrease	-1
A1001T	Decrease	8	-1.15	Decrease	-1
A1173T	Decrease	8	-0.86	Decrease	-1
E1193K	Decrease	6	-0.61	Decrease	-0.87111124
E891K	Decrease	6	-0.61	Decrease	-0.87111124
E1063K	Decrease	6	-0.61	Decrease	-0.87111124
M1267T	Decrease	8	-0.99	Decrease	-0.48382754
M965T	Decrease	8	-0.99	Decrease	-0.48382754
M1137T	Decrease	8	-0.99	Decrease	-0.48315069
L899F	Decrease	F	-0.84	Decrease	-1
L1071F	Decrease	5	-0.84	Decrease	-1
A1303G	Decrease	9	-1.62	Decrease	-1
A1001G	Decrease	9	-1.62	Decrease	-1
A1173G	Decrease	9	-1.62	Decrease	-1
H367N	Decrease	2	-0.47	Decrease	-0.44646271
H65N	Decrease	2	-0.47	Decrease	-0.44646271
H237N	Decrease	2	-0.47	Decrease	-0.446462
A457P	Decrease	0	-0.4	Decrease	-0.67780456
A155P	Decrease	0	-0.4	Decrease	-0.67780456

Amino Acid Change	I-Mutant Prediction	RI	DDG	MUpro Prediction	Confidence Score
A327P	Decrease	0	-0.4	Decrease	-0.67780456
T1142M	Decrease	2	-0.14	Increase	0.65604841
T840M	Decrease	2	-0.14	Increase	0.65604841
T1012M	Decrease	2	-0.14	Increase	0.65604841
R468C	Decrease	5	-1.15	Decrease	-1
R166C	Decrease	5	-1.15	Decrease	-1
R338C	Decrease	5	-1.15	Decrease	-1
Y994H	Decrease	3	-1.16	Decrease	-1
Y692H	Decrease	3	-1.16	Decrease	-1
Y864H	Decrease	3	-1.16	Decrease	-1

DDG: ΔΔG sign **DDG value:** DG (New Protein)-DG (Wild Type) in Kcal/mole, **SVM2 value**: DDG < 0: decrease stability, DDG >0 increase stability
Confidence score: <0 means the variant decreases the protein stability, a score >0 means the variant increases the protein stability.

Table 5. Prediction of stability effect on domain cores and domain-domains interfaces by ELASPIC

Amino Acid Change		ΔGwt	ΔGmut	ΔΔG
S580L	Core	51.6618	50.5871	-0.80729
S450L	Not found			
R1217K	Interface/MSH2ISO2	-114.323	-16.1015	-0.80729
	Interface/MSH3	-118.604	-12.9561	-0.80729
	Interface/MSH2			
	Interface /E9PHA6	-114.28	-18.7611	-0.80729
		-117.916	-16.7864	-0.80729
R915K	Interface /E9PHA6	-110.51	-15.9872	-0.498745
	Interface/MSH2	-110.51	-15.9872	-0.498745
R1087K	Core	365.782	364.934	2.34277
R772W	Core	352.115	352.762	0.438479
R470W	Core	391.33	391.898	1.9065
R642W	Not found			
Y850C	Core	352.201	353.583	1.00591
Y548C	Core	392.244	393.062	2.35951
Y720C	Not found			
R772Q	Core	193.348	193.279	0.443288
R470Q	Core	391.383	392.028	2.69696
R642Q	Not found			
H1248D	Interface/MSH2ISO2	-113.634	353.583	2.38482
	Interface/MSH3	-119.204	-17.012	0.841595
	Interface/MSH2	-113.495	-12.7809	1.10642
	Interface /E9PHA6	-115.405	-18.438	3.1506
H946D	Interface /E9PHA6	-109.962	-16.6652	1.63824
	Interface/MSH2	-109.962	-16.6652	1.63824
H1118D	Not found			
E1214A	Interface/MSH3	-116.623	-15.8175	0.0462715
E912A	Core	392.176	391.441	1.97575
E1084A	Not found			
T1219I	Interface/MSH2ISO2	-115.005	-16.6629	1.07129
	Interface/MSH3	-119.16	-12.6915	1.22503
	Interface/MSH2	-113.954	-18.422	1.23678
	Interface /E9PHA6	-112.954	-17.8123	-0.839946
T917I	Interface /E9PHA6	-109.89	-16.9491	0.821522
	Interface/MSH2	-109.89	-16.9491	0.821522

Amino Acid Change		ΔGwt	ΔGmut	ΔΔG
T1089I	Not found			
K1140R	Interface/MSH2	-114.255	-20.0906	0.564315
	Interface /E9PHA6	-113.545	-19.1065	0.90146
K838R	Core	390.618	389.218	1.93367
K1010R	Not found			
C463W	Core	391.953	427.732	0.645258
G1139S	Core	352.201	355.425	-0.302907
G837S	Core	392.243	398.001	0.532736
G1009S	Not found			
A1303T	Interface/MSH2ISO2	-114.739	-16.6662	1.17137
	Interface/MSH3	-119.16	-12.6867	-0.181608
	Interface/MSH2	-113.936	-18.1328	0.672278
	Interface /E9PHA6	-112.955	-17.8119	1.14587
A1001T	Interface/MSH2	-109.907	-16.7324	0.638454
	Interface /E9PHA6	-109.907	-16.7324	0.638454
A1173T	Not found			
E1193K	Core	351.754	357.314	1.02374
E891K	Core	392.069	391.393	0.230509
E1063K	Not found			
M1267T	Core	350.434	353.032	0.719642
M965T	Core	391.273	393.054	2.05559
M1137T	Interface/MSH2ISO2	-114.728	353.032	0.719642
	Interface/MSH3	-119.16	-16.8874	0.39065
	Interface/MSH2	-113.954	-12.6867	0.481375
	Interface /E9PHA6	-112.955	-18.4277	0.563091
L899F	Interface/MSH2	-109.907	-17.326	0.746877
	Interface /E9PHA6	-109.907	-17.326	0.746877
L1071F	Not found			
A1303G	Interface/MSH2ISO2	-114.739	-16.6662	1.09202
	Interface/MSH3	-119.16	-12.6867	-0.297163
	Interface/MSH2	-113.936	-18.2323	-0.023302
	Interface /E9PHA6	-112.946	-17.8091	0.165379
A1001G	Interface/MSH2	-109.907	-16.7273	0.0340597
	Interface /E9PHA6	-109.907	-16.7273	0.0340597
A1173G	Not found			
H367N	Core	F		
H65N	Core	F		
H237N	Not found			
A457P	Core	123.65	120.66	0.0135101
A155P	Core	119.533	117.053	-0.187891
A327P	Not found			
T1142M	Not found			
T840M	Core	392.186	391.304	-0.91221
T1012M	Not found			
R468C	Core	123.65	124.196	1.03848
R166C	Core	120.35	120.781	0.805692
R338C	Core			
Y994H	Core	352.121	352.639	-0.531703
Y692H	Core	392.228	393.006	2.50887
Y864H	Not found			

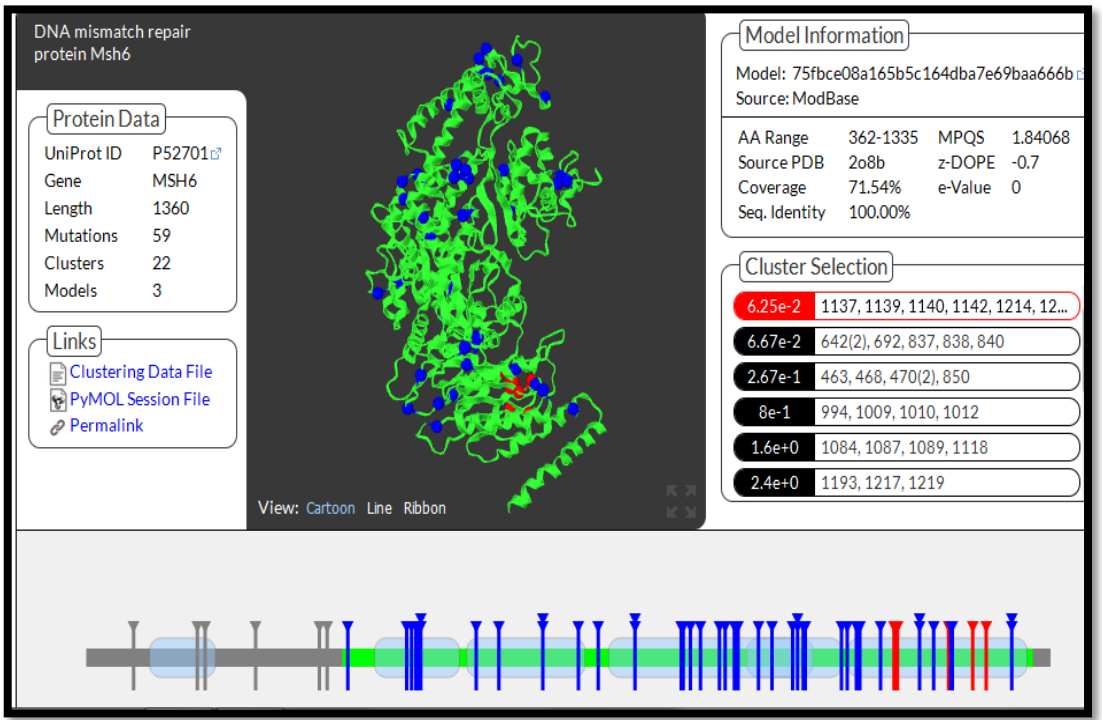

Figure 2. MSH6 protein 3d structure and distribution of highly damaging mutations in protein's domains

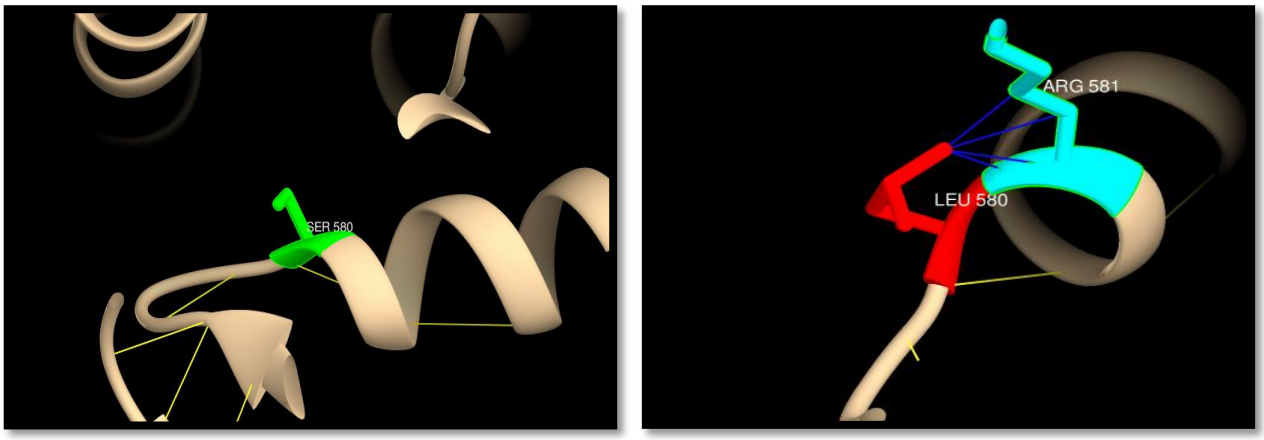

Figure 3. rs41295270 (S580L): Wild residue (green color), mutant residue (red color), hydrogen bond (yellow lines) and clash (blue lines). One hydrogen bond interaction in wild and mutant residue, 4 clashes between mutant residue and ARG581

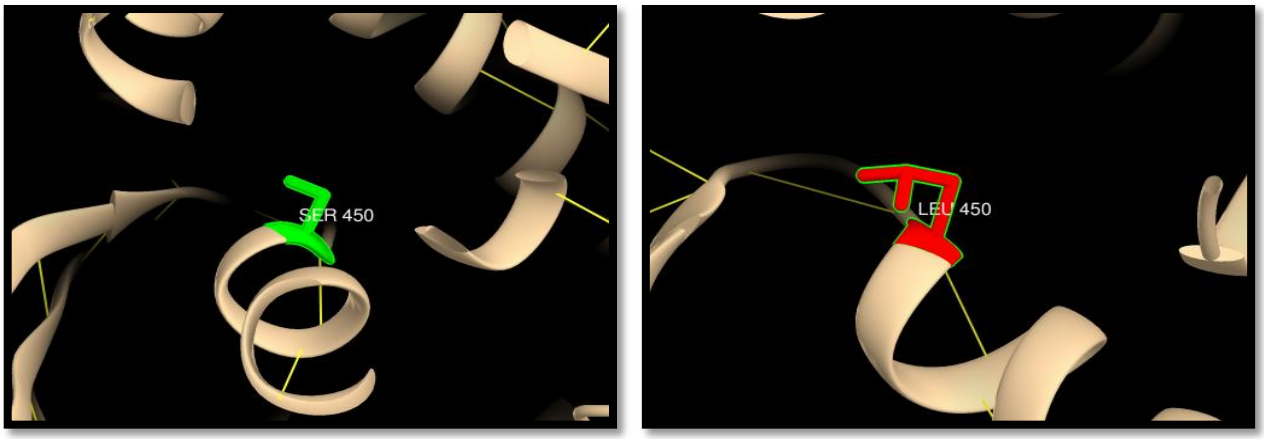

Figure 4. rs41295270 (S450L) (MSH6 ISOFORM2): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow lines). One hydrogen bond interaction in wild and mutant residue

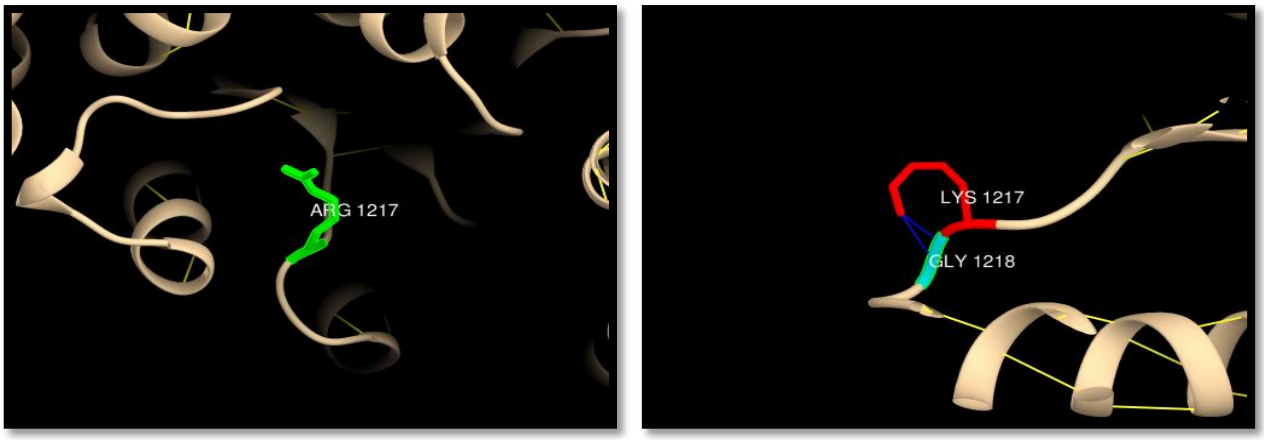

Figure 5. rs63749898 (R1217K): wild residue (green color), mutant residue (red color) and clash (blue lines). 2 clashes between mutant residue and GLY 1218

