Original Article

Identification and analysis of biomarkers for mismatch repair proteins: A bioinformatic approach

Manika Sehgal, Tiratha Raj Singh

Department of Biotechnology and Bioinformatics, Jaypee University of Information and Technology, Waknaghat, Solan, H.P., India.

Address for correspondence:
Tiratha Raj Singh, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan-173234, H.P., India E-mail: tiratharaj@gmail.com

Abstract

Introduction: Mismatch repair is a highly conserved process from prokaryotes to eukaryotes. Defects in mismatch repair can lead to mutations in human homologues of the Mut proteins and affect genomic stability which can result in microsatellite instability (MI). MI is implicated in most human cancers and majority of hereditary nonpolyposis colorectal cancers (HNPCCs) are attributed to defects in MLH1. **Materials and Methods:** In our study we analyzed MLH1 protein and the associated nucleotide and other protein sequences. The protein sequences involved in mismatch repair in different organisms have been found to be evolutionary related. Several other related proteins to MLH1 have also been identified through protein–protein interactions. All associated proteins are either mismatch repair proteins or associated with MLH1 in various pathways. Pathways information was also confirmed through MMR and other pathways in KEGG. QSite Finder showed that the active site of MLH1 protein involves residues from the conserved pattern and is involved in ligand–protein interactions and could be a useful site. To analyze linkage disequilibrium (LD) and common haplotype patterns in disease association, we performed statistical haplotype analysis on HapMap genotype data of SNPs genotyped in population CEU on chromosome 3 for MLH1. **Results:** Various markers have been found and LD plot was also generated. Two distinct blocks have been identified in LD plot which can be independent region of action, and there is involvement of 7 and 17 markers in first and second blocks, respectively. **Conclusion:** Overall correlation of 0.95 has been found among all interactions of genotyped SNPs which is significant.

Key words: Haplotype, hyper conservation, linkage disequilibrium, mismatch repair system and proteins, multiple sequence alignment

INTRODUCTION

DNA mismatch repair is a process that takes place in the cells of almost every living organism, both prokaryotic and eukaryotic because of its evolutionary importance. The first evidence for mismatch repair was obtained from *Streptococcus pneumonia* and then work on *Escherichia coli* had identified a number of genes that, when mutationally inactivated, cause hypermutable strains.[1,2] Three of these proteins are essential in detecting the mismatch and directing repair machinery to it – MutS, MutH and MutL (MutS is a homologue of HexA and MutL of HexB). MLH1 heterodimerizes with PMS2 to form MutL alpha, a component of the postreplicative DNA mismatch repair system (MMR). Defects in MLH1 are a cause of mismatch repair cancer syndrome (MMRCS) also known as Turcot syndrome or brain tumor-polyposis syndrome1 (BTPS1),[3] Muir-Torre syndrome (MuToS) also abbreviated MTS and susceptibility to endometrial cancer (ENDMC).[4] Poor efficacy of DNA polymerase enzyme or the DNA being exposed to ionizing radiations (gamma rays, X-rays, ultraviolet rays), highly reactive oxygen radicals and various chemicals in the environment also produces aberrations in the DNA. If the genetic information encoded in the DNA is to remain uncorrupted, these chemical changes must be corrected to avoid various mutations. The DNA repair ability of a cell is vital to the

Access this article online	
Quick Response Code:	Website: www.jnsbm.org
	DOI: 10.4103/0976-9668.101887

integrity of its genome and thus to its normal functioning and that of the organism.

Mismatch repair enzymes function to recognize these errors and correct them. After replication, these enzymes travel down the new DNA molecules and are able to identify mistakes by the "bulge" that results from a mismatched pair. When an error is discovered, the mismatch repair enzymes then activate other enzymes that complete the DNA repair. There are various disorders that occur due to the mutations in this mismatch repair proteins and affect genomic stability, which can result in microsatellite instability (MI).[5] MI is implicated in most human cancers and majority of hereditary nonpolyposis colorectal cancers (HNPCC) are attributed to defects in MLH1.[6,7] It is also evident that DNA damage and repair are essential processes to understand the mechanisms of cancer, ageing and various human genetic diseases.[8] Therefore there is a need to analyze these proteins and their roles in various disorders. Our approach involves diversified analysis of the structural, functional and evolutionary aspects of these proteins.

