Title

High-throughput sequencing and in-silico analysis confirm pathogenicity of novel MSH3 variants in African American colorectal cancer

Authors

Rashid, Mudasir

Rashid, Rumaisa

Gadewal, Nikhil

et al.

Publication Date

2024-03-01

DOI

10.1016/j.neo.2024.100970

Peer reviewed

Neoplasia 49 (2024) 100970

Contents lists available at ScienceDirect

Neoplasia

journal homepage: www.elsevier.com/locate/neo

Original Research

High-throughput sequencing and *in-silico* analysis confirm pathogenicity of novel *MSH3* variants in African American colorectal cancer

Mudasir Rashid [a], Rumaisa Rashid [a], Nikhil Gadewal [b], John M. Carethers [c,d], Minoru Koi [c], Hassan Brim [a], Hassan Ashktorab [a,*]

[a] Department of Medicine, Gastroenterology Division, Department of Pathology and Cancer Center, Howard University College of Medicine, Washington, DC 20059, USA
[b] Bioinformatics and Computational Biology Facility, Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Kharghar, Navi Mumbai, MH 410210, India
[c] Division of Gastroenterology and Hepatology, Department of Medicine, UC San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA
[d] Moores Cancer Center, and Herbert Wertheim School of Public Health and Human Longevity Science, UC San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA

ARTICLE INFO

Keywords:
MSH3
Non-synonymous mutation
African American
Colorectal carcinoma
Cancer disparities
Mismatch repair

ABSTRACT

The maintenance of DNA sequence integrity is critical to avoid accumulation of cancer-causing mutations. Inactivation of DNA Mismatch Repair (MMR) genes (e.g., *MLH1* and *MSH2*) is common among many cancers, including colorectal cancer (CRC) and is the driver of classic microsatellite instability (MSI) in tumors. Somatic *MSH3* alterations have been linked to a specific form of MSI called elevated microsatellite alterations at selected tetranucleotide repeats (EMAST) that is associated with patient poor prognosis and elevated among African American (AA) rectal cancer patients. Genetic variants of *MSH3* and their pathogenicity vary among different populations, such as among AA, which are not well-represented in publicly available databases. Targeted exome sequencing of *MSH3* among AA CRC samples followed by computational bioinformatic pipeline and molecular dynamic simulation analysis approach confirmed six identified *MSH3* variants (c.G1237A, c.C2759T, c.G1397A, c.G2926A, c.C3028T, c.G3241A) that corresponded to MSH3 amino-acid changes (*p*.E413K; *p*.S466N; *p*.S920F; *p*.E976K; *p*.H1010Y; *p*.E1081K). All identified *MSH3* variants were non-synonymous, novel, pathogenic, and show loss or gain of hydrogen bonding, ionic bonding, hydrophobic bonding, and disulfide bonding and have a deleterious effect on the structure of MSH3 protein. Some variants were located within the ATPase site of MSH3, affecting ATP hydrolysis that is critical for MSH3′s function. Other variants were in the MSH3-MSH2 interacting domain, important for MSH3's binding to MSH2. Overall, our data suggest that these variants among AA CRC patients affect the function of MSH3 making them pathogenic and likely contributing to the development or advancement of CRC among AA. Further clarifying functional studies will be necessary to fully understand the impact of these variants on MSH3 function and CRC development in AA patients.

Introduction

The use of high-throughput sequencing and computational techniques are being used to identify tumor mutational signatures (TMSs) associated with tumor suppression and oncogenes. These TMSs have been linked to the development and progression of various types of cancer, including colorectal cancer (CRC) [1–6]. The mutational signatures have been recorded in several databases, such as, The Cancer Genome Atlas (TCGA) and Catalogue of Somatic Mutations in Cancer (COSMIC), which are valuable resources for cancer researchers [7,8]. Several studies have shown that mutations and dysfunction of MMR genes are considered as an early driver event linked with CRC pathogenesis [9–11]. The human DNA mismatch repair (MMR) system is made of several proteins that work together as heterodimers, such as MutSα (a heterodimer of MSH2-MSH6), MutSβ (a heterodimer of MSH2-MSH3) and MutLα (a heterodimer of MLH1-PMS2) that direct repair of single nucleotide mispairs and slippage mistakes generated during DNA replication. The MutSα complex has a high recognition fidelity for single base-mismatches, and MutSβ for small and large insertion-deletion (I/D) slippages. The ratio of MutSα/MutSβ maintains genomic stability [12,13]. Deregulation of MutSα/MutSβ ratio results in microsatellite instability (MSI) and/or Elevated Microsatellite

* Corresponding author.
E-mail address: hashktorab@howard.edu (H. Ashktorab).

https://doi.org/10.1016/j.neo.2024.100970
Received 5 November 2023; Received in revised form 19 December 2023; Accepted 8 January 2024

Alterations at Selected Tetranucleotide repeats (EMAST) genotypes [14]. Several lines of evidence have well-documented the role of MMR genes (*MSH2, MSH6, MLH1, PMS2*) and to some extent *MSH3* in CRC pathogenesis [11,15–17].

Over-expression of MSH3 imbalances the comparative levels of MutSβ and MutSα. High levels of MutSβ sequester more MSH2 and gradually depletes MutSα which leads to tumor with instable mono- and dinucleotide microsatellites, also referred to as microsatellite instability high tumors (MSI-H) with increased rates of somatic point mutations. MSH3 deficiency has been identified in inflammatory hamartomatous polyps in the gastrointestinal (GI) tract [18] as well as in non-dysplastic colonic tissue from ulcerative colitis patients [19]. Loss of MSH3 function is strongly linked with the biomarker EMAST, a type of MSI at dinucleotide (CA) and tetranucleotide repeats microsatellites (AAAG or ATAG) and observed in 50 % of sporadic CRC [20], much more frequent than the 10-15 % of MSI-high tumors driven by MLH1/MSH2 alterations. Somatic MSH3 dysunction can occur by mis-localization/shuttling (from nucleus to cytoplasm) with inflammation resulting in EMAST and associated with patient poor prognosis. EMAST has been shown to be more prevalent (50 %) among African American (AA) CRC patients than Caucasian CRC patients [21–23]. Other reports have shown that biallelic germline MSH3 mutations cause a colorectal oligopolyposis syndrome for which normal tissues demonstrate EMAST [24,25]. Additionally, approximately 30 % of sporadic MSI-high CRCs contain secondary somatic frameshift mutations within *MSH3* at an intrinsic coding polyadenine microsatellite sequence [24]. Recent reports have shown that MSH3 shuttling, impairing its nuclear function, most frequently occurs with inflammation, oxidative and/or hypoxia conditions resulting in EMAST phenotype [23,26,27]. An identified MSH3 nuclear localization signal (NLS) and nuclear export signal (NES) sequences rich in lysine residues are partly associated with MSH3 shuttling [23]. Acetylation/deacetylation on N-terminal residues (K98, K99, K103, K122, and K123) that are proximal or within NLS of MSH3 regulate shuttling from nucleus to cytoplasm [28,29]. HDAC3 deacetylates these lysine residues with unknown function [28], suggesting that intercellular localization of MSH3 is partially dependent on HDAC3 activity on the N- terminal residues of MSH3. It is unclear why acetylated MSH3 protein translocate from the nucleus and what is the fate of cytoplasmic acetylated MSH3 whether it undergoes degradation or stabilizes and performs any kind of function in the cytoplasm has not yet been explored. Recent reports have shown that the E3 ubiquitin ligase synoviolin 1 (SYVN1) interacts directly with MSH3 and degrades it in the cytoplasm in age-related cataract (ARC)) [30].

In addition to cytoplasmic shuttling of MSH3, AA patients with CRC show reduced survival compared Caucasian (CA) patients with CRC. Due to the correlation between MSH3 dysfunction and decreased survival rates in AA patients, as well as the documented evidence of increased MSH3 dysfunction among African American CRC patients [23,31,32], we investigated alternative mechanisms of MSH3 dysfunction in this specific subgroup, recognizing that not all AA-CRC patients exhibit such dysfunction. Here, we used an *in-silico* approach and identified six *MSH3* variants among AA CRC patients, all of which are non-synonymous, novel, and pathogenic with deleterious effects on the predicted function and structure of MSH3 protein involving loss or gain of hydrogen bonding, ionic bonding, hydrophobic bonding, and disulfide bonding. Some variants were located within the ATPase site of the MSH3 altering hydrolysis of ATP that is critical for the protein's function.

Material and methods

We previously performed targeted exome sequencing of 140 AA colon specimens that included a subset of 54 AA-CRC in which we reported hundreds of novel and deleterious variants in different genes [31]. Of these, 6 novel non-synonymous *MSH3* variants were evaluated for further analysis to predict their functional impact such as the novelty of these variants was assessed by searching cosmic, gnome AD exome allele frequency dbSNP, and their pathogenicity were predicted using SIFT, PolyPhen-2, PROVEAN, and MutPred2 and other bioinformatics web-based databases/tools of which the details are mentioned in Supplementary Table 2.