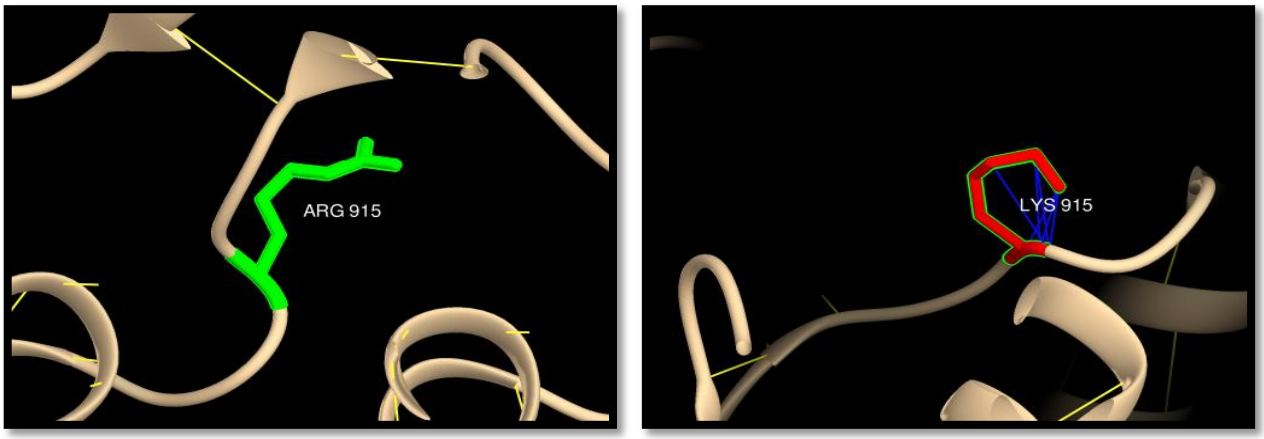

Figure 6. rs63749898 (R9157K) (MSH6 ISOFORM4): wild residue (green color), mutant residue (red color), and clash (blue lines). 6 clashes of mutant residue with itself

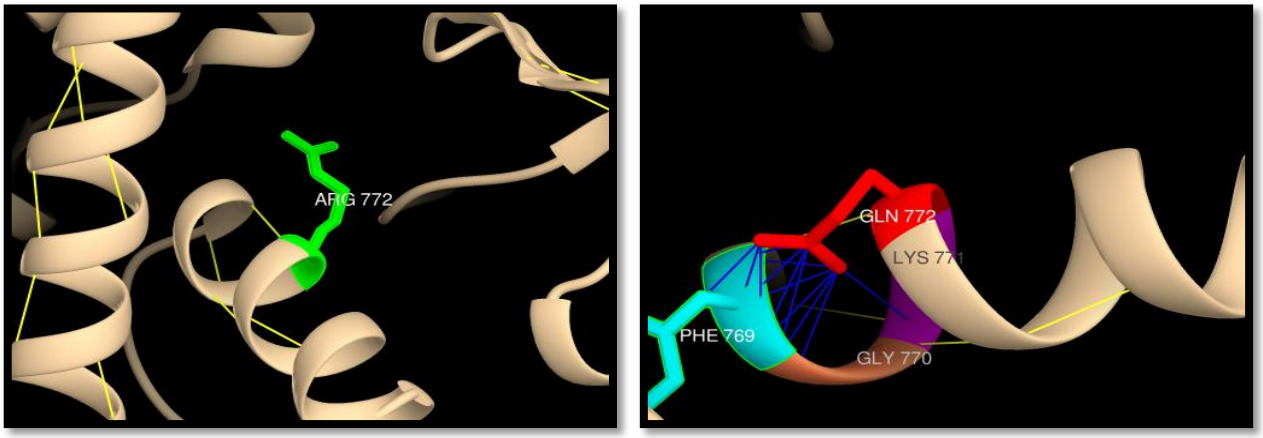

Figure 7. rs63750725 (R772Q): wild residue (green color), mutant residue (red color), hydrogen bond (yellow lines) and clash (blue lines). One hydrogen bond interaction in wild and mutant residue, one clash between mutant residue and LYS771,one clash with GLY770 and 11 clashes with PHE769

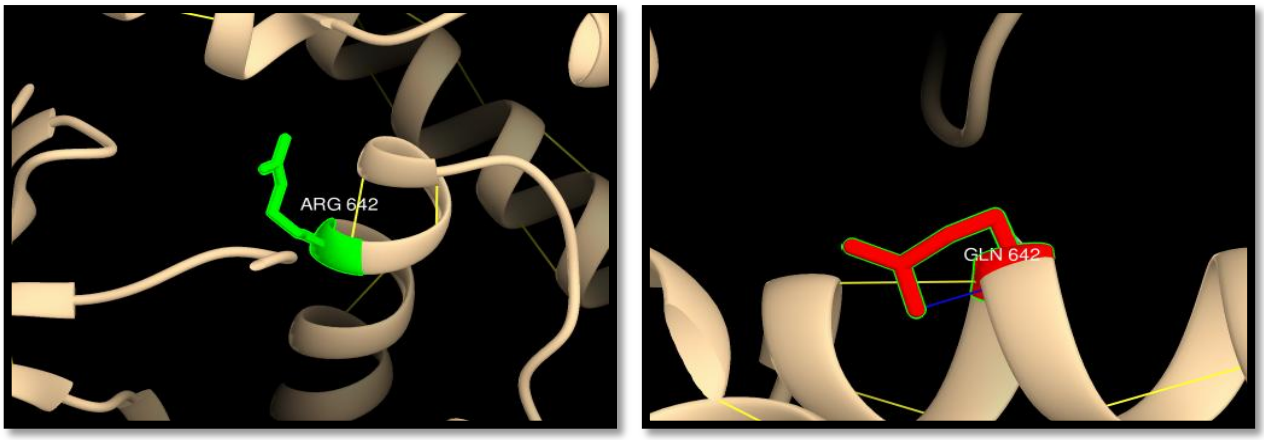

Figure 8. rs63750725 (R642Q) (MSH6 ISOFORM3): wild residue (green color), mutant residue (red color), hydrogen bond (yellow lines) and clash (blue lines). One hydrogen bond interaction in wild and mutant residue, one clash of mutant residue with itself

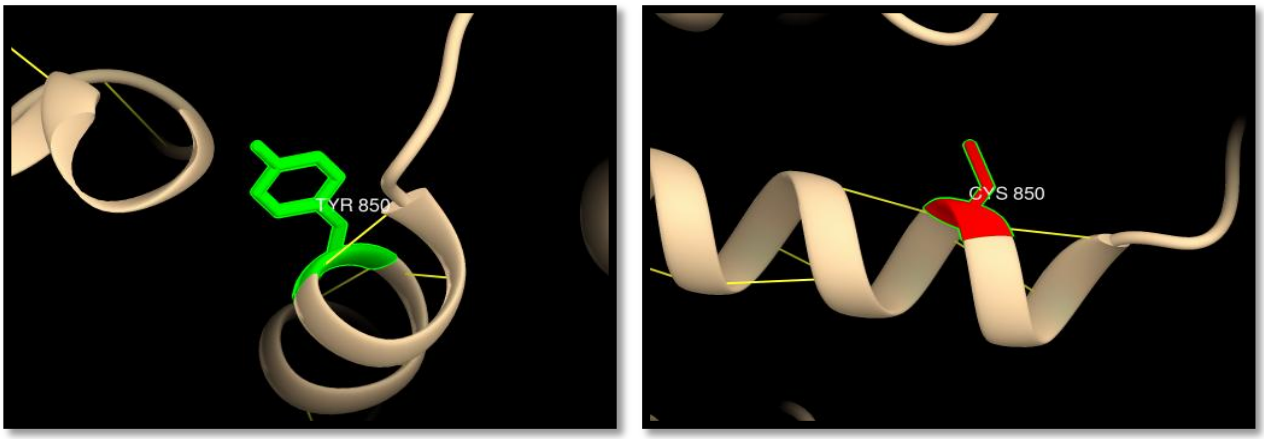

Figure 9. rs63750389; (Y850C): wild residue (green color), mutant residue (red color), hydrogen bond (yellow lines) and clash (blue lines). 2 hydrogen bonds interaction in wild and mutant residue

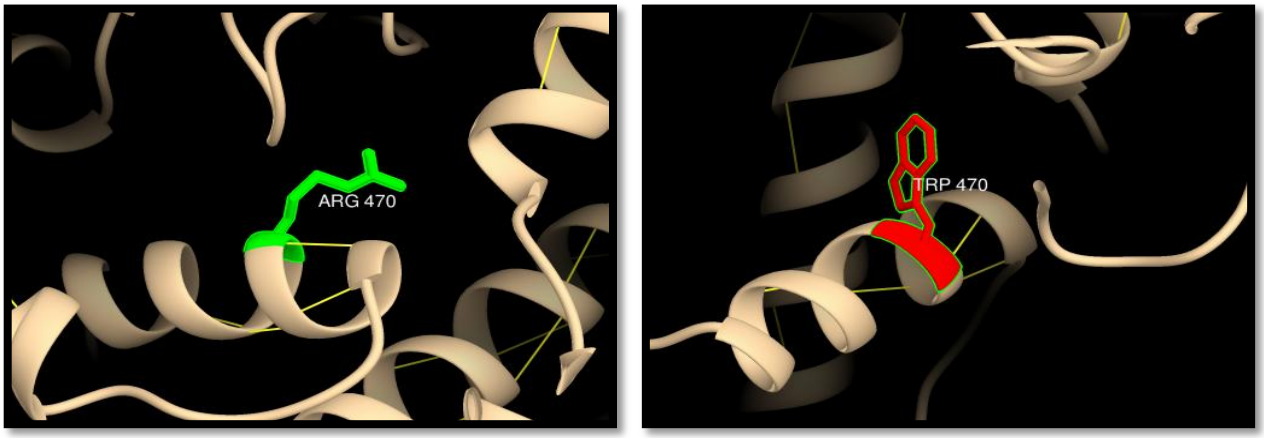

Figure 10. rs63750138 (R470W) (MSH6 ISOFORM4): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow lines). One hydrogen bond interaction in wild and mutant residue

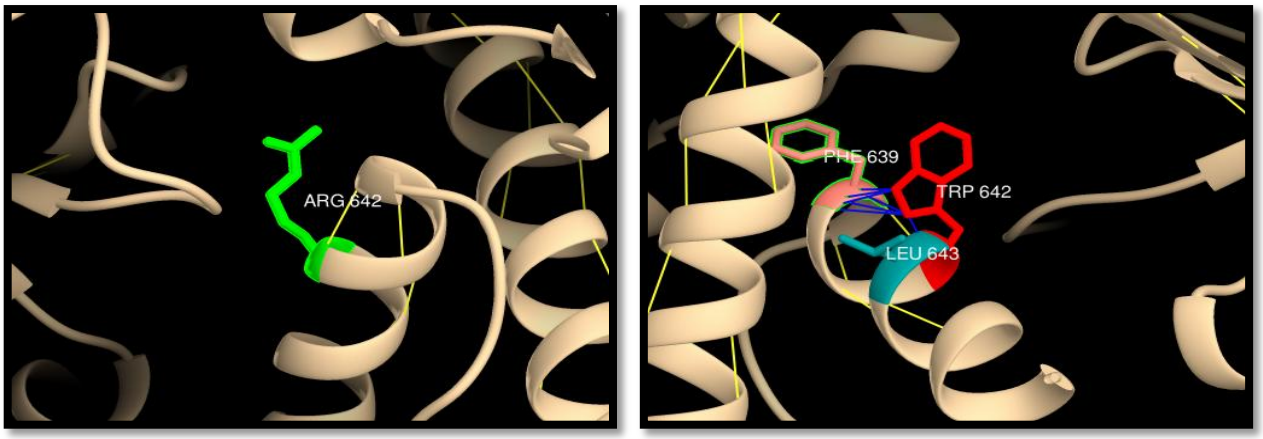

Figure 11. rs63750138 (R642W) (MSH6 ISOFORM3): wild residue (green color), mutant residue (red color), hydrogen bond (yellow lines) and clash (blue line) One hydrogen bond in wild and mutant residue, one clash of mutant residue with LEU643 and 6 clashes with PHE639

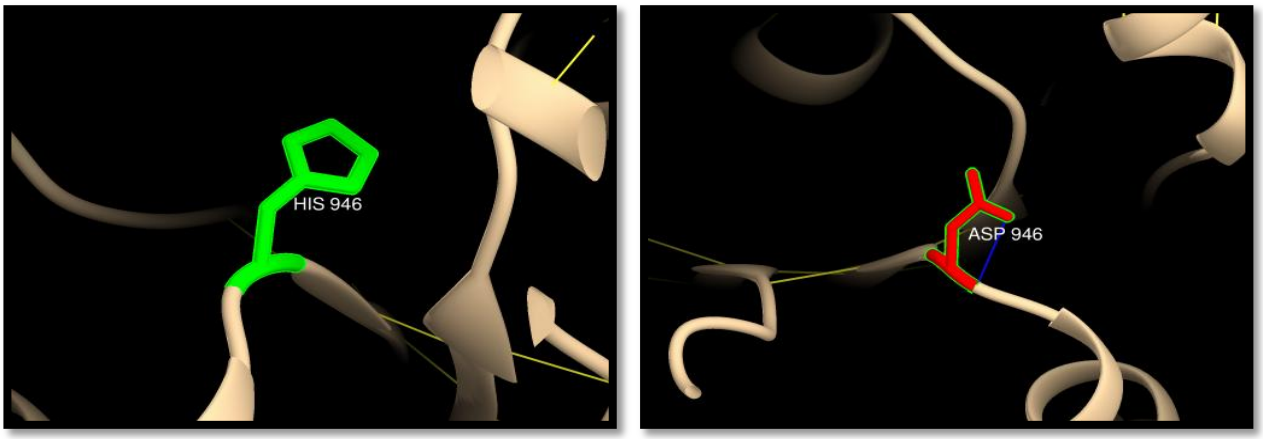

Figure 12. rs63750882 (H946D) (MSH6 ISOFORM4): wild residue (green color), mutant residue (red color) and clash (blue line). One clash of mutant residue with itself