In our study we analyzed the MLH1 protein and other associated proteins. DNA repair is initiated by MutS alpha (MSH2-MSH6) or MutS beta (MSH2-MSH6) binding to a dsDNA mismatch, then MutL alpha is recruited to the heteroduplex.[9] Assembly of the MutL-MutS-heteroduplex ternary complex in presence of RFC and PCNA is sufficient to activate endonuclease activity of PMS2.[10,11] It introduces single-strand breaks near the mismatch and thus generates new entry points for the exonuclease EXO1 to degrade the strand containing the mismatch. DNA methylation would prevent cleavage and therefore assure that only the newly mutated DNA strand is going to be corrected. MutL alpha (MLH1-PMS2) interacts physically with the clamp loader subunits of DNA polymerase III suggesting that it may play a role to recruit the DNA polymerase III to the site of the MMR. Also implicated in DNA damage signaling, a process which induces cell cycle arrest and can lead to apoptosis in case of major DNA damages. The MLH1 protein which is a mismatch repair protein present in many species had a common signature motif - GFRGE[AG]L.

The ability to recognize and repair damaged DNA is common to all forms of life, and numerous DNA repair pathways have evolved to repair almost all possible DNA lesions. The comparative and functional genome study of the organisms helps us to identify conserved regions and various related disorders.[12] There is a strong relationship between DNA repair pathways and human genetic disorders as these disorders represents defects in several associated genes e.g. in case of cancer and multi-system defects specifically in the immune and neurological systems.[13] The various protein–protein interactions which are involved in many complex networks and pathways are essential for understanding the metabolic and cellular processes and can further serve as novel targets for therapeutic interventions. There is a growing interest in understanding haplotype structures in the human genome using identified genetic markers as haplotype structures may provide critical information on human evolutionary history and the identification of genetic variants underlying various human traits.[14] Therefore, a DNA mismatch repair protein i.e. MLH1 which is involved in various disorders has been extensively analyzed in this study.

MATERIALS AND METHODS

Various *in silico* approaches and computational tools have been applied for the biological analysis of MLH1 protein. First, the protein sequence of MLH1 protein in humans was retrieved from NCBI which was cross referenced from Uniprot and Swissprot databases. The MLH1 protein sequences from various other organisms like *Saccharomyces cerevisiae*, *Rattus norvegicus, Bos taurus, Mus musculus*, etc. were also retrieved from NCBI and then these sequences were aligned together using Multiple Sequence Alignment tools like MAFFT[15] and MUSCLE.[16] Conserved motifs in these sequences were compared and confirmed through PROSITE database.[17]

A phylogenetic tree providing evolutionary relatedness of sequences was also obtained through Treefinder with GTR-GI model and 10,000 replicates[18] and the Phylogenetic Web Repeater (POWER).[19] Various protein–protein interactions with MLH1 were obtained from STRING,[20] BIND,[21] IntAct,[22] and MINT[23] PPI databases. The MLH1 sub-cellular localization was obtained from various tools like PSORT,[24] LOCATE,[25] BaCelLo[26] and MultiLoc,[27] which was found related to various disease pathways in KEGG database[28] [Table 1]. The protein structure of MLH1 was also found and downloaded from Protein Data Bank (PDB) and the active site residues were obtained from QSITE Finder.[29]

Linkage Disequilibrium (LD) is used in the study of population genetics for the non-random association of alleles at two or more loci.[30] Various measures have been proposed for characterizing the statistical association between alleles at different loci. Most common measures are D' and r^2 and both range between 0 and 1. D' is a measure of LD between two genetic markers. D' = 1 (complete LD) indicates that two SNPs have not been separated by recombination, while D' <1 (incomplete LD) indicates that the ancestral LD was disrupted during the history of the population. Only D' value near one is a reliable measure of LD extent.