Prediction of novelty of non-synonymous MSH3 mutations

Cosmic v96 (https://cancer.sanger.ac.uk/cosmic, the Catalogue of Somatic Mutations In Cancer) is a vast in-depth resource for studying the effects of somatic mutations in human cancer. It is predicted based on Cancer Mutation Census Genome Version GRCh37 (CMC). The CMC integrates data from manual curation and computational analyses from all coding somatic mutations in COSMIC with biological and biochemical information from various sources like ClinVar significance, dN/dS ratios, and variation frequencies in healthy populations (gnom AD).

Prediction of deleterious effect of non-synonymous mutations of MSH3

Condel (http://bbglab.irbbarcelona.org/fannsdb/home) It is an online tool used to evaluate the effect of non-synonymous single nucleotide variation (SNVs), it combines several tools (Polyphen2, SIFT, FATHMM and Mutation Assessor) and provides a cumulative consensus deleterious score. The prediction result score ranging from zero means neutral and one means deleterious.

M-CAP (http://bejerano.stanford.edu/MCAP) is a pathogenic online predictor for rare non-synonymous mutations in the human genome using integrated analysis from various previous pathogenicity algorithms like Polyphen-2, SIFT, and CADD. The thresholds (> 0.025) are considered as pathogenic variants.

REVEL (https://sites.google.com/site/revelgenomics/) is a new ensemble approach that combines the results from 13 different tools (MutPred, VEST 3.0, FATHMM v2.3, Polyphen-2, MutationTaster, MutationAssessor, LRT, SIFT, PROVEAN, GERP++, phyloP, SiPhy, and phastCons.) to estimate the pathogenicity of missense variants. The threshold value for pathogenicity is >0.5.

VarCards (http://159.226.67.237/sun/varcards/welcome) is a combined genetic and clinical 23 *in silico* predictive algorithms. It provides functional consequences and allele frequencies in different populations using various databases such as SIFT, Polyphen-2_HDIV, Polyphen-2_HVAR, MutationTaster, LRT MutationAssessor, FATHMM, FATHMM_MKL, PROVEAN, MetaSVM, VEST3, MetaLR, CADD, M-CAP, MetaLR, DANN, Eigen, GenoCanyon, fitCons, phyloP, GERP++ phastCons, SiPhy, REVEL and ClinPred.

Structural stability and gain and loss of chemical bonding

The structural stability of the protein may be altered because of mutations. We used web-based tools to examine the energy stability of harmful nonsynonymous nucleotide changes.

HOPE (https://www3.cmbi.umcn.nl/hope/method/) HOPE gathers data from BLAST, 3D structures computations, sequence annotation from Uniports, and homology modeling to study the structural and functional effects of a point mutation.

MUpro (http://mupro.proteomics.ics.uci.edu/) A web-based tool which predicts and determines stability alterations based on sequence-based single-site mutation. Based on the criteria score between 1 and -1, a score more than 0 (>0) predicts increases the protein stability. Conversely, a score less than 0 (<0) predicts mutation decreases the protein stability.

Putative molecular mechanisms of the six amino-acid substitutions

MutPred2 (http://mutpred.mutdb.org/index.html) a web-based machine learning approach that combines genomic and molecular data to assess the pathogenicity of mutated amino acids in proteins. It includes 53,180 pathogenic and 206,946 putatively neutral (unlabeled)

variants obtained from the HGMD, Swiss Var, dbSNP DB., and other databases were used mentioned in detail in Table 3 and Supplementary Table 4. it predicts the functional consequences of amino acid substitutions in proteins based on several factors, including the location of the substitution within the protein structure, the physicochemical properties of the substituted amino acid, and the evolutionary conservation of the affected residue.

Loop modeling and molecular dynamics simulation

The missing residues in the co-crystalized structure of MSH2 and MSH3 (PDB ID: 3THX) were modeled using pdb-fixer tool. The wild-type and mutant complexes (*p.*E413K, *p.*S466N, *p.*S920F, *p.*E976K, *p.*H1010Y and *p.*E1081K) were modelled (Fig. 3A) and subjected to molecular dynamics simulation using GROMACS v2020 [32] with implementation of OPLS-AA/L force field [33]. The systems were solved using the TIP3P water model in a cubic box with periodic boundary conditions and counterions were added to neutralize the system. The systems were first energy minimized using steepest descent algorithm with a tolerance of 1000kJ/mol/nm. Electro-static interactions were calculated using Particle Mesh Ewald (PME) summation with 1nm cut offs for columbic interactions and van der Waal interactions were calculated with a distance cut-off of 1.4nm. Later, systems were equilibrated by applying positional restraints on the structure using NVT followed by NPT ensemble for 1000ps each. Temperature of 300K was coupled using Berendsen thermostat while pressure at 1bar, coupled by the Parrinello–Rahman algorithm. The equilibrated systems were then subjected to 300ns of production run with time-step integration of 2fs. The trajectories were saved at every 2 ps (picosecond) and analyzed using analysis tools from GROMACS. The Root Mean Square deviation (RMSD), Root Mean Square Fluctuations (RMSF) and intermolecular hydrogen bonds between MSH2-MSH3 were calculated for all the complexes. Additionally, we analyzed the intermolecular interaction between the ATPase binding domain of MSH3 and MSH2, an index file (.ndx file) was created representing 700 to 850 amino acids of MSH2 and 850 to 1100 amino acids of MSH3 respectively. Gromacs commands mindist and H-bond were used to calculate the minimum distance between any pair of atoms and number of hydrogen bonds in MSH2-MSH3 complex. The Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method was used to calculate the binding free energies between MSH2-MSH3 complexes. The trajectories from 100ns to 300ns of all the complexes were extracted for MMPBSA analysis using gmx_MMPBSA tool [34].

Principal component analysis (PCA), normal mode analysis (NMA), dynamics cross-correlation matrix (DCCM) and residue interaction network (RIN)

The principal component analysis (PCA) of the Wild Type and the six mutants were carried out using 'bio3d' R-package [35]. The trajectories from Gromacs in xtc format were converted to dcd format using mdconvert tool from MDTraj package [36]. The principal component was plotted for C-Alpha atoms of MSH2 and MSH3 complex for wild-type and mutants. The normal mode analysis (NMA) and dynamics cross-correlation matrix (DCCM) was performed using DynaMut tool [37]. The input structure was extracted from the Gromacs trajectory for each complex having minimum potential energy. The force field selected for the analysis was ANM (Anisotropic Network Model) to obtain superimposed non-trivial modes of MSH2 and MSH3 complexes (Fig. 3) and dynamics cross-correlation matrix for each complex (Fig. 4). The residue interaction network (RIN) of the wild-type and the mutants were generated using webserver RING3 [34]. Due to restriction of input trajectory file size of <200MB, the trajectories of wild-type and mutants were reduced to 86 frames which covers 300ns of simulation time. Supplementary table 4, shows the 100 % contact frequency of the H-bond, ionic and van der Waals interactions between MSH2 and MSH3 throughout the simulation. The conformation dependent contact map of H-bond and van der Waals is plotted for wild-type and mutants (Supplementary Fig. 2).

Results

Validation of the six non-synonymous variants of MSH3 and establishment of their novel status

Using targeted exome sequencing on a subset of 54 African American colorectal cancer (AA-CRC) samples, we identified six different non-synonymous variants in the *MSH3* gene with different frequencies ranging from 0.036 to 0.071 (Table 1). These variants included specific nucleotide substitutions such as G-A and C-T alterations in exon 8 (substitution at nucleotide 1237 from GAG to AAG), exon 9 (substitution at nucleotide 1397 from AGC to AAC), exon 20 (substitution at nucleotide 2759 from TCC to TTC), exon 21 (substitution at nucleotide 2926 from GAA to AAA), exon 22 (substitution at nucleotide 3028 from CAT to TAT), and exon 23 (substitution at nucleotide 3241 from GAA to AAA). The distribution of these variants was found to be present in different domains (MutS-II and MutS_V) of the MSH3 protein, particularly in the ATPase and MSH3-MSH2 interaction regions, which are important for the protein's function (Fig. 1A). Additionally, these variants were validated using targeted Sanger sequencing, which further confirmed the presence of the identified variants of *MSH3* and were consistent with the targeted exome sequencing data (Fig. 1B-1G). In sum, the study's findings demonstrate the potential of targeted exome sequencing in identifying genetic variants of *MSH3* and might be associated with AA-CRC.