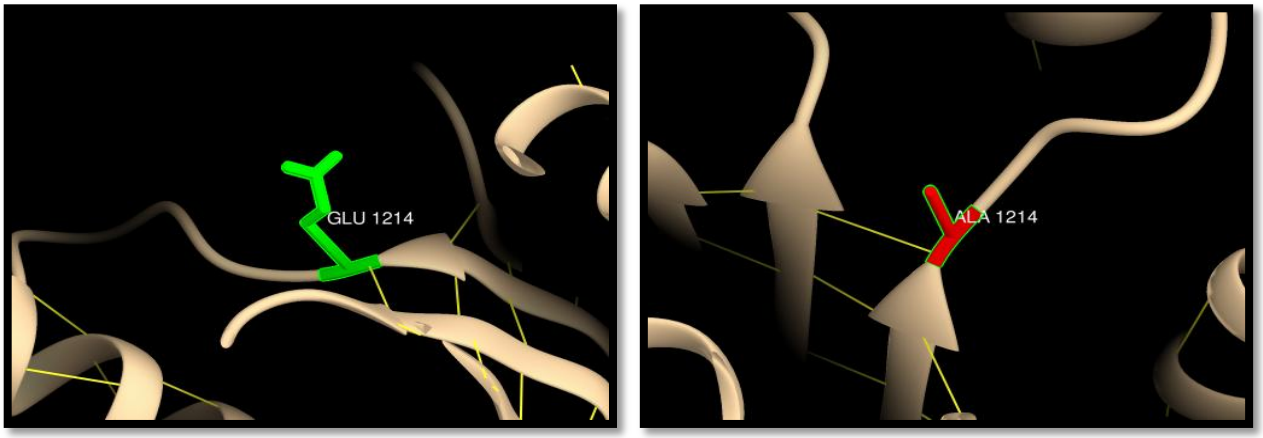

Figure 13. rs63750914 (E1214A): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow lines). One hydrogen bond interaction in both wild and mutant residue

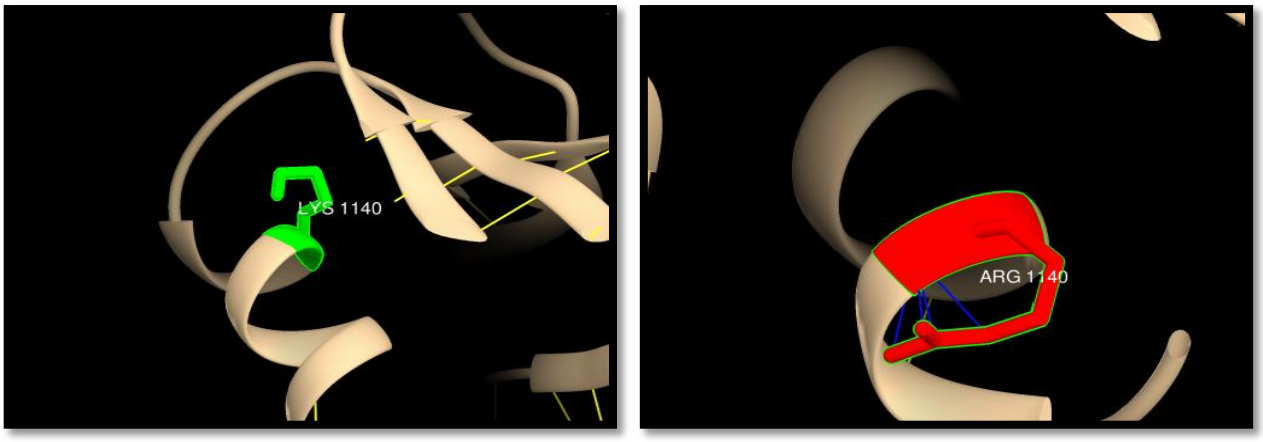

Figure 14. rs63750969 (K1140R): wild residue (green color), mutant residue (red color) and clash (blue line). 4 clashes in mutant residue

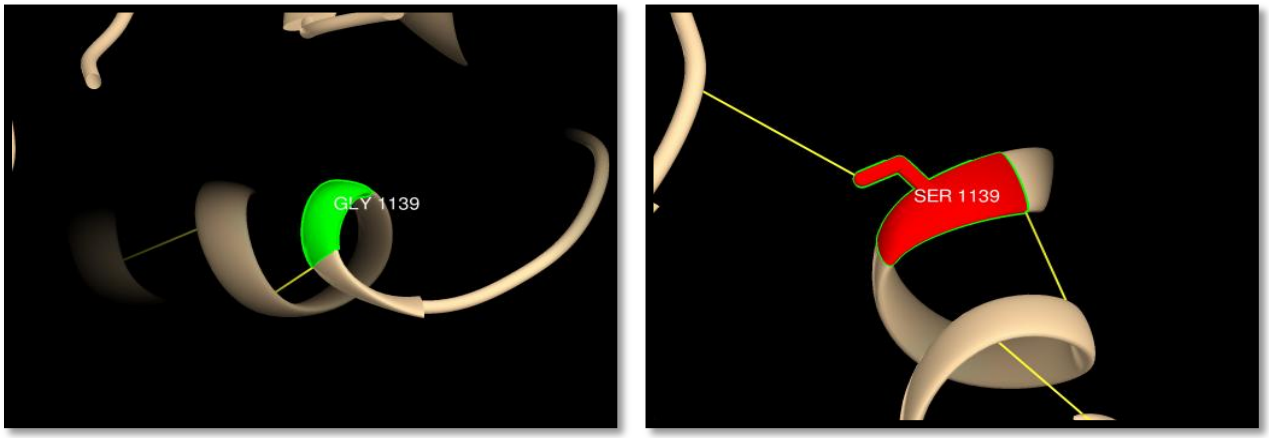

Figure 15. rs63751063 (G1139S): wild residue (green color), mutant residue (red color), hydrogen bond (yellow line). One hydrogen bond in wild residue and 2 bonds in the mutant

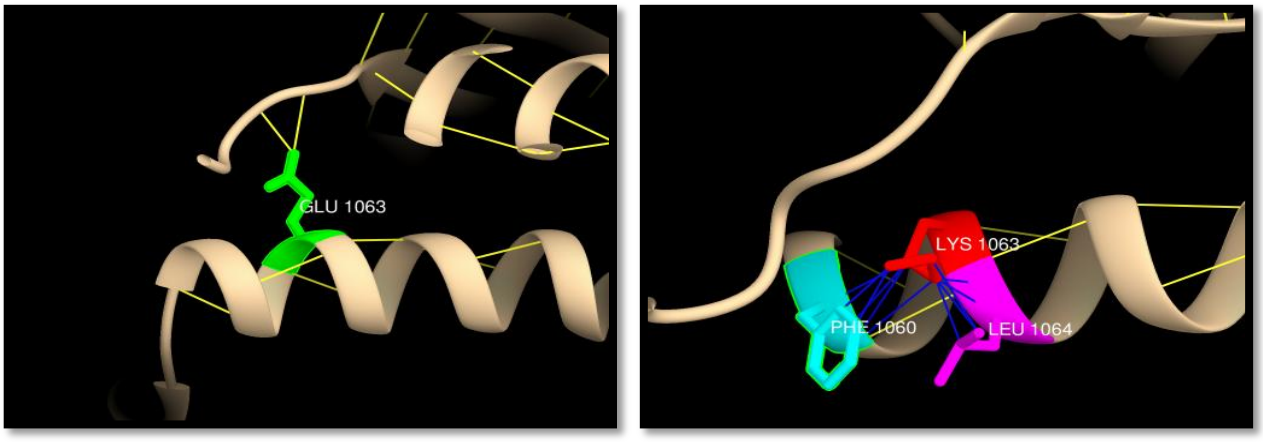

Figure 16. rs63751328 (E1063K) (MSH6 ISOFORM3) wild residue (green color), mutant residue (red color), clash (blue line) and hydrogen bond (yellow line). 3 hydrogen bonds in wild residue and 2 in mutant,6 clashes of mutant residue with PHE1060 and 5 clashes with LEU1064

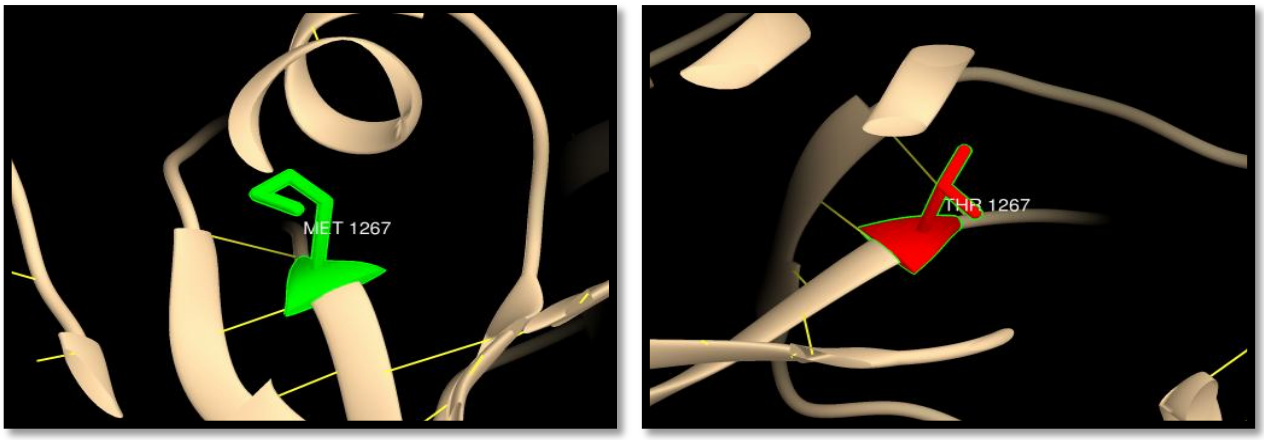

Figure 17. rs148445930 (M1267T): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow line). One hydrogen bond in both wild and mutant residue

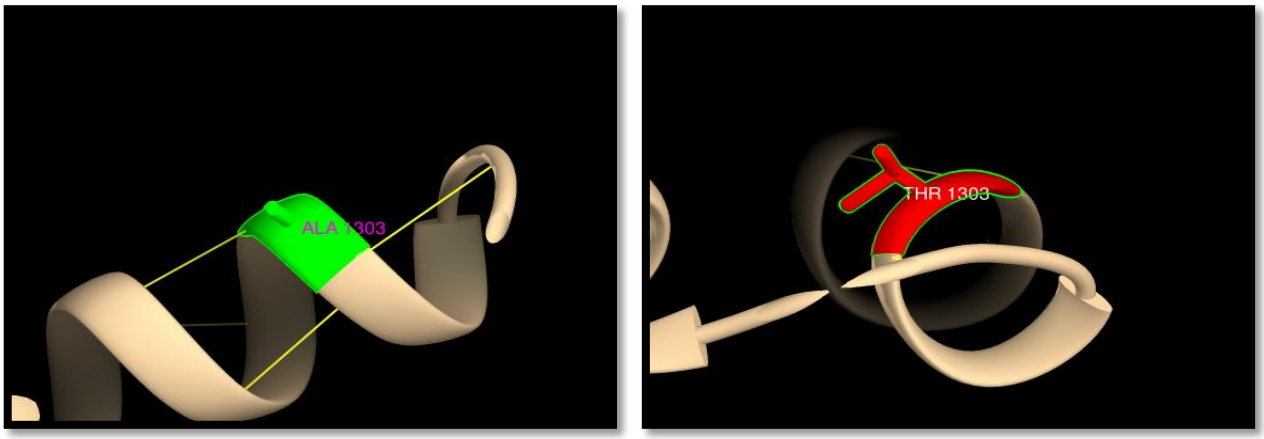

Figure 18. rs201060668 (A1303G): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow line). One hydrogen bond in both wild and mutant residue

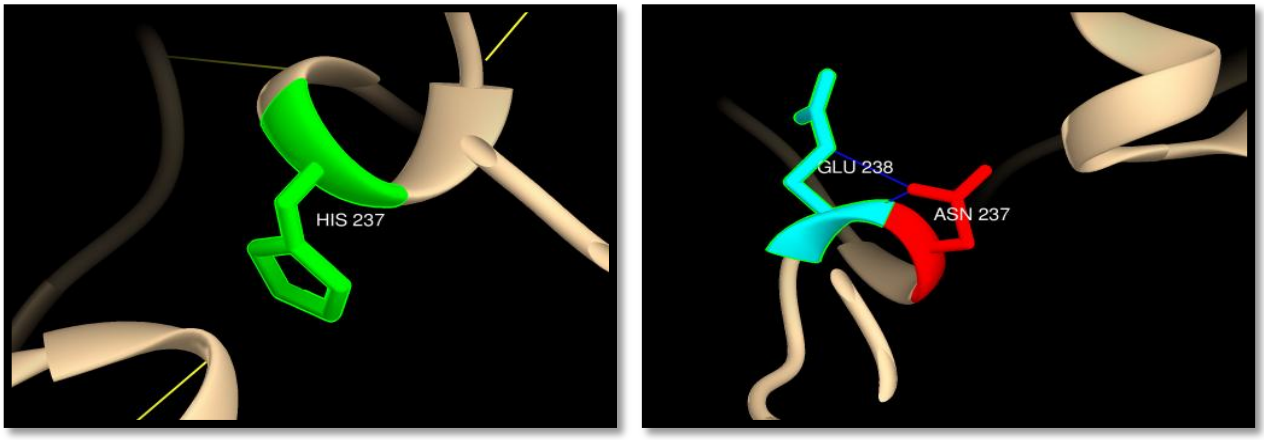

Figure 19. rs201193496 (H237N): wild residue (green color), mutant residue (red color) and clash (blue line). 2 clashes between mutant residue and GLU238