Table 1: Various disease pathways through KEGG

Entry ID	Name of pathway	Important selective description	Pathway class
hsa05210	Colorectal cancer – Homo sapiens	One of the major mechanisms of genomic instability in sporadic CRC progression is microsatellite instability (MSI), results from inactivation of the DNA mismatch repair genes MLH1 and/or MSH2 by hypermethylation of their promoter, and secondary mutation of genes with coding microsatellites.	Human diseases; Cancers
hsa03430	Mismatch repair – Homo sapiens	In *E. coli*, the mismatch-activated MutS-MutL-ATP complex licenses MutH to incise the nearest unmethylated GATC sequence. Several human MMR proteins have been identified based on their homology to *E. coli* MMR proteins. These include human homologues of MutS and MutL. Although *E. coli* MutS and MutL proteins are homodimers, human MutS and MutL homologs are heterodimers.	Genetic information processing; Replication and repair
hsa03460	Fanconi anemia pathway – Homo sapiens	The Fanconi anemia pathway is required for the efficient repair of damaged DNA, especially inter-strand cross-links (ICLs). DNA ICL is directly recognized by FANCM and associated proteins that recruit the FA core complex. The FA core complex monoubiquitinates FANCD2 and FANCI. The monoubiquitinated FANCD2/FANCI becomes an active form and interacts with a series of DNA repair proteins and facilitates downstream repair pathways.	Genetic information processing; Replication and repair
hsa05200	Pathways in cancer – Homo sapiens	Implications of MSI, results from inactivation of the DNA mismatch repair genes MLH1, MSH2 and other homologues.	Human diseases; cancers
hsa05213	Endometrial cancer – Homo sapiens	Two types of endometrial carcinoma are distinguished with respect to biology and clinical course. The morphologic differences between Type 1 and Type 2 cancers are mirrored in their molecular genetic profile with type I showing defects in DNA-mismatch repair and mutations in PTEN, K-ras, and beta-catenin.	Human diseases; cancers

Figure 1: MAFFT-generated Partial MSA of MLH1 of various species

r^2 is also a measure of LD between two genetic markers. $r^2 = 1$ (Perfect LD) for SNPs that have been separated by recombination or have the same allele frequencies. We have here applied haplotype block and haplotype tagger analysis to reveal the information regarding LD.[31] The haplotype analysis was performed using Haploview.

RESULTS

In multiple sequence alignment (MSA) performed by MAFFT and MUSCLE [Figure 1], a conserved signature motif for mismatch repair proteins GFRGE[AG]L, is shown within the rectangle. This proves that the protein sequence involved in mismatch repair in different organisms have been found to be evolutionary related as there is a common conserved motif in MLH1 protein of these species, which is a DNA mismatch repair protein's MutL/HexB/PMS1 signature motif. From the PSORT subcellular localization tool, the MLH1 protein was found to be nuclear which was also confirmed by other available servers. A phylogenetic tree was reconstructed using Treefinder with consensus analysis for 10000 replicates on GTR-GI model with optimum values which shows

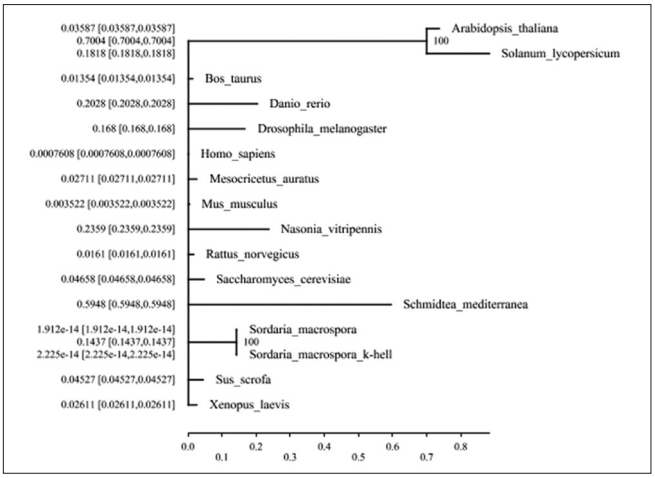

Figure 2: Phylogenetic tree (Consensus) reconstructed from Tree finder

the evolutionary relationship of sequences in this study [Figure 2]. Tree is in harmony with available phylogenies of involved studies but with different marker data. *Arabidopsis* with *Solanum* is an interesting aspect of the tree as this pair is most distant and justifies its presence with these two species as separate and far clade from rest of the species evolutionarily. Further longer branches of *Drosophila* (Insecta) and *Schmidtea* (Platyhelminthes) confirms their position between plants and higher organisms. Positions of fungus (Ascomycetes), zebra fish (Cyprinidae), and *Nasonia* (Insecta) with longer branches than rodents and mammals gave perfect shape to this phylogenetic tree. Tree is in agreement with the available standard phylogenies but distinction of two rooted separated blocks is a unique feature among species in this study.