Computational evaluation of newly discovered non-synonymous MSH3 variants for novelty and pathogenicity

The mutation landscape and allelic frequency of *MSH3* sequencing data of African Americans (AA) from several online databases such as Genome Aggregation Database (gnom AD; consist of 123,136 exomes and 15,496 genomes from 7 populations worldwide) [6]; Exome Aggregation Consortium (ExAC; consists of 60,706 exomes from 7 populations) [38]; 1000 Genomes Project (genomic data for 2,504 individuals from 5 populations) [39] and dbSNP [40] were extensively searched and used. According to the data, none of the 6 variants of exons E8, E9, E20, E21, E22, and E23 have been recorded in any of the databases for African American populations. It is interesting to note that these variants are not present in any other population either, indicating that they are new and unique to *MSH3* in our African American patients. This contrasts with the common *MSH3* variants found in Caucasian populations and *MSH3* variants assigned with dbSNP ID in African Americans. The clinical significance of these *MSH3* variants is uncertain, as determined by the InterVar database, and they have not yet been assigned an ID in the dbSNP database. (Supplementary Table 1). The significance of this information sheds light on the genetic diversity and highlights the need for further research in understanding the impact of these variants and the uncertain clinical significance and to determine their potential effects and implications for AA patients. Furthermore, the Cosmic database (currently contains information on over 11 million coding mutations across more than 1 million tumor samples from over 20,000 genes across different cancer types) was analyzed to confirm the novelty status of six different identified variants (c.G1237A; c.C2759T; c.G1397A; c.G2926A; c.C3028T and c.G3241A) corresponding to amino-acid residues substitution (p.E413K; p.S920F; p.S466N; p.E976K; p.H1010Y; p.E1081K) of *MSH3* gene nucleotide and amino-acid positions, respectively in different normal and cancer tissues. Data analysis revealed the variants as novel and not previously reported in any population and any database including AA-CRC (Fig. 2A-F).

However, in whole exome sequencing research, accurate and precise prediction of the deleteriousness of novel and nonsynonymous variants is essential for separating pathogenic mutations from background

Table 1
Non-synonymous variants in MSH3 gene.

Locus	Ref	Alt	Annotation	Amino Acid Change	Observed	Frequency	Domain
79974809	G	A	SNV	Exon8: c. G1237A: p. E413K	2	2/54=0.037	MutS-II
80021328	G	A	SNV	Exon9: c. G1397A: p. S466N	3	3/54=0.055	MutS-II
80109506	C	T	SNV	Exon20:c.C2759T: p. S920F	4	4/54=0.074	MutS-V
80150061	G	A	SNV	Exon21: c. G2926A: p. E976K	2	2/54=0.037	MutS-V
80160659	C	T	SNV	Exon22:c.C3028T: p.H1010Y	2	2/54=0.037	MutS-V
80169045	G	A	SNV	Exon23: c. G3241A: p. E1081K	2	2/54=0.037	MutS-V

Note: Ref: Reference; Alt: Alteration; SNV: Nonsynonymous variation; adenine (A), cytosine (C), guanine (G), and thymine (T)

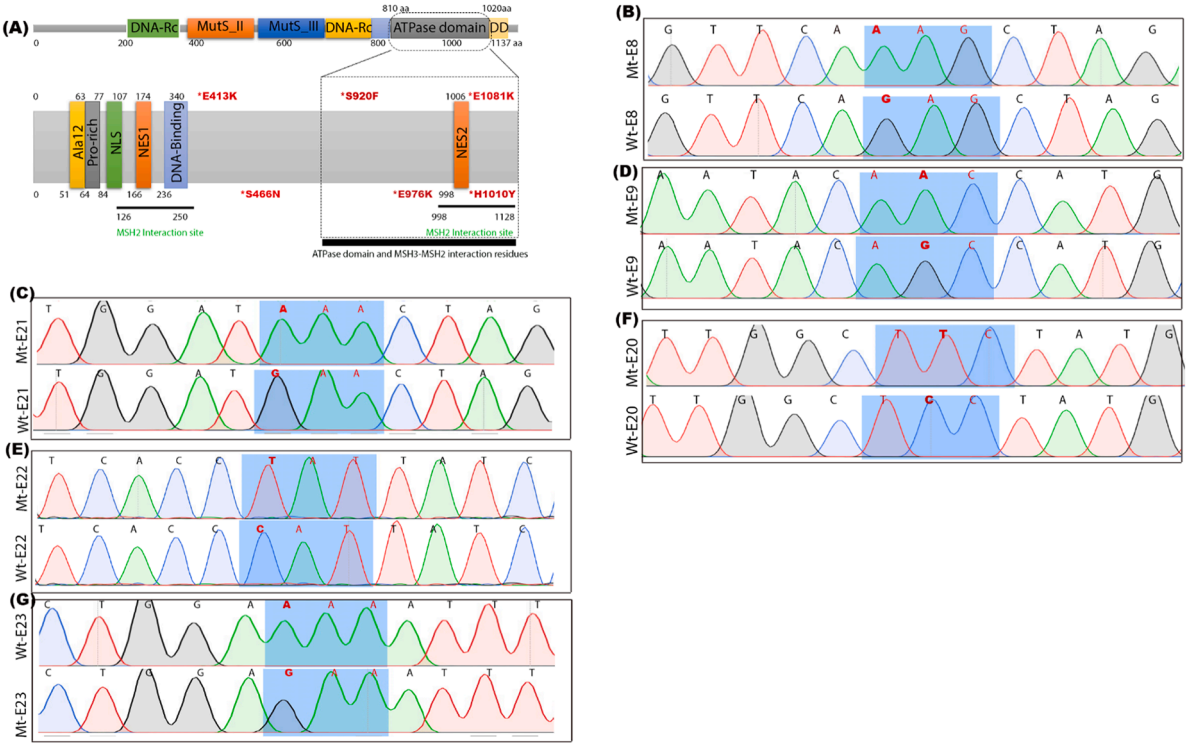

Fig. 1. Targeted Sanger sequencing of MSH3 gene in AA-CRC: (A) Schematic diagram of all identified six different variants of MSH3 (p.E413K; p.S920F; p.S466N; p. E976K; p.H1010Y and p.E1081K) were located in different domains (MutS_II and MutS_V) of MSH3, highlighted with bold red color with asterisk mark* and red dotted box with ATPase domain and MSH2-MSH3 interaction site (B) The chromatogram representation of wild (Wt) and mutants type (Mt) exon 8 (E8) substitution (GAG to AAG)(1B); exon 9 (E9) substitution (AGC to AAC) (1C); exon 20 (E20) substitution (TCC to TTC) (1D); exon 21 (E21) substitution (GAA to AAA) (1E); exon 22 (E22) substitution (CAT to TAT)(1F); exon 23 (E23) substitution (GAA to AAA) (1G) respectively in MSH3 gene were confirmed by Sanger sequencing.

polymorphisms. Even though numerous approaches for predicting deleteriousness have been established, their predictions occasionally vary with one another. The novelty of the six variants was confirmed by multiple database searches. Therefore, we thoroughly assessed multiple algorithms including 23 Function Prediction Algorithms, 4 Conservation Scores and 2 Ensemble Scores (see Materials and Methods). Data indicated that the overall calculated D:A ratio (Number of algorithms predicting variants to be Deleterious: Total *in-silico* algorithms) for each mutation like 21:23, 15:23, 21:23, 22:23, 22:23 and 13:23 for p.E413K, p.S466N, p.S920F, p.E976K, p.H1010Y and p.E1081K, respectively (Table 2). Furthermore, the damaging score (proportion of algorithms predicted to be deleterious (damaging score of loss-of-function variant is deemed to be 1) was analyzed for all six variants showing more than 0.9 value damaging score, suggesting loss-of-function variant for p.E413K, p.S920F, p.E976K and p.H1010Y. Also, extreme score for all variants was calculated (the loss-of-function and damaging (damaging score greater than 0.5) non-synonymous with allele frequency less that 0.0001 are regarded as extreme variants. Interestingly, data suggested that all the variants were extreme and were damaging in nature. The detailed description of all different databases' search results is provided (Supplementary Table 2). Overall, based on the criteria assigned to each database, it was revealed that these unique variants of the MSH3 protein are predicted to be of harmful and deleterious nature. Similar results were corroborated with VarCards, an integrated genetic and clinical database, which found all six identified *MSH3* variants as novel, pathogenic and deleterious [41]. The significance of these databases provides valuable resources for researchers and clinicians to interpret genetic pathogenic variants from background polymorphisms and improve our understanding of the genetic basis of human diseases for *in-vivo* studies.

Impact of MSH3 variants on MSH2 protein complex and ATPase domain of MSH3 using molecular dynamics simulation (MDS)

To assess amino-acid substitutions' effects on MSH3 protein structure and stability between wild-type and mutated MSH3 residues, we used multiple bioinformatic tools and molecular dynamics simulations (MDS). We explore the potential structural effects of the identified six *MSH3* variants (p.E413K, p.S466N, p.E976K, p.S920F, p.E1081K, and p. H1010Y) with their cognate interacting partner MSH2 protein and ATPase domain interaction using MDS via GROMACS v2020. The wild-type and MSH3 variants were modelled (PDB ID: 3THX) with wild type