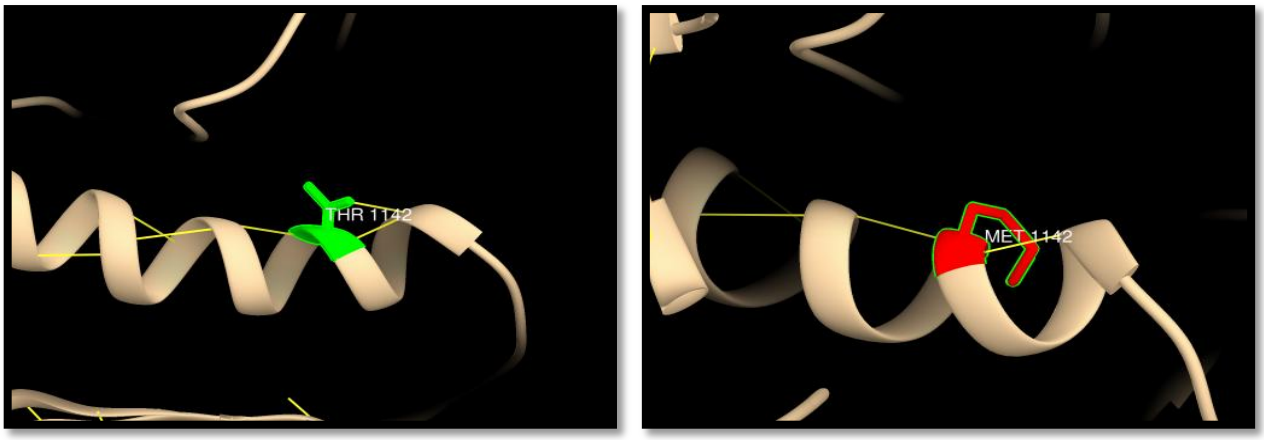

Figure 20. rs267608089 (T1142M): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow line). Three hydrogen bonds in wild residue and two bonds in mutant residue

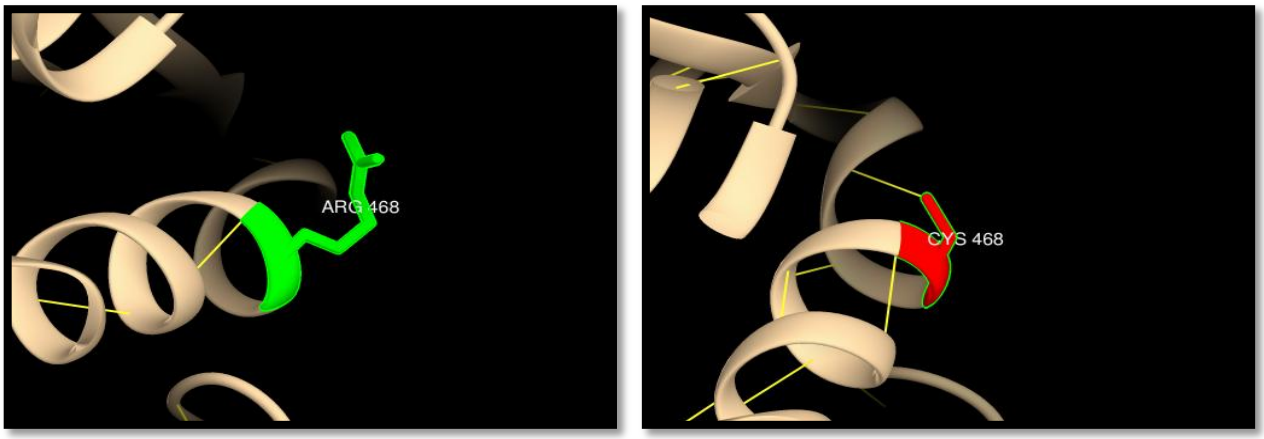

Figure 21. rs369456858 (R468C): wild residue (green color), mutant residue (red color) and hydrogen bond (yellow line). One hydrogen bond in wild residue and two bonds in mutant residue

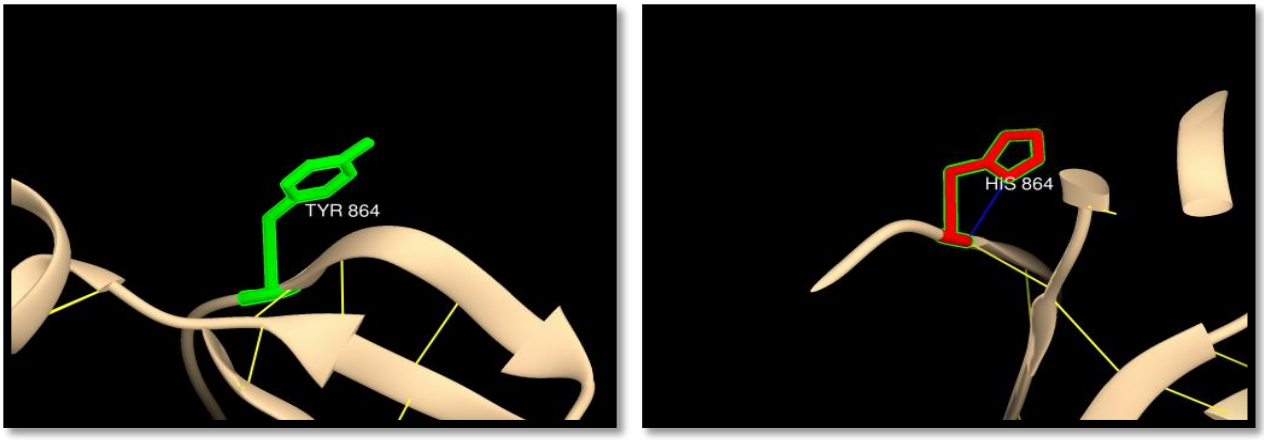

Figure 22. rs373622047 (Y864H) (MSH6 ISOFORM3): wild residue (green color), mutant residue (red color), hydrogen bond (yellow line) and clash (blue line). One hydrogen bond in wild and mutant residue, one clash in the mutant residue

Table 6. Shows the SNPs predicted by PolymiRTS to induce disruption or creation of mirRNA binding site

Location	dbSNP ID	Ancestral Allele	Allele	miR ID	Conservation	MiRSite	Function Class	context+ score change
48034022	rs200412142	A	A	hsa-miR-330-3p	5	GCTTTGAgttgac	D	-0.093
				hsa-miR-888-5p	5	GcTTTGAGTtgac	D	-0.091
			G	hsa-miR-186-3p	5	GCTTTGGGttgac	C	-0.174
48034033	rs184571812	G	C	hsa-miR-1253	3	ACTTCTCAcaaag	C	-0.105
				hsa-miR-6770-5p	3	ACTTCTCAcaaag	C	-0.107
				hsa-miR-758-3p	12	ActtcTCACAAAg	C	-0.1

"D": the derived allele disrupts a conserved microRNA site, **"C":** the derived allele creates a new microRNA site.

3.10. Influence of SNPS at the 3UTR on miR Binding Sites by PolymiRTS Database

Regarding the analysis of SNPs at the 3`UTR region using PolymiRTS database, out of 63 SNPS 2 functional SNPs were predicted to affect miRSite; rs200412142 was found to has 2 alleles (A and G) contained 2 (D) and 1(C) functional classes on 3 miRSite while rs184571821 SNP had 2 alleles (C) contained 3(C) functional class had 3 miRSites; (D) functional class disrupts a conserved miRNA site while (C) is a target binding site that can create a new microRNA site, Table (6).

4. Discussion

The human MSH6 protein was first reported in 1995 as G/T mismatch Binding Protein (GTBP), binding partner of hMSH2 to form the MutSα complex. The hMSH6 gene product is a 160 kDa protein that is unstable without heterodimerization with hMSH2, and consequently utilizes 80%–90% of available hMSH2. MSH2 and MSH6 share five similar domains, but with sufficient differences to give MSH6 several distinct functions. MSH6 also has a unique N-terminal disordered domain that is absent in its MSH2 partner. The hMutSα heterodimer binds to DNA mispairs and short insertion deletion loops (IDLs) [13, 60-64]. Mutations of MSH6 gene had been reported to be associated with many cancers but mainly with HNPCC syndrome, and to date the exact mechanism of how these mutations promote tumor genesis remains controversial, in the present study we presented a computational analysis for the reported MSH6 SNPs using several public softwares and databases in an attempt to understand how do these mutations affect the protein structure and function and hence promote a disease. To sort out tolerant from intolerant nsSNPs ten different prediction algorithm were used; SIFT, Polyphen, PHD-SNPs, SNP&GO, I-Mutant, MUpro, Mutpred, ELASPIC, Mutation 3D and USCF Chimera. GENEMANIA was used to investigate MSH6 gene interactions and its role in net works. PolymiRTs database was used to analyze the impact of SNPs at the 3UTR on microRNAs binding sites. Our findings showed 108 nsSNPs (288 mutations) out of 937 nsSNPs were predicted to be deleterious by SIFT, and 21 nsSNPs (59 mutations) were predicted to be highly damaging by SIFT and Polyphen (table 1,9) and those were selected for further analysis by other insilico tools. 19 nsSNPs (53 mutations) were predicted to be disease related by PHD-SNPs while SNP&GO predicted 14 SNPs (39mutations) to be disease related (table 3). The differences in prediction capabilities refer to the fact that every prediction algorithm uses different sets of sequences and alignments. Mutpred was used to determine the tolerance degree for each amino acid substitution on the basis of physio-chemical properties, the results showed 41 mutations were highly harmful (table 2), Gain of ubiquitination and methylation for the mutation (R→K) at positions 1217,915 and 1087, Loss of solvent accessibility for the mutation (Y→C) at positions 850, 548 and 720, loss of ubiquitination and gain of methylation for the mutation (K→R) at positions 1140, K838 and 1010, Gain of catalytic residue at P466 for the mutation C463W, Gain of ubiquitination for the mutation (E→K) at positions 1193, 891 and 1063, Loss of catalytic residue for the mutation at positions (M→T) 1267 and 965, loss of stability for the mutation (A→G) at positions 1001 and 1173, Loss of catalytic residue for the mutation (A→P) at position 155, Loss of phosphorylation for the mutation (M→T) at positions 1012 and 840, Loss of MoRF binding for the mutation (R→C) at positions 468, 166 and 388, Gain of disorder for the mutation (Y→H) at positions 994, 692 and 864. These results indicate that some nsSNPs account for potential structural and functional changes in MSH6 protein. For further confirmation nsSNPs were submitted to I-Mutant and MUpro and the findings were; In I-Mutant 3.0, 20 nsSNPs (56 mutations) (S→L, R→K, R→W, R→Q, H→D, E→A, T→I, K→R, C→W, G→S, A→T, E→K, M→T, L→F, A→G, H→N, A→P, T→M, R→C, Y→H) decreased the effective stability of the protein while in MUpro 15 nsSNPs (42mutations) were found to decrease the protein stability (R→K, E→A, T→I, C→W, G→S, A→T, E→K, M→T, L→F, A→G, H→N, A→P, R→C, Y→H, Y→C). ELASPIC server was used to classify nsSNPS at the core or interface of the protein. The results showed 26 mutations (S580L, R1087K, R770W, Y850C, G1139S, E1193K, M1267T, H367N, A457P, R468C, R338C, Y994H, T1142M,

R470W, Y548C, R470Q, E912A, K838R, C463W, G837S, E891K, R166C, T840M, R772Q, M965T, Y692H) were in the core of MSH6 protein and 14 mutations (R1217K, H1248D, E1214A, T1219I, K1140R, A1303T, M1137T, A1303G, R915K, H946D, T917I, A1001T, L899F, A1001G) were at the interface. The "core" residues are defined as residues which are exposed in the monomeric protein but buried in the protein complex. Core residues are typically hydrophobic with a composition strongly divergent from the composition of the remainder of the protein surface [65]. Core residues supply the bulk of the energy driving association by hydrophobic interactions [66]. The hydrophobic interactions within the complex cause the core region to become tightly packed upon complex association with little room for conformational variability. For these reasons, the core residues are strongly conserved during evolution [67] and mutations in this region are usually more strongly unfavorable when compared to mutations at the periphery of the interface. In cancer, 3D location of mutations at an interface has served as evidence that protein interactions may be important for metastasis site determination. Mutation3D was used to investigate the distribution of nsSNPs in MutS domains. NsSNPs were distributed in 5 domains of MSH6 protein, MutS I (R470W, R470Q, S540L, C463W, R468C, A 457P), MutS II (R642W, R642Q, S58L,Y548C,Y692H), MutS III (H946D, R915K, Y850C, R772W, R772Q, E912A, K1010R,A100T, M965T, T917I, L899F, E891K, K838R,T1012M, A100G, Y994H, Y864H, T840M), MutS V (H1248D, E1214A, H1118D, R1087K, R1217K, A1303T, M1267T, T1219I, E1193K, A1173T, M1137T, K1140R, E1084A, T1089I, A 1303G, A1173G, T1142M), PWWP (R166C, A155P) while H65N, H237N, A327P and R338C were uncovered mutations, E1063K and L1071F were located at inter domain regions. Mutations within the protein domains are considered high risk mutations which may lead to disturbance or loss of the protein function. 3D structure of MSH6 protein, H bond and clash was shown between wild type and mutant using the visualized chimera program. G1139S, E1193K, E891K, E1063K, T1142M, R468C, Y864H had difference in number of hydrogen bonds between wild and mutant residue indicating that these mutation will disturb the stability of the protein. Cash was detected in the mutant residue (S580L, R1217K, R915K, R772Q, R642Q, R642W, H946D, K1140R, E1063K, H237N, Y864H) indicating a change in the environment of the molecule and hence a change in the structure and function of the protein.

rs41295270 (S580L and S450L): The residue is located in the core of the protein and mutation of this residue can disturb interactions with other molecules or other parts of the protein. The mutant residue is located near a highly conserved position. The wild-type and mutant amino acid differ in size. The mutant residue is bigger than the wild-type and more hydrophobic, the size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. Also difference in hydrophobicity will affect hydrogen bond formation. GSK3 phosphorylation site (MOD_GSK3_1) motif predicted to be at this position is damaged by the mutation, only serine, threonine and tyrosine residues can be phosphorylated and mutation into another residue type will disturb this modification. The mutant residue has 4 clashes with ARG 581 which will change the environment of the molecule. The mutation will cause Loss of disorder (P=0.0825), gain of catalytic residue at S580 (P=0.1945), gain of methylation at R583 (P=0.2312), gain of helix (P = 0.2684), gain of catalytic residue at S450 (P = 0.1945) and gain of methylation at R453 (P = 0.2634).