The pattern of the tree generated by POWER and other

phylogenetic tree generating programs was almost similar with respect to phylogenetic trends of all the species in this study. All the groups and nodes are in agreement with the repetition of particular species with a score of more than 80 except one group where 3 species- Saccharomyces cerevisiae, Sordaria macrospora, Sordaria macrospora k-hell are there while the score of this group is also significant (60–80) as shown in Figure 3.

The MLH1 protein is known to interact with a number of proteins which are involved in DNA repair pathways. From the STRING database, the MLH1 protein is found to be interacting with a number of proteins like msh2, pms2, msh3, exo1, msh6, etc., which has experimental, text-mining, gene fusion, neighborhood, co-expression and other evidences. Some of the important interactions in STRING database have been found similar to the interactions in the BIND database as shown in Figure 4 and Table 2. When these interactions were observed in other databases like IntAct and MINT, certain new interacting proteins were found which are represented in Tables 3 and 4, respectively. Genecards gave the information regarding 102 proteins interacting with MLH1 [Table 5]. Therefore, on comparing all these databases, certain interactions were found common and will be of interest to researchers.

When MLH1 protein was searched in KEGG (Kyoto Encyclopedia for Genes and Genomes) Pathway database, various biochemical pathways had shown the vital role of MLH1 protein in their processes. Various diseases are closely associated with MLH1 protein as this protein is found in the pathways causing many cancers like colorectal cancer, endometrial cancer, etc., and in mismatch repair pathway.[32] Various active site residues were discovered from the MLH1 protein structure so that the putative site should be known in advance where the ligand could probably bind the protein. As we have already seen that this protein is involved in a number of diseases therefore there is a need to

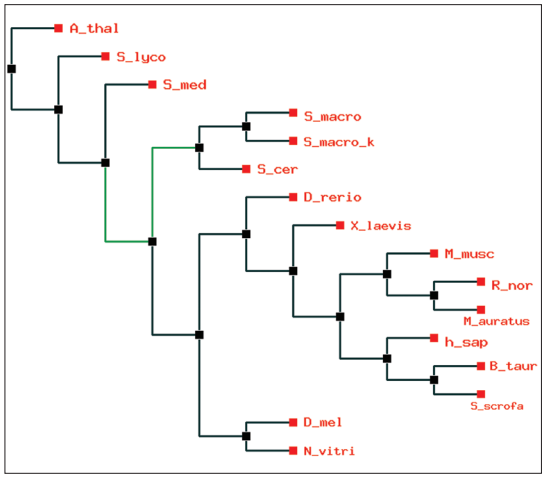

Figure 3: Phylogenetic tree reconstructed from POWER

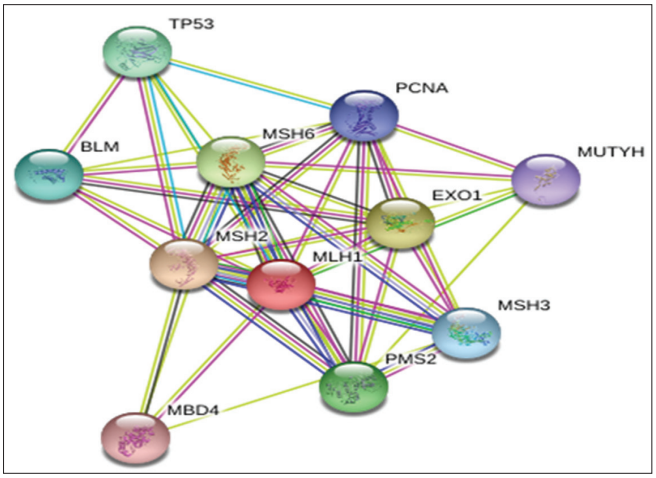

Figure 4: Protein–protein interactions from STRING database

Table 2: Interactions from BIND database

Identifier	Mol A	Mol B	Experimental evidence	Taxonomy
Interaction 12769	**MLH1**	**Blm**	**Immunostaining, Two hybrid test, Other, Immunoprecipitation**	**Homo sapiens**
Interaction 50047	MLH1	Myc	Two hybrid test, Immunoprecipitation, Affinity chromatography	Homo sapiens
Interaction 196577	E2F1	MLH1 promoter	Cross-linking	Homo sapiens
Interaction 12795	MLH1	MED1	Two hybrid test, Immunoprecipitation	Homo sapiens
Interaction 49960	**MLH1**	**Pms1**	**Other**	**Saccharomyces cerevisiae**
Interaction 194854	E2F4	MLH1 promoter	Cross-linking	Homo sapiens
Interaction 12796	MLH1	MSH4	Immunoprecipitation, Affinity chromatography	Homo sapiens
Interaction 316833	H3	MLH1 promoter	Cross linking	Homo sapiens
Interaction 146871	**MLH1**	**Pms2**	**Two hybrid test**	**Drosophila melanogaster**
Interaction 114883	DNA Mismatch Repair Protein Muts	DNA Mismatch Repair Protein Muts	Three-dimensional structure	Thermus aquaticus
Interaction 194397	p130	MLH1 promoter	Cross-linking	Homo sapiens