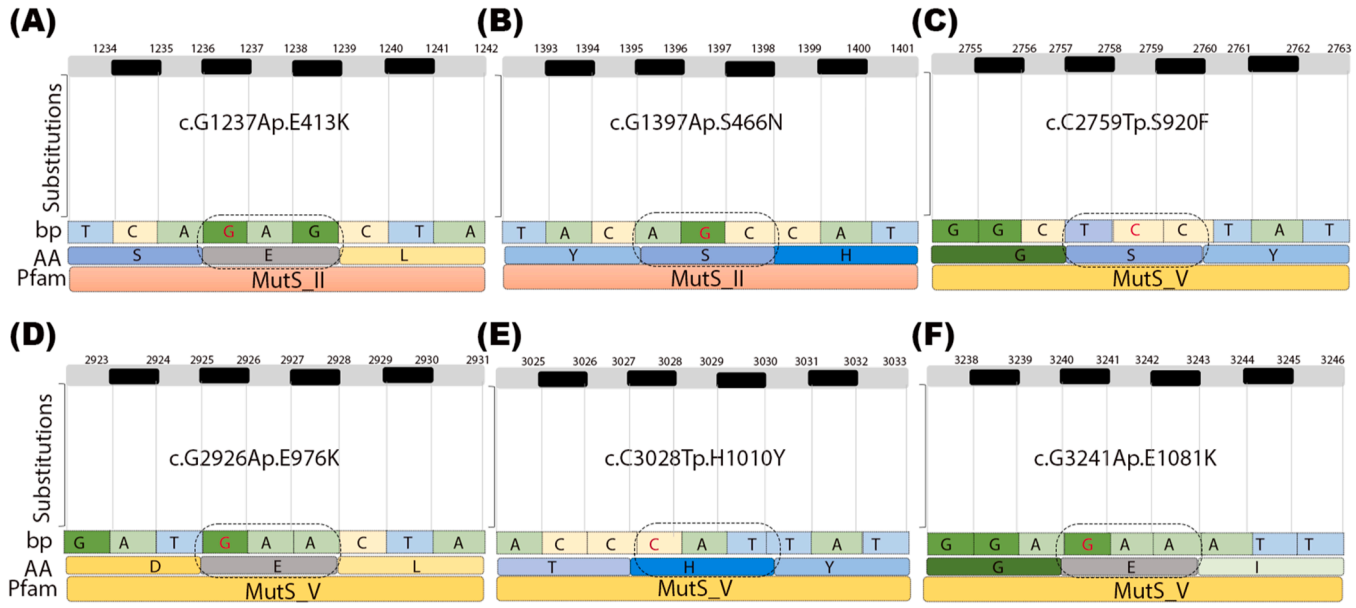

Fig. 2. Novelty of all 6 variants of MSH3 was confirmed by cosmic database: The MSH3 variants in different exons included exon 8, 9, 20, 21, 22, and 23 representing different domains of MSH3 with varied frequency. (A-D) Novelty of all six different variants of MSH3 were confirmed by cosmic DB; X-axis represents DNA sequence in base pairs (bp), protein sequence in amino acids, domain information by Pfams database and Y-axis depicts substitution, c=cDNA and p= represent protein; Red marked (dashed encircle and square box) represents the exon 8 substitution in nucleotide 1237 (GAG to AAG); exon 9 substitution in nucleotide 1397 (AGC to AAC); exon 20 substitution in nucleotide 2759 (TCC to TTC), exon 21 substitution in nucleotide 2926 (GAA to AAA); exon 22 substitution in nucleotide 3028 (CAT to TAT); exon 23 substitution in nucleotide 3241 (GAA to AAA) respectively in MSH3 gene.

Table 2
Summary of *in silico* analysis of the six MSH3 mutations from different algorithms.

Gene (Cytoband)	Location	Ref	Alt	Effect	Amino acid change	D:A algorithms	Damaging score	Extreme
MSH3 (5q14.1)	chr5:79974809-79974809	G	A	SNV	Exon8: c.1237G>A:p.E413K	21:23	0.96	Yes
MSH3 (5q14.1)	chr5:80021328-80021328	G	A	SNV	Exon9: c.1397G>A:p.S466N	15:23	0.70	Yes
MSH3 (5q14.1)	chr5:80109506-80109506	C	T	SNV	Exon20: c.2759C>T:p.S920F	21:23	0.91	Yes
MSH3 (5q14.1)	chr5:80150061-80150061	G	A	SNV	Exon21: c.2926G>A:p.E976K	22:23	1.00	Yes
MSH3 (5q14.1)	chr5:80160659-80160659	C	T	SNV	Exon22: c.3028C>T:p.H1010Y	22:23	1.00	Yes
MSH3 (5q14.1)	chr5:80169045-80169045	G	A	SNV	Exon23: c.3241G>A:p.E1081K	13:23	0.70	Yes

Note: Ref (Reference nucleotide); Alt (Alteration nucleotide); SNV (Single Nucleotide Variant); D:A Algorithms (Number of algorithms predicted to be deleterious: Total in silico algorithms, for example 22:23 means p.E413K is found as damaging in 22 out of 23 algorithms) Damaging score (Proportion of algorithms predicted to be deleterious (damaging score of loss-of-function variant is deemed to be 1); Extreme (The loss-of-function and damaging (damaging score greater than 0.5) non-synonymous with allele frequency less that 0.0001 are regarded as extreme variants [41].

MSH2 complexes and MDS trajectories of MSH2-MSH3 complexes were processed to calculate the root mean square deviation (RMSD) for the conformational difference between the backbones of the MSH2-MSH3 complex from 0ns to 300ns for wild-type and six variants (Fig. 3B-C). Data indicated that MSH2 protein showed a maximum average of 0.64nm RMSD (SD 0.09) compared to 0.49nm RMSD (SD 0.06) of MSH3 protein. Furthermore, in comparison to the wild type, the RMSD of p. E413K, p.S466N, and p.E1081K is higher than p.E976K, p.H1010Y, and p.S920F in MSH2, whereas the only difference in MSH3 is that the RMSD of p.S466N is lower than the wild type. Suggesting, that the change in RMSD is due to the fluctuations in some regions of domains in the protein complexes. Therefore, a comparative Root Mean Square Fluctuations (RMSF) plots of MSH2 and MSH3 were analyzed. In the MSH2 protein, variant p.E413K showed high fluctuations in the region from 470 to 550 residues while variant p.E1081K showed high fluctuations in the residues from 880 to 910. Similarly, in the MSH3 protein mutant, p. E413K showed high fluctuations in the region from 720 to 770 residues while variant p.E1081K showed high fluctuations in the residues from 1030 to 1040. The binding affinity between the MSH2 and MSH3 protein in the complex was predicted using gmx_MMPBSA tool.

Furthermore, the RMSD values were corroborated with the average number of inter-molecular hydrogen bonds between MSH2 and MSH3 per timeframe. The lower number of inter-molecular hydrogen bonds indicated that the molecular interaction between MSH2 and MSH3 is less and may impart conformational changes which leads to higher RMSD and vice versa. We found four of the variants (p.S920F; p.E976K; p.H1010Y; p.E1081K) in domain V of MSH3, corresponding to ATPase activity of MSH3 protein. MDS analysis depicts the number of hydrogen bonds between MSH2 and MSH3 at the interface of ATPase binding domain were 9.32 (SD 2.1), 18.37 (SD 2.27), 13.5 (SD 3.03), 13.4 (SD 1.73) and 13.9 (SD 1.99) for wild-type, S920F, E976K, H1010Y and E1081K complexes respectively. The number of hydrogen bonds