rs63749898 (R1217K, R915K and R1087K): The mutant residue is located near a highly conserved position. The mutant residue is smaller than the wild-type and this may cause a possible loss of external interactions. The wild-type residue forms a hydrogen bond with: glutamic acid at position 1214, glycine at position 1216, the size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did, also difference in hydrophobicity will affect hydrogen bond formation. PKA Phosphorylation site (MOD_PKA_2) motif predicted to be at this position is damaged by the mutation. The mutant residue has 2 clashes with GLY1218. The mutation will cause gain of disorder (P = 0.1598).

rs63750138 (R772W, R470W and R642W): The mutant residue is bigger and more hydrophobic than the wild residue the charge of the buried wild-type (positive) residue is lost by this mutation. The wild-type and mutant amino acids differ in size. The mutant residue is bigger than the wild-type which is buried in the core of the protein but the mutant residue probably will not fit. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturbs correct folding. The mutant residue has 1 clash with LEU643 and 6 clashes with PHE639. The mutation will cause gain of loop (P=0.2045) and loss of MoRF binding (P = 0.2336). This mutation matches a previously described variant (VAR_043958) of Hereditary non-polyposis colorectal cancer 5 (HNPCC5) [MIM: 614350] which is annotated with the severity of the disease.

rs63750389 (Y850C, Y548C and Y720C). The wild-type residue forms a hydrogen bond with: serine at positions 564,262,434 and valine at positions 594,292,464. The wild-type residue forms a salt bridge with: serine at position 564,262,434, valine at position 592,290,464 glutamic acid at position 847,545,717, threonine at position 849,547,719. The wild-type residue is predicted (by KMAD) to be a phosphorylation site and only serine, threonine and tyrosine residues can be phosphorylated and mutation into another residue type will disturb this modification. The mutant residue is smaller than the wild-type residue and this difference will cause an empty space in the core of the protein. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding. This mutation matches a previously described variant (VAR_012963) associated with HNPCC5 and Colorectal

cancer, reported by Ying et al (1999) in Chinese patients with hereditary non polyposis colorectal cancer [68].

rs63750725 (R642W, R772W and R470W). There is a difference in charge and size between the wild-type (positively charged) and mutant (smaller and neutral) amino acid. The wild-type residue forms a hydrogen bond with: glutamic acid at position 1084, 1214 and glycine at position 1086, 1216. The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The mutant residue has 1 clash with LEU643 and PHE639. The mutation will cause gain of loop (P=0.2045) and loss of MoRF binding (P = 0.2336).

rs63750882 (H1248D, H946D and H1118D). There is a difference in charge between the wild-type (neutral) and mutant (negatively charged) amino acid and this can lead to protein folding problems. The mutant residue is smaller than the wild-type residue and this will cause an empty space in the core of the protein. According to the PISA-database, the mutated residue is involved in a multimer's contact; the mutation introduces a smaller residue at this position. The new residue might be too small to make multimer contacts. The motif NEK2 phosphorylation site (MOD_NEK2_1) is damaged by the mutation. The mutant residue at position 946 has 1 clash. The mutation will cause loss of helix (P = 0.1706), loss of disorder (P = 0.1902), gain of loop (P = 0.2045) and loss of sheet (P = 0.302).

rs63750914 (E1214A, E912A and E1084A). The wild-type residue was negatively charged while the mutant residue is neutral and smaller than the wild-type. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturbs correct folding. The wild-type residue forms a hydrogen bond with: arginine at positions 772,470,642. The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The mutation will cause loss of stability (P = 0.2868), loss of sheet (P = 0.302) and Loss of disorder (P = 0.3528).

rs63750949 (T1219I, T917I, and T1089I). There is a difference in size and hydrophobicity between the wild-type and mutant (bigger and more hydrophobic) amino acid. The wild-type residue is predicted (by KMAD) to be a phosphorylation site and mutation into another residue type will disturb this modification. In both the PDB-file and in the PISA-assembly, this residue was found to be involved in a multimer's contact. This is a strong indication that the residue is indeed in contact with other proteins. The mutation introduces a bigger residue at these positions; this can disturb the multimeric interactions. A more hydrophobic residue is introduced and any hydrogen bond that could be made by the wild-type residue to other monomers will be lost and affect the multimeric contacts. The mutation will cause loss of helix (P = 0.3949).

rs63750969 (K1140R, K838R and K1010R): The wild-type residue has interactions with a ligand annotated as ADP The difference in properties between wild-type and mutation can easily cause loss of interactions with the ligand. Because ligand binding is often important for the protein's function, this function might be disturbed by this mutation. According to the PISA-database, the mutated residue is involved in a multimer contact. The mutations introduce bigger residues at these positions which can disturb the multimeric interactions. The wild-type residue is located in a region annotated in UniProt to form an α-helix. The mutations convert the wild-type residue in to residue that do not prefer α-helices as secondary structure.

rs63750985 (C463W): The mutant residue is bigger than the wild-type which is buried in the core of the protein so the mutant residue probably will not fit. The mutantion will cause gain of catalytic residue at P466 (P = 0.0237), gain of MoRF binding (P = 0.0768), loss of stability (P = 0.1456), loss of phosphorylation at T465 (P = 0.3476) and loss of disorder (P = 0.3669).

rs63751063 (G1139S, G837S and G1009S): The mutant residue is bigger than the wild-type residue. The mutatnt residue is not in direct contact with a ligand, however, the mutation could affect the local stability which in turn could affect the ligand-contacts made by one of the neighboring residues. These differences in properties between wild-type and mutant residue can easily cause loss of interactions with the nucleotide ("ATP"). This can directly affect the function of the protein. The wild-type residue forms a salt bridge with: glycine at positions 1138,836 and 1008.The wild-type residue forms a salt bridge with: glycine at these positions. The torsion angles for this residue are unusual. Only glycine is flexible enough to make these torsion angles, so mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure. The mutation will cause loss of helix (P = 0.1299), loss of disorder (P = 0.2501) and loss of stability (P = 0.2938).

rs63751064 (A1303T, A1001T and A1173T): There is a difference in hydrophobicity and size between the wild-type (more hydrophobic) and mutant (bigger) amino acid. The wild-type residue is located in a region annotated in UniProt to form an α-helix. The mutation converts the wild-type residue in a residue that does not prefer α-helices as secondary structure. WDR5 WD40 repeat (blade 5, 6)-binding ligand (LIG_WD40_WDR5_VDV_2) motif is damaged by the mutation. The mutation will cause loss of helix (P = 0.1299) and loss of stability (P = 0.1957) Loss of MoRF binding (P = 0.2535).

rs63751328 (E1193K, E891K and E1063K). There is a difference in charge and size between the wild-type (negatively charged) and mutant (positively charged and bigger) amino acid. The charge of the buried wild-type residue is reversed by this mutation; this can cause repulsion between residues in the protein core. The mutant residue is bigger than the wild-type residue which buried in the core of the protein so the mutant residue probably will not fit. The mutant residue has 6 clashes with PHE 1060 and 5 clashes with LEU1063. The mutation will cause loss of helix (P = 0.1299), loss of disorder (P = 0.2501) and loss of stability (P

= 0.2938). This mutation matches a previously described variant (VAR_043970), found in an endometrial cancer sample; displays marked impairment of heterodimerization with *MSH*2 and of in vitro mismatch repair capacity.

rs148445930 (M1267T, M965T and M1137T): The mutant residue is smaller than the wild-type residue and this might cause a possible loss of external interactions. The hydrophobicity of the wild-type and mutant residue differs and the mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein. The mutation will cause loss of sheet (P = 0.0817), loss of stability (P = 0.251), gain of disorder (P = 0.2718) and gain of loop (P = 0.2754).

rs182024561 (L899F and L1071F): The wild-type residue is highly conserved. The mutant residue is bigger than the wild-type residue is buried in the core of the protein so the mutant residue probably will not fit. The mutation will cause loss of stability (P = 0.2847), loss of disorder (P = 0.3187) and loss of helix (P = 0.3949).

rs201060668 (A1303G, A1001G and A1173G): The mutation introduces a glycine at these positions. Glycine is very flexible and can disturb the required rigidity of the protein at these positions. Mutant residue is smaller than the wild-type residue and this will cause a possible loss of external interactions. There is difference in hydrophobicity between the wild-type (more hydrophobic) and mutant residue and this might cause loss of hydrophobic interactions with other molecules on the surface of the protein. WDR5 WD40 repeat (blade5,6)-binding ligand (LIG_WD40_WDR5_VDV_2) is damaged by the mutation. The mutation will cause loss of stability (P = 0.0472), loss of helix A1000 (P = 0.1682) and loss of MoRF binding (P = 0.2546).

rs201193496 (H367N, H65N and H237N): The mutant residue is smaller than the wild-type residue and this will cause an empty space in the core of the protein. The new residue might be too small to make multimeric contacts. The residue is buried in the core of a domain. The differences between the wild-type and mutant residue might disturb the core structure of this domain. The mutant residue has 2 clashes with GLU238. The mutation will cause gain of relative solvent accessibility (P = 0.0249), gain of solvent accessibility (P = 0.0488), gain of loop (P = 0.0851) and loss of helix (P = 0.1299).

rs267608052 (A457P, A155P and A327P): There is difference in size between the wild-type and mutant residue. The mutant residue is bigger than the wild-type residue which buried in the core of the protein so the mutant residue probably will not fit. The wild-type residue forms a hydrogen bond with: aspartic acid. The size difference between them makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The mutation will cause Loss of stability (P = 0.1203), gain of sheet (P = 0.1945).

rs267608089 (T1142M, T840M and T1012M): The mutant residue is bigger than the wild-type residue which is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein. The mutant residue is more hydrophobic than the wild residue so any hydrogen bond that could be made by the wild-type residue to other monomers will be lost and affect the multimeric contacts. The difference in properties between wild-type and mutation can easily cause loss of interactions with the ligand. Because ligand binding is often important for the protein's function, this function might be disturbed by this mutation. The mutation may cause gain of MoRF binding (P=0.0972).

rs369456858 (468, 166 and 338): The mutant residue is smaller than the wild-type residue and this will cause a possible loss of external interactions. The charge of the wild-type (positive) residue is lost by this mutation which can cause loss of interactions with other molecules. The mutation may cause loss of MoRF binding (P = 0.0099) and loss of disorder (P = 0.1277).

rs373622047 (Y994H, Y692H and Y864H): The mutant residue is smaller than the wild-type residue and this will cause possible loss of external interactions. The wild-type residue is predicted to be located in its preferred secondary structure, a β-strand. The mutant residue prefers to be in another secondary structure; therefore the local conformation will be slightly destabilized. There is difference in the hydrophobicity between the wild-type and mutant residue. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein. The mutation might cause gain of disorder (P = 0.0379), loss of MoRF binding (P = 0.1319) and loss of helix (P = 0.1706).

MicroRNAs (miRNAs) are negative gene regulators acting at the 3'UTR level, modulating the translation of cancer-related genes. Single-nucleotide polymorphisms (SNPs) within the 3'UTRs could impact the miRNA-dependent gene regulation either by weakening or by reinforcing the binding sites. Thus, the alteration of the normal regulation of a given gene could affect the individual's risk of cancer [69]. It is helpful to predict which of the many SNPs could really impact the regulation of a target gene. We used polymiRTS database to predict the impact of SNPs at the 3UTR on micro RNAs binding sites and the results showed 2SNPs out of 63 at the 3UTR disrupt microRNAs binding sites and hence affect the gene expression. rs200412142 was found to has 2 alleles (A and G) contained 2 (D) and 1(C) functional classes on 3 miRSite while rs184571821 SNP had 2 alleles (C) contained 3(C) functional class had 3 miRSites; (D) functional class disrupts a conserved miRNA sites while (C) is a target binding site that can create a new microRNA sites. Recently micro RNAs have became the top important objects for revolutionary molecular research and many experiments had been conducted to assess the potentiality of using micrRNAs as targeted treatments for many cancers.

MSH6 is known of its DNA repairing function, yet GENMANIA revealed a possible role in B cell activation involved in immune response, production of molecular mediator of immune response, immunoglobulin production

and somatic recombination of immunoglobulin genes involved in immune response, suggesting vital roles in immune system which might be an interesting area for future research.

5. Conclusions

21 nsSNPS were predicted by different softwares to be the most damaging mutations for MSH6 protein altering physiochemical properties of the protein; size, charge and hydrophobicity leading to loss or disturbance of the protein internal and external interactions and eventually loss of the protein's function. 14 nsSNPs (R1217K, H1248D, E1214A, T1219I, K1140R, A1303T, M1137T, A1303G, R915K, H946D, T917I, A1001T, L899F, A1001G) were located at the interface of the MSH6 protein interfering with its relation with MSH2ISO2, MSH3, MSH2 and E9PHA6. Interactions of MSH6 with these proteins are critical for its MMR function and any structural alterations that interfere or harm these interactions would probably increase susceptibility to tumors formation and progression. 2 SNPs at the 3UTR; rs200412142 and rs184571821 introduced a change in the micro RNA binding site at the 3UT which might result in deregulation of the gene function.

ACKNOWLEDGMENTS

We thank the National Center for Neurological Sciences (NCNS) for the technical support.