Table 3: Some important interactions from IntAct database

Name molecule A	Links molecule A	Name molecule B	Links molecule B	Interaction detection method	Interaction AC
|MLH1	P38920 EBI-11003	MLH2	Q07980 EBI-33369	Inferred by curator, Co-immunoprecipitation, Two hybrid	EBI-2342182 EBI-969865 EBI-969909
MLH1	**P40692 EBI-744248**	**AP2B1**	**P63010 EBI-432924**	**Two hybrid pooling approach**	**EBI-753892**
MLH1	**P40692 EBI-744248**	**FRMD6**	**Q96NE9 EBI-741729**	**Two hybrid pooling approach**	**EBI-757414**
MLH1	Q8T0N1 EBI-499087	PMS2	O76417 EBI-498159	Two hybrid fragment pooling approach	EBI-504370
MLH1	P38920 EBI-11003	MLH3	Q12083 EBI-31634	Inferred by curator, Co-immunoprecipitation, Two hybrid	EBI-2342193 EBI-969873 EBI-969919
MLH1	**P38920 EBI-11003**	**PMS1**	**P14242 EBI-13561**	**Inferred by curator, Co-immunoprecipitation, Two hybrid**	**EBI-2342163 EBI-969882 EBI-969929**
MLH1	P40692 EBI-744248	ALDOA	P04075 EBI-709613	Two hybrid	EBI-2938911 imex:IM-15124-56
MLH1	P40692 EBI-744248	FLNC	Q14315 EBI-489954	Two hybrid	EBI-2939310 imex:IM-15124-2

Table 4: Interactions from MINT database

Protein	Evidences	Score	Associations	Complex	HT
BRIP1 Homo sapiens (Q9BX63)	11	0.91	10	5	
AP2B1 Homo sapiens (P63010)	**1**	**0.28**	**1**		**1**
BRCA1 Homo sapiens (P38398)	**1**	**0.28**	**1**	**1**	
FASTKD5 Homo sapiens (Q7L8L6)	1	0.28	1		1
FRMD6 Homo sapiens (Q96NE9)	**1**	**0.28**	**1**		**1**
PMS2 Homo sapiens (P54278)	**1**	**0.28**	**1**	**1**	
SNW1 Homo sapiens (Q13573)	1	0.28	1	1	
TRIM29 Homo sapiens (Q14134)	1	0.28	1		1
ZC3H11A Homo sapiens (O75152)	1	0.28	1		1

Table 5: Interacting proteins for MLH1 In genecards

Genecard	External ID(s)	Interaction details
BRCA1	**P38398**	**STRING (score = 0.984) MINT-5115348 MINT-5115319 MINT-5115375 MINT-5115552**
PMS2	**P54278**	**STRING (score = 0.999) EBI-744248, EBI-1162561 MINT-5115348 MINT-5115404 MINT-5115319 MINT-5115375 EBI-744248, EBI-1162561 MINT-5115348 MINT-5115404 MINT-5115319 MINT-5115375**
SOCS1	ENSP00000329418	STRING (score = 0.74)
MSH2	**ENSP00000233146**	**STRING (score = 0.999)**
APBA2	ENSP00000219865	STRING (score = 0.75)
MSH6	**ENSP00000234420**	**STRING (score = 0.994)**
PCNA	**ENSP00000368438**	**STRING (score = 0.993)**
RAD52	ENSP00000351284	STRING (score = 0.787)
BLM	**ENSP00000347232**	**STRING (score = 0.99)**
MUTYH	**ENSP00000352239**	**STRING (score = 0.989)**
MBD4	**ENSP00000249910**	**STRING (score = 0.987)**
PMS1	**ENSP00000343888**	**STRING (score = 0.743)**

All the protein shown in bold in Tables 2-5, are the most common protein–protein interactions found in multiple PPI databases

analyze this protein in detail and the pockets identified [Figure 5] where the drug could bind would help in designing new inhibitors for the protein. This kind of analysis can provide an insight for the therapeutic applications. All of the protein atoms close to a probe-cluster defining various sites are shown in Table 6.