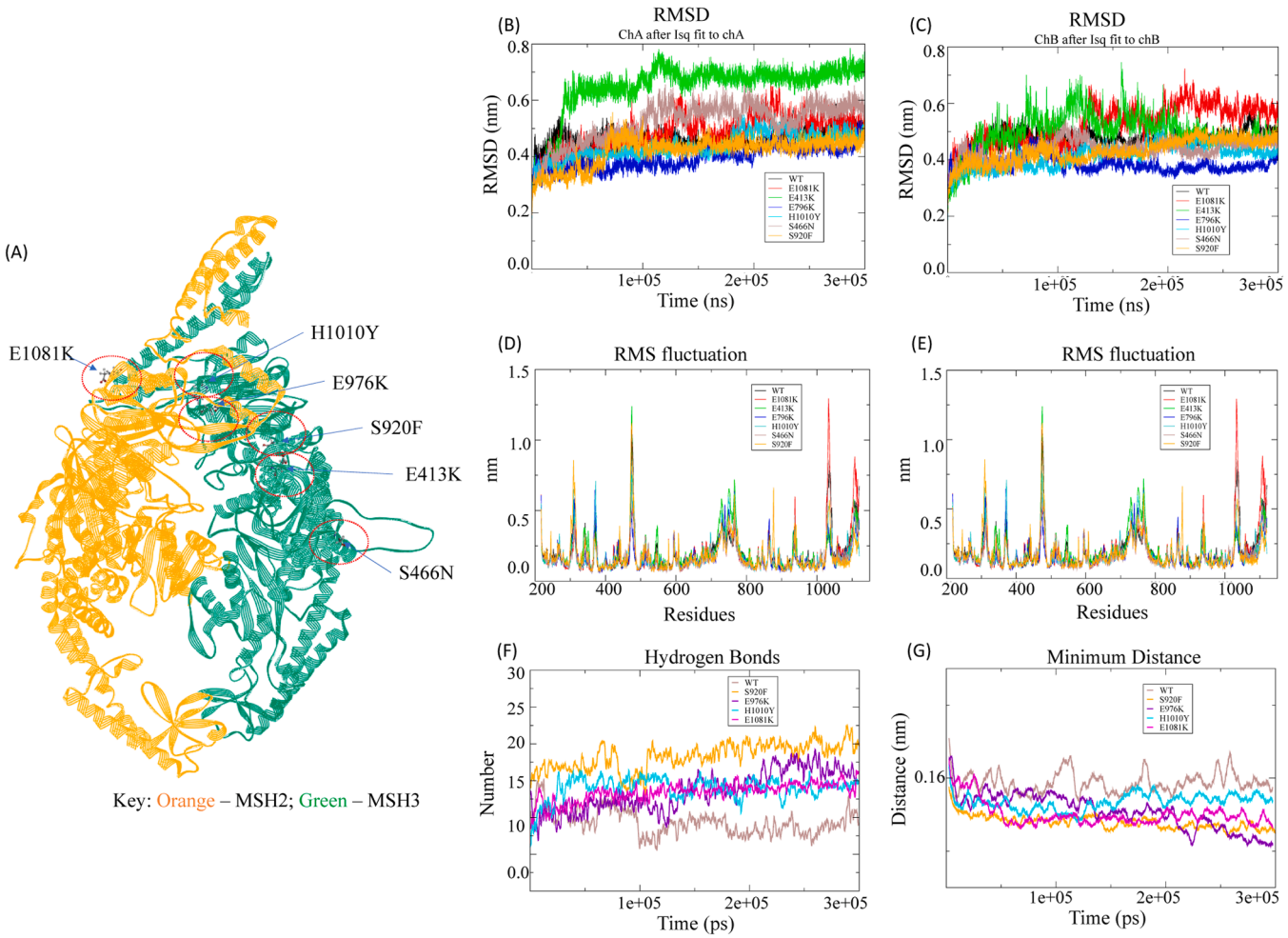

Fig. 3. Structural and bonding impact of identified six novel variants of MSH3 using Homology modeling and molecular dynamics: (3A) The MSH2-MSH3 complex (PDB ID: 3THX) remodeling and molecular dynamic simulation (MDS) via GROMACS v2020 of wild-type and mutant p.E413K, p.S466N, p.S920F, p.E976K, p.H1010Y and p.E1081K, (3B-C) The root mean square deviation (RMSD) of MSH2 and MSH3 from 0ns to 300ns for wild-type and six variants. (3D-E) Root Mean Square Fluctuations (RMSF) of wild-type and variants (p.E413K, p.S466N, p.S920F, p.E976K, p.H1010Y and p.E1081K) were analyzed. Y-axis represents RMSD or RMSF (nm; nanometer or ang:Angstrom respectively) and X-axis depicts time or residue (picosecond: ps or amino acids;AA). (3F-G) Number of hydrogen-bonds and minimum distance was calculated between MSH2-MSH3 complex near ATPase binding domain of Wild-type, S920F, E976K, H1010Y and E1081K; Minimum distance between any pair of atoms from ATPase domain of MSH3 and MSH2 of Wild-type, S920F, E976K, H1010Y and E1081K; Minimum distance at ATPase domain was calculated by mindist command of gromacs in wild-type and mutant p.E413K, p.S466N, p.S920F, p.E976K, p.H1010Y and p.E1081K of MSH3 protein.

between MSH2 and MSH3 near ATPase domain is directly proportional to binding affinities which is reflected in MMPBSA values. The binding energies of wild-type, S920F, E976K, H1010Y and E1081K are -176.08 (SD 9.5), -245.41 (SD 10.8), -195.22 (SD 14.2), -187.31 (SD 7.8) and -182.57 (SD 8.2) kcal/mol respectively at the ATPase interface domain of MSH2-MSH3 (Supplementary Table 3A). Furthermore, the minimum distance between any pair of atoms between the MSH2 and MSH3 at ATPase domain is inversely correlated with average number of hydrogen bonds and MMPBSA values, which means that in wild type the number of hydrogen bonds between the complex is less, the binding affinity is also less, hence the minimum distance between any pair of atoms in the interface is more as compared mutant complexes (Fig. 3F-G). Other bonding patterns (Vander Wall's, electrostatic and total free energy was affected; suggesting these mutations might alter ATPase activity and folding and induce genomic instability or EMAST phenotype.

It is however worth noting that some of the detected variations were not in the ATPase domain but in the MSH2-MSH3 interacting domain. Such variations might not affect MSH3 function but do affect its ability to form stable heterodimer structures with MSH2 and leads to genomic instability. Overall, our findings suggest that these variants may affect the stability and interactions of the MSH3 protein, potentially leading to changes in its function and contributing to the development of diseases or other pathologies. *In-vitro* experiments are currently underway to accurately assess the effect of these variants on MSH3 function in AA-CRC.

Principal component analysis (PCA), normal mode analysis (NMA) and dynamics cross-correlation matrix (DCCM) and residue interaction network (RIN)

The Principal Component Analysis (PCA) method was used to visualize and reduce the complexity of Root-Mean-Square Deviation (RSMD) trajectory data using R-bioconductor package 'bio3d'. The PC1 accounts for the largest proportion of variance followed by the next highest variance by PC2. Supplementary Fig. 1 showed that PC1 is in decreasing order for E1081K (39.8 %), S466N (38.2 %), E413K (37.9 %), WT (33.3 %), H1010Y (31.4 %), E976K (30.2 %), S902F (29.3 %) while PC2 for E413K (19.23 %), WT (14.69 %), E1081K (14.36 %), S902F (12.39 %), E976K (10.43 %), H1010Y (10.42 %), S466N (8.39 %). Considering the PC1 and PC2 suggest that the proportion of variance (more than 50 %) is uncorrelated in E1081K and E413K which is greater than WT, evident by two fluctuating domains in E1081K and E413K as per NMA (Fig. 4B and

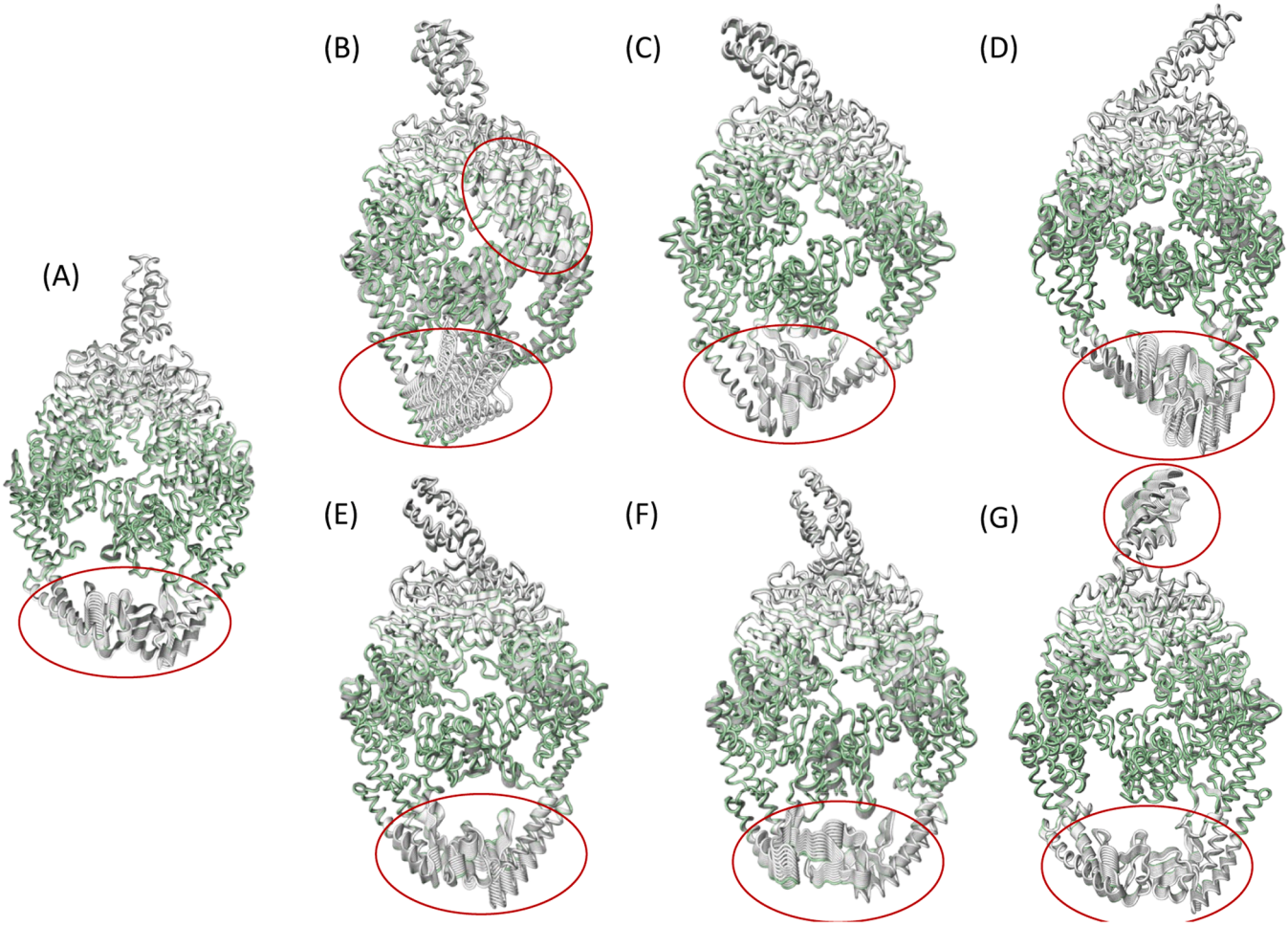

Fig. 4. The Normal Mode Analysis (NMA) representation analysis of seven superimposed non-trivial modes. The color-coded structures represent the mobility observed in the NMA results, with a color in green (indicating the least mobility) to shades of white (indicating the highest mobility) of WT(A), E413K (B), S466N (C), S920F(D), E976K(E), H1010Y (F), and E1081K (G) were analyzed in this context using DynaMut tool.