Appendix

Table 7. Shows the genes co-expressed and sharing a domain with MSH6

Symbol	Description	Co-expression	shared protein
MSH2	mutS homolog 2	Yes	Yes
MSH3	mutS homolog 3	No	Yes
PCNA	proliferating cell nuclear antigen	Yes	No
SP3	Sp3 transcription factor	Yes	No
TSEN34	TSEN34 tRNA splicing endonuclease subunit	No	No
TOP3A	topoisomerase (DNA) III alpha	No	No
WDR76	WD repeat domain 76	Yes	No
MUTYH	mutY homolog	No	No
MSH4	mutS homolog 4	No	Yes
MSH5	mutS homolog 5	No	Yes
MSH5-SAPCD1	MSH5-SAPCD1 readthrough (NMD candidate)	No	Yes
ATR	ataxia telangiectasia and Rad3 related	Yes	No
WRN	Werner syndrome, RecQ helicase-like	No	No
DNA2	DNA replication helicase/nuclease 2	No	Yes
HYOU1	hypoxia up-regulated 1	No	No
MLH1	mutL homolog 1	No	No
RPA2	replication protein A2, 32kDa	No	No
PES1	pescadillo ribosomal biogenesis factor 1	No	No
AIRE	autoimmune regulator	No	No
ATP6V1B2	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B2	No	Yes

Table 8. Shows MSH6 functions and its appearance in network and genome

Feature	FDR	Genes in network	Genes in genome
mismatch repair complex binding	1.52E-12	6	10
mismatch repair	1.52E-12	7	22
DNA recombination	1.49E-10	9	151
DNA-dependent ATPase activity	6.34E-10	7	56
ATPase activity	1.01E-09	9	197

Feature	FDR	Genes in network	Genes in genome
reciprocal DNA recombination	1.91E-09	6	35
reciprocal meiotic recombination	1.91E-09	6	35
purine nucleoside monophosphate catabolic process	1.91E-09	8	139
structure-specific DNA binding	1.91E-09	8	142
purine ribonucleoside monophosphate catabolic process	1.91E-09	8	139
ATP catabolic process	1.91E-09	8	137
ribonucleoside monophosphate catabolic process	1.91E-09	8	139
nucleoside monophosphate catabolic process	1.91E-09	8	140
ATPase activity, coupled	1.91E-09	8	143
meiotic nuclear division	3.24E-09	7	84
ATP metabolic process	3.68E-09	8	158
meiotic cell cycle	3.68E-09	7	87
purine ribonucleoside monophosphate metabolic process	7.91E-09	8	177
purine nucleoside monophosphate metabolic process	7.91E-09	8	177
ribonucleoside monophosphate metabolic process	8.60E-09	8	180
nucleoside monophosphate metabolic process	1.06E-08	8	187
meiosis I	1.06E-08	6	50
nuclear chromosome	7.26E-08	8	239
double-stranded DNA binding	2.33E-07	6	84
regulation of DNA metabolic process	4.36E-07	7	180
DNA secondary structure binding	5.44E-07	4	13
cellular process involved in reproduction	1.07E-06	7	207
nuclear division	4.65E-06	7	257
organelle fission	8.52E-06	7	282
somatic diversification of immunoglobulins	1.34E-05	4	28
somatic diversification of immune receptors	2.58E-05	4	33
base-excision repair	3.09E-05	4	35
regulation of DNA recombination	3.09E-05	4	35
immunoglobulin production	3.78E-05	4	37
negative regulation of DNA metabolic process	2.02E-04	4	56
telomere maintenance	2.59E-04	4	60
telomere organization	2.70E-04	4	61
production of molecular mediator of immune response	4.08E-04	4	68
synaptonemal complex	4.43E-04	3	18
response to UV	9.08E-04	4	84
somatic recombination of immunoglobulin gene segments	9.11E-04	3	23
telomere maintenance via semi-conservative replication	1.02E-03	3	24
nuclear chromosome part	1.12E-03	5	209
telomere maintenance via recombination	1.24E-03	3	26
somatic cell DNA recombination	1.43E-03	3	28
somatic diversification of immune receptors via germline recombination within a single locus	1.43E-03	3	28
DNA replication	1.43E-03	5	222
nuclear cell cycle DNA replication	1.43E-03	3	28
DNA-dependent DNA replication	1.61E-03	4	102
double-strand break repair	2.29E-03	4	112

Feature	FDR	Genes in network	Genes in genome
DNA strand elongation involved in DNA replication	2.41E-03	3	34
mitotic recombination	2.41E-03	3	34
cell cycle DNA replication	2.58E-03	3	35
DNA strand elongation	3.00E-03	3	37
telomere maintenance via telomere lengthening	3.20E-03	3	38
condensed nuclear chromosome	3.68E-03	3	40
single-stranded DNA binding	7.11E-03	3	50
anatomical structure homeostasis	9.35E-03	4	166
PML body	9.68E-03	3	56
Aging	1.71E-02	3	68
response to ionizing radiation	1.99E-02	3	72
response to light stimulus	2.19E-02	4	210
immunoglobulin production involved in immunoglobulin mediated immune response	2.63E-02	2	13
meiotic chromosome segregation	2.63E-02	2	13
condensed chromosome	3.18E-02	3	86
DNA catabolic process	3.24E-02	3	87
postreplication repair	4.24E-02	2	17
Synapsis	4.24E-02	2	17
response to gamma radiation	4.24E-02	2	17
chromosome organization involved in meiosis	4.70E-02	2	18
immune system development	4.76E-02	4	266
ADP binding	5.04E-02	2	19
nucleotide-excision repair, DNA gap filling	5.04E-02	2	19
magnesium ion binding	5.19E-02	3	106
isotype switching	5.30E-02	2	20
somatic recombination of immunoglobulin genes involved in immune response	5.30E-02	2	20
somatic diversification of immunoglobulins involved in immune response	5.30E-02	2	20
response to radiation	5.34E-02	4	281
B cell activation involved in immune response	6.87E-02	2	23
nuclear replication fork	8.69E-02	2	26
DNA conformation change	9.65E-02	3	135
damaged DNA binding	9.85E-02	2	28

Table 9. Shows results of deleterious SNPs predicted by SIFT, Polyphen-2

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs728619	A/C	Y538S	ENSP00000234420	Deleterious	0،002	Possibly Damaging	0،845
rs728619	A/C	Y236S	ENSP00000438580	Deleterious	0،002	Possibly Damaging	0،682
rs728619	A/C	Y408S	ENSP00000446475	Deleterious	0،002	Possibly Damaging	0،933
rs3211299	G/T	S45I	ENSP00000390382	Deleterious	0،018	Probably Damaging	0،999
rs34374438	A/T	K552M	ENSP00000438580	Deleterious	0،004	Probably Damaging	1
rs34374438	A/T	K854M	ENSP00000234420	Deleterious	0،006	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs34374438	A/T	K724M	ENSP00000446475	Deleterious	0،007	Probably Damaging	1
rs41294988	A/T	K13T	ENSP00000234420	Deleterious	0،009	Possibly Damaging	0،948
rs41295268	G/A	R468H	ENSP00000234420	Deleterious	0،003	Probably Damaging	0،996
rs41295268	G/A	R338H	ENSP00000446475	Deleterious	0،003	Probably Damaging	0،996
rs41295268	G/A	R166H	ENSP00000438580	Deleterious	0،004	Probably Damaging	0،994
rs41295270	C/T	S580L	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs41295270	C/T	S450L	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs41295270	C/T	S278L	ENSP00000438580	Deleterious	0،002	Probably Damaging	1
rs41295278	A/G	R1321G	ENSP00000234420	Deleterious	0،015	Possibly Damaging	0،542
rs41557217	A/C	E122D	ENSP00000406248	Deleterious (Warning Low Confidence)	0	Possibly Damaging	0،754
rs55760494	G/C	R300P	ENSP00000234420	Deleterious	0،012	Possibly Damaging	0،46
rs56238300	T/G	L1223W	ENSP00000446475	Deleterious	0،004	Probably Damaging	0،999
rs56238300	T/G	L1353W	ENSP00000234420	Deleterious	0،005	Probably Damaging	0،998
rs56238300	T/G	L1051W	ENSP00000438580	Deleterious	0،005	Probably Damaging	0،999
rs61753795	C/T	R988C	ENSP00000234420	Deleterious	0،001	Probably Damaging	0،994
rs61753795	C/T	R686C	ENSP00000438580	Deleterious	0،001	Probably Damaging	0،975
rs61753795	C/T	R858C	ENSP00000446475	Deleterious	0،001	Probably Damaging	0،999
rs61754782	G/C	Q475H	ENSP00000234420	Deleterious	0،005	Possibly Damaging	0،774
rs61754782	G/C	Q345H	ENSP00000446475	Deleterious	0،007	Possibly Damaging	0،774
rs61754782	G/C	Q173H	ENSP00000438580	Deleterious	0،013	Possibly Damaging	0،732
rs61754783	A/G	M362V	ENSP00000446475	Deleterious	0،01	Possibly Damaging	0،885
rs61754783	A/G	M492V	ENSP00000234420	Deleterious	0،011	Possibly Damaging	0،771
rs61754783	A/G	M190V	ENSP00000438580	Deleterious	0،041	Possibly Damaging	0،57
rs63749857	G/A	G670R	ENSP00000234420	Deleterious	0،003	Probably Damaging	0،998
rs63749857	G/A	G368R	ENSP00000438580	Deleterious	0،003	Probably Damaging	0،995
rs63749857	G/A	G540R	ENSP00000446475	Deleterious	0،003	Probably Damaging	0،996
rs63749890	T/G	F265C	ENSP00000234420	Deleterious	0،005	Probably Damaging	1
rs63749890	T/G	F135C	ENSP00000446475	Deleterious	0،007	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs63749898	G/A	R1217K	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63749898	G/A	R915K	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63749898	G/A	R1087K	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63749919	A/G	Y969C	ENSP00000234420	Deleterious	0،001	Probably Damaging	1
rs63749919	A/G	Y667C	ENSP00000438580	Deleterious	0،001	Probably Damaging	0،999
rs63749919	A/G	Y839C	ENSP00000446475	Deleterious	0،001	Probably Damaging	1
rs63749973	G/A	G264R	ENSP00000438580	Deleterious	0،002	Probably Damaging	1
rs63749973	G/A	G566R	ENSP00000234420	Deleterious	0،005	Probably Damaging	1
rs63749973	G/A	G436R	ENSP00000446475	Deleterious	0،005	Probably Damaging	1
rs63750065	A/G	Y267C	ENSP00000446475	Deleterious	0،008	Probably Damaging	0،994
rs63750065	A/G	Y397C	ENSP00000234420	Deleterious	0،009	Probably Damaging	0،984
rs63750065	A/G	Y95C	ENSP00000438580	Deleterious	0،01	Probably Damaging	0،967
rs63750119	G/T	R1242L	ENSP00000234420	Deleterious	0	Probably Damaging	0،999
rs63750119	G/T	R1112L	ENSP00000446475	Deleterious	0	Probably Damaging	0،999
rs63750138	C/T	R772W	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750138	C/T	R470W	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750138	C/T	R642W	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750143	G/T	R29L	ENSP00000390382	Deleterious	0،018	Possibly Damaging	0،531
rs63750157	T/C	C1158R	ENSP00000234420	Deleterious	0،009	Probably Damaging	0،999
rs63750157	T/C	C1028R	ENSP00000446475	Deleterious	0،011	Probably Damaging	0،999
rs63750157	T/C	C856R	ENSP00000438580	Deleterious	0،013	Probably Damaging	0،996
rs63750253	G/A	R1095H	ENSP00000234420	Deleterious	0،047	Probably Damaging	1
rs63750257	G/C	G846R	ENSP00000438580	Deleterious	0،043	Probably Damaging	1
rs63750257	G/C	G1018R	ENSP00000446475	Deleterious	0،043	Probably Damaging	1
rs63750257	G/C	G1148R	ENSP00000234420	Deleterious	0،045	Probably Damaging	1
rs63750287	C/A	A719D	ENSP00000438580	Deleterious	0،021	Probably Damaging	0،987
rs63750358	G/C	G383A	ENSP00000438580	Deleterious	0،016	Probably Damaging	1
rs63750358	G/C	G555A	ENSP00000446475	Deleterious	0،02	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs63750358	G/C	G685A	ENSP00000234420	Deleterious	0،021	Probably Damaging	1
rs63750370	C/T	T1225M	ENSP00000234420	Deleterious	0،001	Probably Damaging	1
rs63750370	C/T	T923M	ENSP00000438580	Deleterious	0،001	Probably Damaging	1
rs63750370	C/T	T1095M	ENSP00000446475	Deleterious	0،001	Probably Damaging	1
rs63750389	A/G	Y850C	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750389	A/G	Y548C	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750389	A/G	Y720C	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750442	C/G	T798R	ENSP00000438580	Deleterious	0،011	Possibly Damaging	0،708
rs63750442	C/G	T1100R	ENSP00000234420	Deleterious	0،012	Possibly Damaging	0،752
rs63750442	C/G	T970R	ENSP00000446475	Deleterious	0،014	Possibly Damaging	0،885
rs63750462	C/T	P623L	ENSP00000234420	Deleterious	0،026	Possibly Damaging	0،866
rs63750462	C/T	P493L	ENSP00000446475	Deleterious	0،045	Probably Damaging	0،97
rs63750595	G/A	C257Y	ENSP00000438580	Deleterious	0،001	Probably Damaging	0،999
rs63750595	G/A	C559Y	ENSP00000234420	Deleterious	0،002	Probably Damaging	1
rs63750595	G/A	C429Y	ENSP00000446475	Deleterious	0،002	Probably Damaging	1
rs63750617	C/T	R946C	ENSP00000446475	Deleterious	0،025	Probably Damaging	0،973
rs63750617	C/T	R1076C	ENSP00000234420	Deleterious	0،028	Probably Damaging	0،991
rs63750725	G/A	R772Q	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750725	G/A	R470Q	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750725	G/A	R642Q	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750741	T/C	L449P	ENSP00000234420	Deleterious	0،002	Probably Damaging	1
rs63750741	T/C	L319P	ENSP00000446475	Deleterious	0،002	Probably Damaging	1
rs63750741	T/C	L147P	ENSP00000438580	Deleterious	0،003	Probably Damaging	1
rs63750753	C/A	P1087H	ENSP00000234420	Deleterious	0،009	Probably Damaging	0،993
rs63750753	C/A	P785H	ENSP00000438580	Deleterious	0،011	Probably Damaging	0،99
rs63750753	C/A	KP957H	ENSP00000446475	Deleterious	0،016	Probably Damaging	0،998
rs63750753	C/G	P1087R	ENSP00000234420	Deleterious	0،014	Possibly Damaging	0،921
rs63750753	C/G	P785R	ENSP00000438580	Deleterious	0،016	Possibly Damaging	0،813