4G). This result suggests that E413K variant imparts more conformational changes in MSH3 protein while E1081K impart more conformational changes in MSH2-MSH3 proteins' interaction.

Furthermore, we utilized DynaMut webserver to conduct NMA and DCCM analysis, which generates a three-dimensional matrix depicting the time-dependent relationships among protein residues (wild and mutants).The dynamics cross-correlation matrix (DCCM) of E1081K and E413K also showed more anti-correlated residues as compared to WT (Fig. 5A, 5B, and 5G). The mutant S466N has higher PC1 than WT, but the lower PC2 contribution leads to more correlated residues than WT as seen in Fig. 5A and 5C. The mutants E976K, S902F and H1010Y have no significant difference in the proportion of variance compared to WT, which are corroborated with the correlation of residues in the DCCM plots (Fig. 5A, 5D,5E, 5F) and normal mode analysis (Fig. 4A, 4D, 4E, 4F).

The RING3 webserver was used to access the intermolecular interactions between MSH2 and MSH3 (Supplementary Table 4). The variant E1081K has the least number of H-bond compared to WT, which is evident by two fluctuating domains in NMA (Fig. 4G). E413K also has two fluctuating domains in NMA, but the conformational changes in the center region of MSH3 (Fig. 4B) may lead to increase in number of hydrogen bonds. S466N has maximum number of hydrogen bonds and van der Waals interaction compared to WT, which is evident by the less anti-correlated residues in DCCM plot of S466N complex (Fig. 5C).

The conformational dependent contact map of all the complexes shows the distinct change of intermolecular interaction (H-bond and Van der Waals) in 86 frames flanking by five residues from the site of mutation (Supplementary Fig. 2). Surprisingly, the change in charge from negative to positive for E413K variant led to no change in frequency of H-bond but increase in frequency of Van der Waals seen. This may be due to the formation of new H-bonds by change in conformation seen in the two domains. For S466N, there is increase in frequency of H-bond and Van der Waals interaction due to more hydrophilic nature of asparagine. For S902F, no change frequency of H-bond but increase in frequency of Van der Waals, which is due to more hydrophobic nature of phenylalanine. For H1010Y is a decrease in frequency of H-bond but increase in frequency of van der Waals due positive charge on histidine and more hydrophobicity of tyrosine. For variant E1081K and E967K, decrease in frequency of both H-bond and Van der Waals interaction due to change in overall charge of the amino-acid.

Putative molecular mechanisms of the six novel variants

Various web-based tools such as GERP++, phyloP, fitCons, SiPhy and PolyPhen-2 were used to predict the evolutionary conservation of amino acids. Interestingly, we observed that all six novel and pathogenic missense variants of MSH3 are conserved evolutionary from lower to higher hierarchy (Supplementary Table 5 and Supplementary Fig. 3). Furthermore, their putative molecular mechanisms were also analyzed using MutPred2 database. MutPred2 predicts the functional consequences of amino acid substitutions in proteins using a combination of evolutionary conservation analysis, structural analysis, and machine

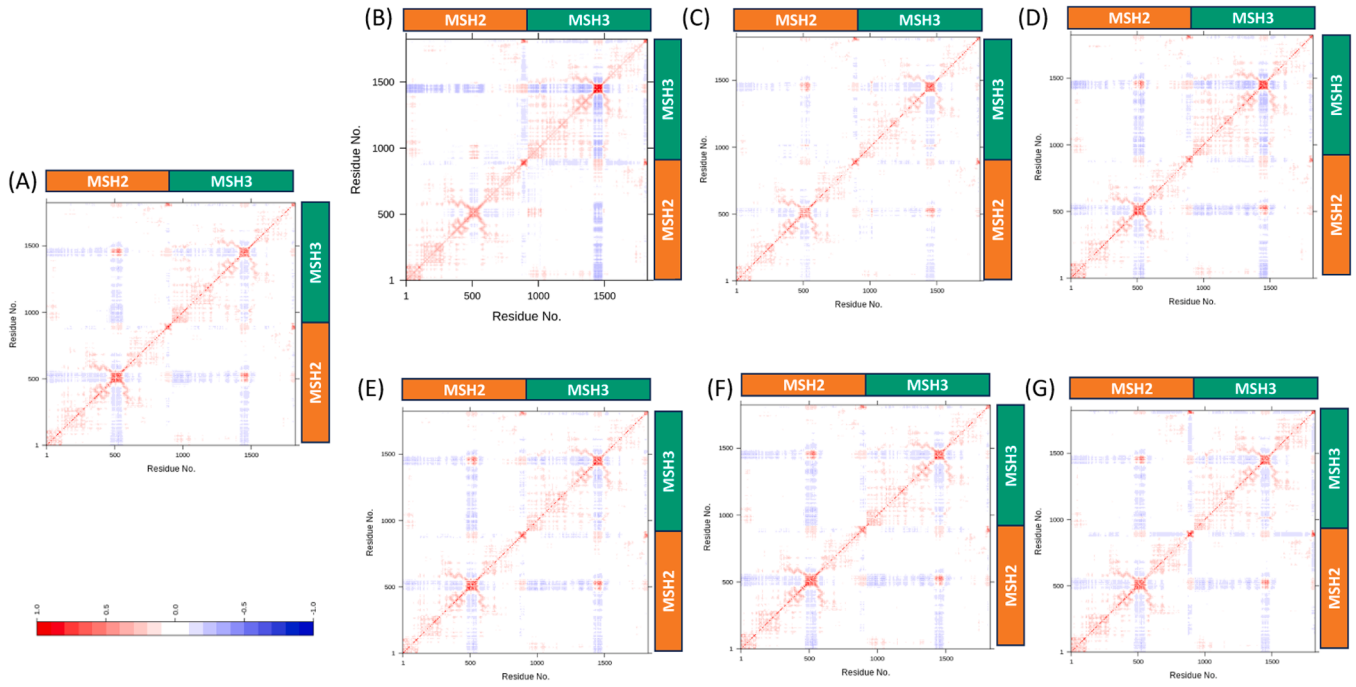

Fig. 5. The representation of dynamics cross-correlation matrix (DCCM) analysis plot: The DCCM matrix was computed for the Cα atoms during 300 ns of molecular dynamics simulations, and the resulting matrix illustrates the degrees of correlated and anti-correlated motions. These motion relationships are visually represented by a color gradient ranging from red to blue, where red indicates positive correlations and blue signifies negative correlations. The different protein variants with different mutations, namely Wild type of WT(A), E413K (B), S466N (C), S920F(D), E976K(E), H1010Y (F), and E1081K (G) were analyzed in this context using DynaMut tool.

learning algorithms and provides information on the potential impact of substitutions on protein-protein interactions, post-translational modifications, and functional domains of the protein. Interestingly, data revealed that loss and gain of relative solvent accessibility of p.S920F and p.E976K respectively, also loss of catalytic site at S920, T981 and T981 in p.S920F, p.E976K and p.H1010Y of MSH3 protein respectively was observed. Similarly, we found loss of allosteric site at Y921 and p.Y1011 upon substituting of S920F and p.H1010Y in MSH3 protein respectively. Additionally, loss of allosteric site at Y921 in p.S920F substituting in MSH3 protein suggesting that p.E976K, p.H1010 and p.S920F (Table 3). Data indicated that these variants are very important as they are present in ATPase domain V of MSH3 protein, therefore, we hypothesis that these variants (p.H1010, p.H1081K) might alter the ATPase activity of MSH3 and alters its function that further needs to be deciphered further.

Discussion

There are various challenges to assign a novel mutation (common, benign, rare or pathogenic) in AA genomes compared to other population, because assessing a pathogenic variant depends on searching publicly accessible genomic reference databases, where African data are underrepresented [6,42]. Additionally, the SNPs deposited in public database and the bioinformatic tools used to determine SNPs are from non-African population based [43,44]. Due to great genetic diversity, it is likely to discover many new SNPs and/or variants that have not yet

Table 3
Putative molecular mechanism linked with pathogenic six novel mutations of MSH3.