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs63750753	C/G	P957R	ENSP00000446475	Deleterious	0،025	Probably Damaging	0،963
rs63750804	A/T	D729V	ENSP00000438580	Deleterious	0،002	Probably Damaging	0،981
rs63750804	A/T	D1031V	ENSP00000234420	Deleterious	0،003	Probably Damaging	0،999
rs63750804	A/T	D901V	ENSP00000446475	Deleterious	0،003	Probably Damaging	0،997
rs63750836	C/T	T982M	ENSP00000438580	Deleterious	0،002	Probably Damaging	1
rs63750836	C/T	T1154M	ENSP00000446475	Deleterious	0،002	Probably Damaging	1
rs63750836	C/T	T1284M	ENSP00000234420	Deleterious	0،003	Probably Damaging	1
rs63750878	G/T	S285I	ENSP00000234420	Deleterious	0،038	Possibly Damaging	0،698
rs63750882	C/G	H1248D	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750882	C/G	H946D	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750882	C/G	H1118D	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750897	C/G	S373C	ENSP00000446475	Deleterious	0،005	Probably Damaging	0،999
rs63750897	C/G	S503C	ENSP00000234420	Deleterious	0،006	Probably Damaging	0،998
rs63750897	C/G	S201C	ENSP00000438580	Deleterious	0،009	Probably Damaging	0،998
rs63750914	A/C	E1214A	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750914	A/C	E912A	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750914	A/C	E1084A	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750949	C/T	T1219I	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750949	C/T	T917I	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750949	C/T	T1089I	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750969	A/G	K1140R	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63750969	A/G	K838R	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750969	A/G	K1010R	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63750985	C/G	C463W	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63750985	C/G	C765W	ENSP00000234420	Deleterious	0،001	Probably Damaging	1
rs63750985	C/G	C635W	ENSP00000446475	Deleterious	0،001	Probably Damaging	1
rs63751005	T/C	V207A	ENSP00000438580	Deleterious	0،004	Probably Damaging	1
rs63751005	T/C	V509A	ENSP00000234420	Deleterious	0،006	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs63751005	T/C	V379A	ENSP00000446475	Deleterious	0،006	Probably Damaging	1
rs63751063	G/A	G1139S	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63751063	G/A	G837S	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63751063	G/A	G1009S	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63751064	G/A	A1303T	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63751064	G/A	A1001T	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63751064	G/A	A1173T	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63751113	G/A	R976H	ENSP00000234420	Deleterious	0	Probably Damaging	0،999
rs63751113	G/A	R846H	ENSP00000446475	Deleterious	0	Probably Damaging	0،998
rs63751113	G/A	R674H	ENSP00000438580	Deleterious	0،002	Probably Damaging	0،998
rs63751258	G/T	K99N	ENSP00000234420	Deleterious	0،006	Probably Damaging	1
rs63751328	G/A	E1193K	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs63751328	G/A	E891K	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs63751328	G/A	E1063K	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs63751405	T/C	L435P	ENSP00000234420	Deleterious	0،001	Probably Damaging	0،998
rs63751405	T/C	L133P	ENSP00000438580	Deleterious	0،001	Probably Damaging	1
rs63751405	T/C	L305P	ENSP00000446475	Deleterious	0،001	Probably Damaging	0،999
rs115386788	G/T	R858L	ENSP00000446475	Deleterious	0،007	Possibly Damaging	0،812
rs142949377	C/G	P134R	ENSP00000406248	Deleterious (Warning Low Confidence)	0	Probably Damaging	0،99
rs143517321	G/A	E639K	ENSP00000234420	Deleterious	0،03	Possibly Damaging	0،943
rs143517321	G/A	E509K	ENSP00000446475	Deleterious	0،031	Possibly Damaging	0،943
rs143520357	G/T	D641Y	ENSP00000438580	Deleterious	0،011	Probably Damaging	0،964
rs143520357	G/T	D813Y	ENSP00000446475	Deleterious	0،011	Probably Damaging	0،994
rs143520357	G/T	D943Y	ENSP00000234420	Deleterious	0،012	Probably Damaging	0،994
rs146359682	A/C	N854H	ENSP00000446475	Deleterious	0،015	Possibly Damaging	0،658
rs146359682	A/C	N984H	ENSP00000234420	Deleterious	0،02	Possibly Damaging	0،658
rs147453999	A/T	T1243S	ENSP00000234420	Deleterious	0،016	Possibly Damaging	0،946
rs148445930	T/C	M1267T	ENSP00000234420	Deleterious	0	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs148445930	T/C	M965T	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs148445930	T/C	M1137T	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs148517241	G/C	D98H	ENSP00000406248	Deleterious (Warning Low Confidence)	0،048	Probably Damaging	0،984
rs149159527	C/T	T521I	ENSP00000234420	Deleterious	0،004	Probably Damaging	1
rs149159527	C/T	T391I	ENSP00000446475	Deleterious	0،004	Probably Damaging	1
rs149159527	C/T	T219I	ENSP00000438580	Deleterious	0،005	Probably Damaging	1
rs150632241	A/G	E1187G	ENSP00000234420	Deleterious	0،006	Probably Damaging	1
rs150632241	A/G	E1057G	ENSP00000446475	Deleterious	0،008	Probably Damaging	0،999
rs150632241	A/G	E885G	ENSP00000438580	Deleterious	0،01	Probably Damaging	0،998
rs151086192	G/C	D365H	ENSP00000438580	Deleterious	0،003	Probably Damaging	1
rs151086192	G/C	D667H	ENSP00000234420	Deleterious	0،004	Probably Damaging	1
rs151086192	G/C	D537H	ENSP00000446475	Deleterious	0،004	Probably Damaging	1
rs182024561	C/T	L1201F	ENSP00000234420	Deleterious	0	Probably Damaging	0،999
rs182024561	C/T	L899F	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs182024561	C/T	L1071F	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs187491488	G/C	V1253L	ENSP00000234420	Deleterious	0،012	Possibly Damaging	0،756
rs187491488	G/C	V1123L	ENSP00000446475	Deleterious	0،012	Possibly Damaging	0،516
rs190075874	A/C	K866T	ENSP00000234420	Deleterious	0،033	Probably Damaging	0،971
rs192740549	G/C	L1226F	ENSP00000446475	Deleterious	0،022	Possibly Damaging	0،914
rs192740549	G/C	L1356F	ENSP00000234420	Deleterious	0،023	Possibly Damaging	0،788
rs200447622	C/G	S247C	ENSP00000438580	Deleterious	0،025	Possibly Damaging	0،879
rs200944853	A/C	T6P	ENSP00000446475	Deleterious	0	Probably Damaging	0،998
rs201060668	C/G	A1303G	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs201060668	C/G	A1001G	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs201060668	C/G	A1173G	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs201191389	C/G	P976R	ENSP00000438580	Deleterious	0،047	Probably Damaging	0،996
rs201191389	C/T	P976L	ENSP00000438580	Deleterious	0،019	Possibly Damaging	0،819
rs201191389	C/T	P1278L	ENSP00000234420	Deleterious	0،023	Possibly Damaging	0،943

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs201191389	C/T	P1148L	ENSP00000446475	Deleterious	0،025	Probably Damaging	0،993
rs201193496	C/A	H367N	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs201193496	C/A	H65N	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs201193496	C/A	H237N	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs201892477	A/T	I464F	ENSP00000234420	Deleterious	0،009	Probably Damaging	0،993
rs201892477	A/T	I162F	ENSP00000438580	Deleterious	0،011	Probably Damaging	0،973
rs201892477	A/T	I334F	ENSP00000446475	Deleterious	0،013	Probably Damaging	0،997
rs202066386	T/A	V1253E	ENSP00000234420	Deleterious	0،001	Probably Damaging	0،999
rs202066386	T/A	V951E	ENSP00000438580	Deleterious	0،001	Probably Damaging	0،999
rs202066386	T/A	V1123E	ENSP00000446475	Deleterious	0،001	Probably Damaging	0،999
rs202127474	T/C	I493T	ENSP00000438580	Deleterious	0،003	Possibly Damaging	0،55
rs202127474	T/C	I795T	ENSP00000234420	Deleterious	0،004	Possibly Damaging	0،851
rs202127474	T/C	I665T	ENSP00000446475	Deleterious	0،004	Possibly Damaging	0،886
rs202219685	A/T	R411W	ENSP00000234420	Deleterious	0،001	Probably Damaging	0،996
rs202219685	A/T	R281W	ENSP00000446475	Deleterious	0،001	Probably Damaging	0،998
rs202219685	A/T	R109W	ENSP00000438580	Deleterious	0،003	Probably Damaging	0،995
rs267608038	G/T	V96F	ENSP00000390382	Deleterious	0،013	Probably Damaging	0،99
rs267608038	G/T	V195F	ENSP00000234420	Deleterious	0،027	Probably Damaging	0،986
rs267608052	G/C	A457P	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs267608052	G/C	A155P	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs267608052	G/C	A327P	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs267608054	G/T	D1026Y	ENSP00000234420	Deleterious	0،004	Probably Damaging	0،999
rs267608054	G/T	D724Y	ENSP00000438580	Deleterious	0،005	Probably Damaging	0،999
rs267608054	G/T	D896Y	ENSP00000446475	Deleterious	0،005	Probably Damaging	0،998
rs267608067	T/A	W475R	ENSP00000438580	Deleterious	0،014	Probably Damaging	1
rs267608067	T/A	W647R	ENSP00000446475	Deleterious	0،02	Probably Damaging	1
rs267608067	T/A	W777R	ENSP00000234420	Deleterious	0،021	Probably Damaging	1
rs267608089	C/T	T1142M	ENSP00000234420	Deleterious	0	Probably Damaging	1

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs267608089	C/T	T840M	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs267608089	C/T	T1012M	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs267608100	C/G	D879E	ENSP00000438580	Deleterious	0،012	Probably Damaging	1
rs267608100	C/G	D1051E	ENSP00000446475	Deleterious	0،017	Probably Damaging	1
rs267608100	C/G	D1181E	ENSP00000234420	Deleterious	0،018	Probably Damaging	1
rs267608140	T/A	L1052Q	ENSP00000438580	Deleterious	0،028	Possibly Damaging	0،489
rs267608140	T/A	L1354Q	ENSP00000234420	Deleterious	0،04	Possibly Damaging	0،836
rs267608140	T/A	L1224Q	ENSP00000446475	Deleterious	0،045	Possibly Damaging	0،836
rs368437140	G/A	D857N	ENSP00000234420	Deleterious	0،02	Probably Damaging	1
rs368437140	G/A	D727N	ENSP00000446475	Deleterious	0،021	Probably Damaging	0،999
rs368437140	G/A	D555N	ENSP00000438580	Deleterious	0،023	Probably Damaging	0،999
rs369042519	C/T	A762V	ENSP00000438580	Deleterious	0،014	Possibly Damaging	0،949
rs369042519	C/T	A934V	ENSP00000446475	Deleterious	0،028	Possibly Damaging	0،929
rs369042519	C/T	A1064V	ENSP00000234420	Deleterious	0،031	Possibly Damaging	0،836
rs369456858	C/T	R468C	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs369456858	C/T	R166C	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs369456858	C/T	R338C	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs370157832	G/A	R144H	ENSP00000406248	Deleterious (Warning Low Confidence)	0،047	Probably Damaging	0،978
rs370157832	G/A	R113H	ENSP00000446475	Deleterious	0،048	Possibly Damaging	0،951
rs370412074	C/T	S860F	ENSP00000234420	Deleterious	0،013	Probably Damaging	0،992
rs370412074	C/T	S730F	ENSP00000446475	Deleterious	0،013	Probably Damaging	0،969
rs370412074	C/T	S558F	ENSP00000438580	Deleterious	0،016	Possibly Damaging	0،897
rs370505117	C/T	R1024W	ENSP00000234420	Deleterious	0،003	Probably Damaging	1
rs370505117	C/T	R722W	ENSP00000438580	Deleterious	0،003	Probably Damaging	0،998
rs370505117	C/T	R894W	ENSP00000446475	Deleterious	0،004	Probably Damaging	1
rs372103816	A/G	Y1066C	ENSP00000234420	Deleterious	0،002	Probably Damaging	0،999
rs372103816	A/G	Y936C	ENSP00000446475	Deleterious	0،002	Probably Damaging	0،999
rs372103816	A/G	Y764C	ENSP00000438580	Deleterious	0،003	Probably Damaging	0،999

SNP ID	Nucleotide Change	Amino Acid Change	Protein ID	SIFT Prediction	TI	Polyphen-2 Result	PSIC SD
rs372352774	A/G	K26E	ENSP00000475605	Deleterious	0،02	Possibly Damaging	0،706
rs372990379	C/G	S806C	ENSP00000234420	Deleterious	0،025	Possibly Damaging	0،901
rs373622047	T/C	Y994H	ENSP00000234420	Deleterious	0	Probably Damaging	1
rs373622047	T/C	Y692H	ENSP00000438580	Deleterious	0	Probably Damaging	1
rs373622047	T/C	Y864H	ENSP00000446475	Deleterious	0	Probably Damaging	1
rs373721483	T/A	C481S	ENSP00000438580	Deleterious	0،003	Probably Damaging	0،992
rs373721483	T/A	C783S	ENSP00000234420	Deleterious	0،007	Probably Damaging	0،998
rs373721483	T/A	C653S	ENSP00000446475	Deleterious	0،007	Probably Damaging	0،996
rs374041375	A/G	N124D	ENSP00000406248	Deleterious (Warning Low Confidence)	0	Probably Damaging	0،992
rs374070511	C/T	P808S	ENSP00000438580	Deleterious	0،049	Probably Damaging	0،997
rs375966384	C/T	P991L	ENSP00000234420	Deleterious	0،034	Probably Damaging	0،999
rs375974046	C/T	T239I	ENSP00000446475	Deleterious	0،018	Possibly Damaging	0،935
rs375974046	C/T	T67I	ENSP00000438580	Deleterious	0،019	Possibly Damaging	0،897
rs375974046	C/T	T369I	ENSP00000234420	Deleterious	0،021	Possibly Damaging	0،835
rs376220212	G/A	R577H	ENSP00000234420	Deleterious	0،014	Probably Damaging	0،998
rs376220212	G/A	R447H	ENSP00000446475	Deleterious	0،023	Probably Damaging	0،998
rs376220212	G/A	R275H	ENSP00000438580	Deleterious	0،043	Probably Damaging	0،998
rs376243329	C/T	R1095C	ENSP00000234420	Deleterious	0،005	Probably Damaging	1
rs376243329	C/T	R965C	ENSP00000446475	Deleterious	0،009	Probably Damaging	1
rs376243329	C/T	R793C	ENSP00000438580	Deleterious	0،023	Probably Damaging	1
rs377216828	C/T	R113C	ENSP00000446475	Deleterious	0،021	Possibly Damaging	0،951
rs377216828	C/T	R144C	ENSP00000406248	Deleterious (Warning Low Confidence)	0،023	Probably Damaging	0،994
rs377216828	C/T	R243C	ENSP00000234420	Deleterious	0،037	Possibly Damaging	0،951
rs377542011	A/C	Q889P	ENSP00000234420	Deleterious	0،038	Probably Damaging	0،971
rs377542011	A/C	Q759P	ENSP00000446475	Deleterious	0،044	Possibly Damaging	0،905

PolyPhen-2 result: POROBABLY DAMAGING (more confident prediction) / POSSIBLY DAMAGING (less confident prediction), **PSIC SD:** Position-Specific Independent Counts software if the score is ≥ 0.5, **Tolerance Index**: Ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is ≤ 0.05, and tolerated if the score is > 0.05. **RI**: Reliability Index

REFERENCES

[1] http://www.britannica.com/science/single-nucleotide-polymorphism

[2] Carninci P, Kasukawa T, Katayama S et al. 2005. The transcriptional and scape of the mammalian genome. Science. 309:1559–1563.