Substitution	MutPred2 score	Affected PROSITE and ELM Motifs	Molecular mechanisms with P-values <= 0.05	Probability	P-value
p.E413K	0.538	ELME000053	-	-	-
p.S466N	0.511	ELME000197, ELME000336	Altered Transmembrane protein;	0.14;	0.02;
			Gain of Sulfation at Y465	0.01	0.05
p.S920F	0.874	None	Loss of Relative solvent accessibility.	0.3;	0.009;
			Loss of Allosteric site at Y921;	0.21;	0.04;
			Altered Metal binding;	0.16;	0.02;
			Loss of Catalytic site at S920	0.12	0.03
p.E976K	0.885	ELME000333, PS00486	Altered Metal binding;	0.65;	0.0001;
			Altered Ordered interface;	0.27;	0.04;
			Gain of Allosteric site at R979;	0.27;	0.005;
			Loss of Catalytic site at T981;	026;	0.004;
			Gain of Relative solvent accessibility;	0.24;	0.04;
				0.12	0.03;
p.H1010Y	0.777	ELME000142, ELME000336	Altered Metal binding;	0.5;	0.0038;
			Altered Ordered interface;	0.32;	0.001;
			Loss of Allosteric site at Y1011;	0.3;	0.003;
			Loss of Catalytic site at H1010	0.27	0.003
p.E1081K	0.413	-	-	-	-

Note: >0.5 MutPred2 score pathogenic and < 0.5 is benign; PROSITE motifs and Eukaryotic Linear Motif (ELM) database describing protein domains functional significances. Probability score threshold of 0.50 would suggest pathogenicity.

been reported in existing public databases [45]. Furthermore, the identification of genes that can be used for therapeutic purposes in African genomic research is also missing. *In-silico* prediction algorithms can aid in functional analysis of variants from various populations, The allele frequency of different exons (E8, E9, E20, E21, E22, E23) of *MSH3* variants in different population were analyses as shown in Table 2, and we found six novel non-synonymous variants of *MSH3* in AA-CRC samples. Multiple bioinformatic tools (Supplementary Table 2) and molecular dynamics simulations (MDS) were used to assess amino-acid substitutions' effects on MSH3 protein structure and stability between wild type (Wt.) and mutated (Mt) MSH3 residues. We found that six different non-synonymous variants in different exons of MSH3 gene (c. G1237A, c.C2759T, c.G1397A, c.G2926A, c.C3028T, c.G3241A) were deleterious in nature. Previous reports have shown that a protein's ability to function may be impacted by amino-acid substitutions brought on by variation in the protein-coding region, which can also result in pathogenicity [46,47]. In our study, these novel pathogenic variants in MSH3 protein (*p.*E413K; *p.*S466N; S *p.*920F; *p.*E976K; *p.*H1010Y; *p.* E1081K) immensely alters protein stability, structure and chemical bonding which affect its function and increase risk of diseases such as CRC in AA individuals.

Reports have shown that studying the nature of the protein's function requires reliable predictions of the variant's impact on the stability of protein structure calculated by unfolding Gibbs free energy change ($\Delta\Delta G = G$) in Kcal/mol [48,49]. Similarly, MSH3 wild type and variant structure stability was calculated by ($\Delta\Delta G = \Delta G$ mutant – ΔG wild type) using different tools (MUpro, and DynaMut) and showed instability indicating that the variant was more pathogenic. Based on our results, p.E413K; p.S466N; p.E976K; and p.E1081K substitutions showed negative ΔG and positive ΔS which indicated instability of MSH3 protein and might lead to rapid degradation compared to wild type MSH3 protein. However, the other two variants showed a positive ΔG and a negative ΔS for p.S920F and p.H1010Y variants suggesting MSH3 might attain more stability and new interacting partners and hence show deleterious effects. Previous reports have shown that Walker A (ATP binding) and Walker B (ATP hydrolysis) motifs of ATPase domain have mutations (G769A, D870A, G795A) that will likely lead to non-functional MSH3 and lead to EMAST instability in the genome [50]. Since, we found variants (p.E976K; p.E976K; p.H1010Y; p.E1081K) in MSH3 were very close or within the domain V (Walker B) which has ATPase activity and other bonding patterns (Van der Waals, electrostatic and total free energy was affected, suggesting these variants might alter ATPase activity and folding and induce genomic instability or EMAST phenotype. The study detected variations in the MSH3 protein that may affect its stability and interactions, which may potentially contribute to the development of diseases like CRC in African American (AA) populations. In-vitro experiments are underway to better understand the impact of these variants on MSH3 function in AA-CRC. Overall, the study provides important insights into the structural impact of *MSH3* variants, which could help in understanding the underlying mechanisms of MSH3-based DNA mismatch repair deficiency and its associated AA-CRC phenotypes and outcomes. We have initiated CRISPR-Cas9 knock-in these mutations in model systems having MSH3 in wild type-like SW620 CRC cell line and organoids generated from African American (AA) CRC parents; with time we will delineate the mechanism of these mutations associated with either EMAST or other mal-functions affiliated with these mutations as we have successfully generated these mutations. Functional assays are underway to fully understand the impact of these variants on MSH3 function and CRC development in AA patients.

Conclusion

The combination of molecular dynamics simulations and bioinformatic analyses confirmed that the six novel AA CRC MSH3 variants could affect the structure and function of the MSH3 protein. Therefore, it is quite likely that these unique novel mutations will change the way the MSH3 protein functions as a disease-related trait with EMAST phenotype in AA colon carcinogenesis.

All the Supplementary data (figure and tables) are provided at the end of manuscript.

CRediT authorship contribution statement

Mudasir Rashid: Data curation, Formal analysis, Writing – review & editing. **Rumaisa Rashid:** Data curation. **Nikhil Gadewal:** Data curation, Software. **John M. Carethers:** Writing – review & editing. **Minoru Koi:** Data curation. **Hassan Brim:** Writing – review & editing. **Hassan Ashktorab:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Authors thanks to Dr. Rohit Kumar (Department of Medicine, Baylor College of Medicine, Houston, TX, USA) for his intensive discussion regarding *in-silico* analysis of exome and genomic data. This work was supported by the United States Public Health Service (R01 CA258519). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2024.100970.

References

[1] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.L. Borresen-Dale, et al., Signatures of mutational processes in human cancer, Nature 500 (7463) (2013) 415–421.
[2] B. Afsari, A. Kuo, Y. Zhang, L. Li, K. Lahouel, L. Danilova, A. Favorov, T. A. Rosenquist, A.P. Grollman, K.W. Kinzler, et al., Supervised mutational signatures for obesity and other tissue-specific etiological factors in cancer, Elife 10 (2021).
[3] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, P.J. Campbell, M.R. Stratton, Deciphering signatures of mutational processes operative in human cancer, Cell rep. 3 (1) (2013) 246–259.
[4] P. Georgeson, T.A. Harrison, B.J. Pope, S.H. Zaidi, C. Qu, R.S. Steinfelder, Y. Lin, J. E. Joo, K. Mahmood, M. Clendenning, et al., Identifying colorectal cancer caused by biallelic MUTYH pathogenic variants using tumor mutational signatures, Nature commun. 13 (1) (2022) 3254.
[5] F. De Nicola, F. Goeman, M. Pallocca, F. Sperati, L. Pizzuti, E. Melucci, B. Casini, C. A. Amoreo, E. Gallo, M.G. Diodoro, et al., Deep sequencing and pathway-focused analysis revealed multigene oncodriver signatures predicting survival outcomes in advanced colorectal cancer, Oncogenesis. 7 (7) (2018) 55.
[6] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, et al., Analysis of protein-coding genetic variation in 60,706 humans, Nature 536 (7616) (2016) 285–291.
[7] S.A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, C.G. Cole, S. Ward, E. Dawson, L. Ponting, et al., COSMIC: somatic cancer genetics at high-resolution, Nucleic acids res. 45 (D1) (2017) D777–D783.
[8] L.B. Alexandrov, J. Kim, N.J. Haradhvala, M.N. Huang, A.W. Tian Ng, Y. Wu, A. Boot, K.R. Covington, D.A. Gordenin, E.N. Bergstrom, et al., The repertoire of mutational signatures in human cancer, Nature 578 (7793) (2020) 94–101.
[9] E. Vilar, S.B. Gruber, Microsatellite instability in colorectal cancer-the stable evidence, Nat. rev. Clin. oncol. 7 (3) (2010) 153–162.
[10] H.T. Nguyen, H.Q. Duong, The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy, Oncol. lett. 16 (1) (2018) 9–18.
[11] N. Pecina-Slaus, A. Kafka, I. Salamon, A. Bukovac, mismatch repair pathway, genome stability and cancer, Front. Mol. Biosci. 7 (2020) 122.
[12] N.V. Romanova, G.F. Crouse, Different roles of eukaryotic MutS and MutL complexes in repair of small insertion and deletion loops in yeast, PLoS genetics 9 (10) (2013) e1003920.
[13] G.M. Li, Mechanisms and functions of DNA mismatch repair, Cell Res. 18 (1) (2008) 85–98.

[14] J.M. Park, S. Huang, D. Tougeron, F.A. Sinicrope, MSH3 mismatch repair protein regulates sensitivity to cytotoxic drugs and a histone deacetylase inhibitor in human colon carcinoma cells, PLoS. One 8 (5) (2013) e65369.
[15] V. Lee, A. Murphy, D.T. Le, L.A. Diaz Jr., Mismatch repair deficiency and response to immune checkpoint blockade, The oncologist 21 (10) (2016) 1200–1211.
[16] A. Xavier, M.F. Olsen, L.A. Lavik, J. Johansen, A.K. Singh, W. Sjursen, R.J. Scott, B. A. Talseth-Palmer, Comprehensive mismatch repair gene panel identifies variants in patients with Lynch-like syndrome, Molecular gene. genomic med. 7 (8) (2019) e850.
[17] R. Liccardo, M. Lambiase, A. Nolano, M. De Rosa, P. Izzo, F. Duraturo, Significance of rare variants in genes involved in the pathogenesis of Lynch syndrome, Int. j. molec. med. 49 (6) (2022).
[18] S.C. Huang, J.K. Lee, E.J. Smith, R.T. Doctolero, A. Tajima, S.E. Beck, N. Weidner, J.M. Carethers, Evidence for an hMSH3 defect in familial hamartomatous polyps, Cancer 117 (3) (2011) 492–500.
[19] K. Munakata, T. Kitajima, S.S. Tseng-Rogenski, M. Uemura, H. Matsuno, K. Kawai, Y. Sekido, T. Mizushima, Y Toiyama, T. Yamada, M. Mano, E. Mita, M. Kusonoki, M. Mori, J.M. Carethers, Inflammation-associated microsatellite alterations caused by MSH3 dysfunction are prevalent in ulcerative colitis and increase with neoplastic development, Clin. Transl. Gastroenterol. 10 (2019) e00105.
[20] A.C. Haugen, A. Goel, K. Yamada, G. Marra, T.P. Nguyen, T. Nagasaka, S. Kanazawa, J. Koike, Y. Kikuchi, X. Zhong, et al., Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer, Cancer res. 68 (20) (2008) 8465–8472.
[21] J.M. Carethers, M. Koi, S.S. Tseng-Rogenski, EMAST is a form of microsatellite instability that is initiated by inflammation and modulates colorectal cancer progression, Genes 6 (2) (2015) 185–205.
[22] S. Venderbosch, S. van Lent-van Vliet, A.F. de Haan, M.J. Ligtenberg, M. Goossens, C.J. Punt, M. Koopman, I.D. Nagtegaal, EMAST is associated with a poor prognosis in microsatellite instable metastatic colorectal cancer, PLoS. One 10 (4) (2015) e0124538.
[23] S.S. Tseng-Rogenski, K. Munakata, D.Y. Choi, P.K. Martin, S. Mehta, M. Koi, W. Zheng, Y. Zhang, J.M. Carethers, The human DNA mismatch repair protein msh3 contains nuclear localization and export signals that enable nuclear-cytosolic shuttling in response to inflammation, Mol. Cell Biol. 40 (13) (2020).
[24] R. Adam, I. Spier, B. Zhao, M. Kloth, J. Marquez, I. Hinrichsen, J. Kirfel, A. Tafazzoli, S. Horpaopan, S. Uhlhaas, et al., Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis, Am. j. hum. genetics 99 (2) (2016) 337–351.
[25] M. Koi, B.H. Leach, S.S. Tseng-Rogenski, C.A. Burke, J.M. Carethers, Compound heterozygous MSH3 germline variants and associated tumor somatic DNA mismatch repair dysfunction, NPJ Precis. Oncol. 8 (2024) 12.
[26] S.S. Tseng-Rogenski, H. Chung, M.B. Wilk, S. Zhang, M. Iwaizumi, J.M. Carethers, Oxidative stress induces nuclear-to-cytosol shift of hMSH3, a potential mechanism for EMAST in colorectal cancer cells, PLoS. One 7 (11) (2012) e50616.
[27] S.S. Tseng-Rogenski, Y. Hamaya, D.Y. Choi, J.M. Carethers, Interleukin 6 alters localization of hMSH3, leading to DNA mismatch repair defects in colorectal cancer cells, Gastroenterology 148 (3) (2015) 579–589.
[28] G.M. Williams, V. Paschalis, J. Ortega, F.W. Muskett, J.T. Hodgkinson, G.M. Li, J. W.R. Schwabe, R.S. Lahue, HDAC3 deacetylates the DNA mismatch repair factor MutSbeta to stimulate triplet repeat expansions, Proceed. Nat. Acad. Sci. United States of America 117 (38) (2020) 23597–23605.
[29] K. Debacker, A. Frizzell, O. Gleeson, L. Kirkham-McCarthy, T. Mertz, R.S. Lahue, Histone deacetylase complexes promote trinucleotide repeat expansions, PLoS biol. 10 (2) (2012) e1001257.
[30] X. Chen, G. Zhang, P. Li, J. Yu, L. Kang, B. Qin, Y. Wang, J. Wu, J. Zhang, M. Qin, et al., SYVN1-mediated ubiquitination and degradation of MSH3 promotes the apoptosis of lens epithelial cells, The FEBS j. 289 (18) (2022) 5682–5696.
[31] H. Ashktorab, H. Azimi, S. Varma, E.L. Lee, A.O. Laiyemo, M.L. Nickerson, H. Brim, Driver genes exome sequencing reveals distinct variants in African Americans with colorectal neoplasia, Oncotarget. 10 (27) (2019) 2607–2624.
[32] S. Pall, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess, E. Lindahl, Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, J. Chem. Phys. 153 (13) (2020) 134110.
[33] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.
[34] D. Clementel, A. Del Conte, A.M. Monzon, G.F. Camagni, G. Minervini, D. Piovesan, S.C.E. Tosatto, RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles, Nucleic. Acids. Res. 50 (W1) (2022) W651–W656.
[35] B.J. Grant, A.P. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics. 22 (21) (2006) 2695–2696.
[36] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C. X. Hernandez, C.R. Schwantes, L.P. Wang, T.J. Lane, Pande VS: MDTraj: a modern open library for the analysis of molecular dynamics trajectories, Biophys. J. 109 (8) (2015) 1528–1532.
[37] C.H. Rodrigues, D.E. Pires, D.B. Ascher, DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability, Nucleic. Acids. Res. 46 (W1) (2018) W350–W355.
[38] K.J. Karczewski, B. Weisburd, B. Thomas, M. Solomonson, D.M. Ruderfer, D. Kavanagh, T. Hamamsy, M. Lek, K.E. Samocha, B.B. Cummings, et al., The ExAC browser: displaying reference data information from over 60 000 exomes, Nucleic. Acids. Res. 45 (D1) (2017) D840–D845.
[39] C. Genomes Project, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, et al., A global reference for human genetic variation, Nature 526 (7571) (2015) 68–74.
[40] E.M. Smigielski, K. Sirotkin, M. Ward, S.T. Sherry, dbSNP: a database of single nucleotide polymorphisms, Nucleic. Acids. Res. 28 (1) (2000) 352–355.
[41] J. Li, L. Shi, K. Zhang, Y. Zhang, S. Hu, T. Zhao, H. Teng, X. Li, Y. Jiang, L. Ji, et al., VarCards: an integrated genetic and clinical database for coding variants in the human genome, Nucleic. Acids. Res. 46 (D1) (2018) D1039–D1048.
[42] A.B. Popejoy, S.M. Fullerton, Genomics is failing on diversity, Nature 538 (7624) (2016) 161–164.
[43] S. Pabinger, A. Dander, M. Fischer, R. Snajder, M. Sperk, M. Efremova, B. Krabichler, M.R. Speicher, J. Zschocke, Z. Trajanoski, A survey of tools for variant analysis of next-generation genome sequencing data, Briefings bioinform. 15 (2) (2014) 256–278.
[44] R. Bao, L. Huang, J. Andrade, W. Tan, W.A. Kibbe, H. Jiang, G. Feng, Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing, Cancer informatics 13 (Suppl 2) (2014) 67–82.
[45] K. Lebeko, N. Manyisa, E.R. Chimusa, N. Mulder, C. Dandara, A. Wonkam, A genomic and protein-protein interaction analyses of nonsyndromic hearing impairment in cameroon using targeted genomic enrichment and massively parallel sequencing, Omics : j. integrative biol. 21 (2) (2017) 90–99.
[46] F.J. Kaye, R.A. Kratzke, J.L. Gerster, J.M. Horowitz, A single amino acid substitution results in a retinoblastoma protein defective in phosphorylation and oncoprotein binding, Proc. Natl. Acad. Sci. u S. a 87 (17) (1990) 6922–6926.
[47] Y. Bromberg, B. Rost, Correlating protein function and stability through the analysis of single amino acid substitutions, BMC. Bioinformatics. 10 Suppl 8 (Suppl 8) (2009) S8.
[48] T. Mavroconstanti, S. Johansson, I. Winge, P.M. Knappskog, J. Haavik, Functional properties of rare missense variants of human CDH13 found in adult attention deficit/hyperactivity disorder (ADHD) patients, PLoS. One 8 (8) (2013) e71445.
[49] M. Nailwal, J.B. Chauhan, In silico analysis of non-synonymous single nucleotide polymorphisms in human DAZL gene associated with male infertility, Systems biol. reproduct. med. 63 (4) (2017) 248–258.
[50] C. Kumar, G.M. Williams, B. Havens, M.K. Dinicola, J.A. Surtees, Distinct requirements within the Msh3 nucleotide binding pocket for mismatch and double-strand break repair, J. Mol. Biol. 425 (11) (2013) 1881–1898.