[3] Liu J, Gough J, and Rost B. 2006. Distinguishing protein-coding from non-coding RNAs through support vector machines. PLOS Genetics. 2:4-29.

[4] Ng. P. C. and Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. Annual Review of Genomics and Human Genetics.7:61–80.

[5] Smith E.P, Boyd J, Frank G.R et al. 1994. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. The New England Journal of medicine. 331:1056– 1061.

[6] Lander E. S. 1996. The new genomics: global views of biology. Science.274:536–539.

[7] Thomas R, R McConnell, J Whittacker, P Kirkpatrick, Bradley J, and Sandford R. 1999. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. The American Journal of Human Genetics. 65:39–49.

[8] Doniger S.W, Kim H. S, Swaine D, et al. 2008. Acatalog of neutral and deleterious polymorphism in yeast. PLOS Genetics. 4: 8.

[9] Billur H. E, Hofree M and Carter H. 2015. Identifying mutation specific cancer Pathways using a structurally resolved protein interaction network. Pacific Symposium on Biocomputing. 20: 84-95.

[10] Jeffrey R. B and Yang Z. 2015. Predicting the Effect of Mutations on Protein Protein Binding Interactions through Structure-Based Interface Profiles. PLOS Computational Biology. DOI: 10.1371. 1-2

[11] http://www.ncbi.nlm.nih.gov/gene. Accessed in January 2016.

[12] http://ghr.nlm.nih.gov/gene/MSH6. Accessed in February 2016.

[13] Drummond J.T, Li G.M, Longley M.J and Modrich P. 1995. Isolation of an hMSH2–p160 heterodimer that restores DNA mismatch repair to tumor cells. Science. 268:1909–1912.

[14] Palombo F, Gallinari P, Iaccarino I, Lettieri T, Hughes M, D'Arrigo A, Truong O, Hsuan J.J and Jiricny J. 1995. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science. 268:1912–1914.

[15] Palombo F, Iaccarino I, Nakajima E, Ikejima M, Shimada T and Jiricny J. 1996. hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr. Biol. 6:1181–1184.

[16] Genschel J, Littman S.J, Drummond J.T and Modrich P. 1998. Isolation of MutSβ from human cells and comparison of the mismatch repair specificities of MutSβ and MutSα. J. Biol. Chem. 273:19895–19901.

[17] Marsischky G.T, Filosi N, Kane M.F and Kolodner R. 1996. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev.10: 407–420.

[18] Edelmann W, Yang K., Umar A, Heyer J, Lau K., Fan K, Liedtke W, Cohen P.E, Kane M.F, Lipford J.R. et al. 1997 Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell, 91, 467–477.

[19] Edelmann W, Umar A, Yang K, Heyer J, Kucherlapati M, Lia M, Kneitz B, Avdievich E, Fan K, Wong E et al. 2000. The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res. 60:803–807.

[20] Sia E.A, Kokoska R.J, Dominska M, Greenwell P and Petes T.D. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol.17:2851–2858.

[21] Lynch H.T, de la Chapelle A. 2003. Hereditary colorectal cancer. N Engl J Med. 348(10):919-932.

[22] Watson P, Vasen H.F, Mecklin J.P, et al. 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. Int J Cancer. 123(2): 444-449.

[23] Vasen H.F, Watson P, Mecklin J.P and Lynch H.T. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology. 116(6): 1453-1456.

[24] Rodriguez-Bigas M.A, Boland C.R, Hamilton S.R, et al. 1997. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst. 23(89): 1758-1762.

[25] Terdiman J.P, Gum J.R Jr, Conrad P.G, et al. 2001. Efficient detection of hereditary nonpolyposis colorectal cancer gene carriers by screening for tumor microsatellite instability before germline genetic testing. Gastroenterology. 120(1): 21-30.

[26] Graham R.P, Kerr S.E, Butz M.L, Thibodeau S.N, Halling K.C, Smyrk T.C, Dina M.A, Waugh V.M and Rumilla K.M. 2015. Heterogenous MSH6 loss is a result of microsatellite instability within MSH6 and occurs in sporadic and hereditary colorectal and endometrial carcinomas. Am J Surg Pathol. 39(10):1370-6.

[27] Connor A.A, Katzov-Eckert H, Whelan T, Aronson M, Lau L, Marshall C, Charames G.S, Pollett A, Gallinger S and Lerner-Ellis J. 2015. Identification of a novel MSH6 germline variant in a family with multiple gastro-intestinal malignancies by next generation sequencing. Fam Cancer. 14(1): 69-75.

[28] Castellsagué E, Liu J, Volenik A, Giroux S, Gagné R, Maranda B, Roussel-Jobin A, Latreille J, Laframboise R, Palma L, Kasprzak L, Marcus VA, Breguet M, Nolet S, El-Haffaf Z, Australie K, Gologan A, Aleynikova O, Oros-Klein K, Greenwood C, Mes-Masson AM, Provencher D, Tischkowitz M, Chong G, Rousseau F and Foulkes W.D. 2015. Characterization of a novel founder MSH6 mutation causing Lynch syndrome in the French Canadian population. Clin Genet. 87(6):536-42.

[29] Kato M, Takano M, Miyamoto M, Sasaki N, Goto T, Tsuda H and Furuya K. 2015. DNA mismatch repair-related protein

loss as a prognostic factor in endometrial cancers. J Gynecol Oncol. 26(1):40-5.

[30] Pritchard C.C, Morrissey C, Kumar A, Zhang X, Smith C, Coleman I, Salipante S.J, Milbank J, Yu M, Grady W.M, Tait J.F, Corey E, Vessella RL, Walsh T7, Shendure J and Nelson P.S. 2014. Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. Nat Commun. 25(5):4988.

[31] Hoell J.I, Gombert M, Ginzel S, Loth S, Landgraf P, Käfer V, Streiter M, Prokop A, Weiss M, Thiele R and Borkhardt A. 2014. Constitutional mismatch repair-deficiency and whole-exome sequencing as the means of the rapid detection of the causative MSH6 defect. Klin Padiatr. 226 (6-7):357-61.

[32] Nguyen S.A, Stechishin O.D, Luchman H.A, Lun X.Q, Senger D.L, Robbins S.M, Cairncross J.G and Weiss S. 2014. Novel MSH6 mutations in treatment-naïve glioblastoma and anaplastic oligodendroglioma contribute to temozolomide resistance independently of MGMT promoter methylation. Clin Cancer Res. 20(18): 4894-903.

[33] Stark A.M, Doukas A, Hugo H.H, Hedderich J, Hattermann K, Maximilian Mehdorn H and Held-Feindt J. 2015. Expression of DNA mismatch repair proteins MLH1, MSH2, and MSH6 in recurrent glioblastoma. Neurol Research. 37(2): 95-105.

[34] Alvino E, Passarelli F, Cannavò E, Fortes C, Mastroeni S, Caporali S, Jiricny J, Cappellini GC, Scoppola A, Marchetti P, Modesti A and D'Atri S. 2014. High expression of the mismatch repair protein MSH6 is associated with poor patient survival in melanoma. Am J Clin Pathol. 142(1):121-32.

[35] Peltomäki P. 2001. Deficient DNA Mismatch Repair: A Common Etiologic Factor For Colon Cancer. Hum. Mol. Genet. 10:735–740.

[36] Wijnen J, De Leeuw W, Vasen H, Van Der Klift H, Møller P and Stormorken A. 1999. Familial Endometrial Cancer In Female Carriers Of Msh6 Germline Mutations. Nat. Genet. 23:142–144.

[37] Wagner A, Hendriks Y, Meijers-Heijboer E.J, de Leeuw W.J.F, Morreau H, Hofstra R, Tops C, Bik E, Bröcker-Vriends A.H.J.T, van der Meer C. et al. 2001. Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. J. Med. Genet.38: 318–322.

[38] Berends M.J, Wu Y, Sijmons R.H, Mensink R.G, Van der Sluis T, Hordijk-Hos J.M, De Vries E.G, Hollema H, Karrenbeld A, Buys C.H, Van der Zee A.G, Hofstra R.M and Kleibeuker J.H. 2002. Molecular and clinical characteristics of msh6 variants: an analysis of 25 index carriers of a germline variant. Am j hum genet. 70:26–37.

[39] Schweizer P, Moisio A-L, Kuismanen S.A, Truninger K, Vierumaki R, Salovaara R, Arola J, Butzow R, Jiricny J, Peltomaki P and Nystrom-Lahti M. 2001. Lack of MSH2 and MSH6 characterizes endometrial but not colon carcinomas in hereditary nonpolyposis colorectal cancer. Cancer Res. 61:2813–2815.

[40] http://informatics.sdsu.edu/bioinformatics/. Accessed in march 2015

[41] http://www.biotecharticles.com/Bioinformatics-Article/Importance-and-Applications-of-Bioinformatics-in-Molecular-Medicine-214.html

[42] Mohamed Dafaalla, Tarig Mabrouk, Omer Ali, Ibrahim Elkhidir, Mohamed Adel Taha, Mohamed A. Abdelrahim, Osman Arbab, Musaab M. alfaki, Abbashar M. Hussein and Mohamed A. Hassan. 2015. Computational Analysis of the Single nucleotide Polymorphisms that Affect Superoxide Dismutase Reveals Important Domain Related to Protein. American Journal of Biomedical Research. 3(4): 58-65.

[43] Shahenaz S.Salih MAH, Ibtihal M. Abdelhag, Wafaa M. Abdalla, Altaf S. Mosad and Mohamed M. Hassan. 2014. Computational Detection of Deleterious Single Nucleotide Polymorphisms in Human Adenomatous Polyposis Coli Gene the Gate-Keeper of Colorectal Carcinoma. International Journal of Computational Bioinformatics and In Silico Modeling. 3(6):531-537.

[44] http://pages.genemania.org/. Accessed in December 2015

[45] SIFT HELP [Internet]. Available from: http://sift.bii.a-star.edu.sg/www/SIFT_help.html#SIFT.

[46] Ivan Adzhubei, Daniel M. Jordan, Sunyaev S.R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current Protocols in Human Genetics. 7(20): 1-41.

[47] Bava KA, Gromiha MM, Uedaira H, Kitajima K and Sarai A. 2004. ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res. 32: D120-D121.

[48] J. Cheng A. Randall, P. Baldi. 2005. Prediction Of Protein Stability Changes For Single-Site Mutations Using Support Vector Machines. Proteins. 62(4): 1125-1132.

[49] J. Cheng, A. Z. Randall, M. Sweredoski, and P. Baldi. 2005. Scratch: A Protein Structure and Structural Feature Prediction Server. Nucleic Acids Res. 33 (Server Issue): W72-76.

[50] Capriotti E, Calabrese R and Casadio R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics. 22; 2729-2734.

[51] Capriotti E, Fariselli P, Calabrese R. and Casadio R. 2005. Predicting protein stability changes from sequences using support vector machines. Bioinformatics. 21 (Suppl 2): ii54-ii58.

[52] Capriotti E, Fariselli P and Casadio R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research. 33 (Web server issue): W306-W310.

[53] Altschul S. F, Madden T. L, Schaffer A. A, Zhang J, Zhang. Z, Miller. W and Lipman D. J. Gapped. 1997. BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25: 3389-3402.

[54] Calabrese R, Capriotti E, Fariselli P, Martelli P. L and Casadio R. 2009. Functional Annotations Improve the Predictive Score of Human Disease-related Mutations in Proteins. Human Mutation. 30: 1237-1244.

[55] Li B, Krishnan V.G, Mort M.E, Xin F, Kamati K.K, Cooper D.N, Mooney S.D and Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 25: 2744–2750

[56] Mutation3D: Cancer Gene Prediction. 2016. Through Atomic Clustering of Coding Variants in the Structural Proteome. Hum Mutat. doi: 10.1002/humu.22963.

[57] Nielsen M, Lundegaard C, Lund O and Petersen TN. 2010. CPHmodels- 3.0 - Remote homology modeling using structure guided sequence profiles. Nucleic Acids Res. 38, doi:10.1093/nar/gkq535.

[58] http://www.cgl.ucsf.edu/chimera/. Accessed in June 2016

[59] Jesse D, Ziebarth YC, Anindya Bhattacharya and Anlong Chen. 2012. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. Nucleic Acids Research. 40 (Database issue):D216-D221.

[60] Papadopoulos N, Nicolaides NC, Liu B, Parsons R, Lengauer C, Palombo F, D'Arrigo A, Markowitz S, Willson JK, Kinzler KW, et al. 1995. Mutations of GTBP in genetically unstable cells. Science. 268:1915–1917.

[61] Alani E. 1996. The Saccharomyces cerevisiae Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. Mol. Cell. Biol. 16:5604–5615.

[62] Culligan K, Meyer-Gauen G, Lyons-Weiler J, Hays J. 2000. Evolutionary Origin, Diversification and Specialization of Eukaryotic MutS Homolog Mismatch Repair Proteins. Nucleic Acids Res. 28:463–471.

[63] Modrich P. 2006. Mechanisms in eukaryotic mismatch repair. The Journal of Biological Chemistry. 281:30305–30309.

[64] Matton N, Simonetti J, Williams KJ. 2000. Identification of mismatch repair protein complexes in HeLa nuclear extracts and their interaction with heteroduplex DNA. Journal of Biological Chemistry. 275:17808–17813.

[65] Talavera D, Robertson DL, Lovell SC. 2011. Characterization of protein-protein interaction interfaces from a single species. PloSone. 6: e21053. doi:10.1371 /journal.pone 0021053 PMID:21738603.

[66] Andreani J, Faure G, Guerois R. Versatility and in variance in the evolution of homologous heteromeric interfaces. 2012. PLoS computational biology. 8: e1002677. doi:10.1371/journal.pcbi.1002677 PMID: 22952442.

[67] Guharoy M, Chakrabarti P. 2005. Conservation and relative importance of residues a cros protein-protein interfaces. Proceedings of the National Academy of Sciences of the United States of America. 102: 15447–52.PMID:16221766.

[68] Ying Wu, Maran J.W. Berends, Rob G.J. Mensink , Claudia Kempinga, Rolf H. Sijmons, Ate G.J. van der Zee, Harry Hollema, Jan H. Kleibeuker, Charles H.C.M. Buys and Robert and M.W. Hofstra. 1999. Association of Hereditary Nonpolyposis Colorectal Cancer–Related Tumors Displaying Low Microsatellite Instability with MSH6 Germline Mutations. Am.J.Hum.Genet. 56(5): 1291-1298.

[69] Landi D, Barale R, Gemignani F and Landi S. 2011. Prediction of the biological effect of polymorphisms within microrna binding sites. Methods.Mol.Biol.676: 197-210.