According to some recent studies it has been found that chromosomes are structured in a way that each chromosome can be divided into many blocks named haplotypes.[33] Knowledge of local linkage disequilibrium (LD) and common haplotype patterns in disease association has potential to make them comprehensive and

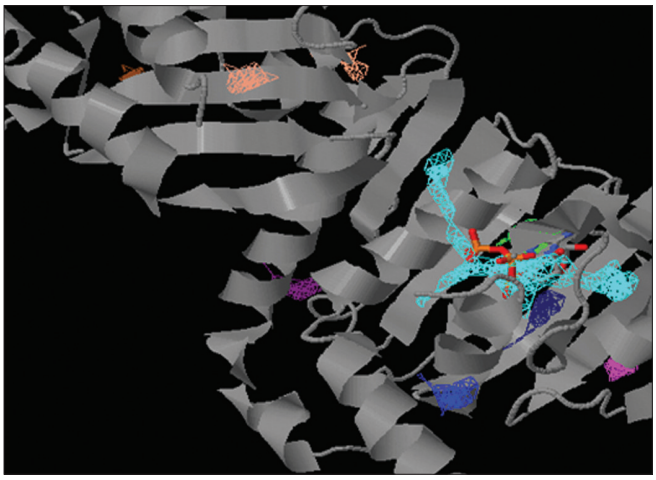

Figure 5: MLH1 protein with colored active sites

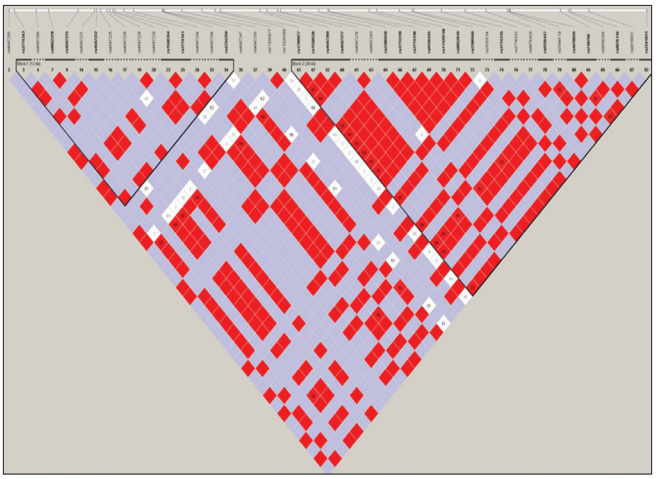

Figure 6: LD plot generated from haploview

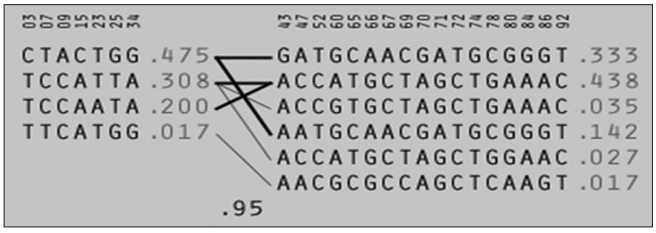

Figure 7: Haplotypes from haploview

efficient.[34] Haplotype tagging refers to the methods of selecting minimal number of SNPs that uniquely identify common haplotypes (>5% in frequency). Principal use of tagging is to select a 'good' subset of SNPs to be typed in all the studied individuals. We performed haplotype analysis on HapMap genotype data of SNPs genotyped in population CEU on chromosome 3. LD plot was generated and in haploblock diagram [Figure 6], two distinct blocks have been identified that are the alternative blocks within same loci on LD plot, and a strong correlation between blocks indicates independent site of action which is being proposed by this analysis and there is involvement of 7 markers in first block while 17 markers in second block with significant statistical support. Overall correlation of 0.95 has been found among all interactions of genotyped SNPs which is significant [Figure 7].

Table 6: QSITE FINDER predicted active site residues (selected)

Atom number	Atom type	Residue name	Chain identifier	Residue number
254	CA	GLU	A	37
255	C	GLU	A	37
256	O	GLU	A	37
257	CB	GLU	A	37
657	CA	GLY	A	98
658	C	GLY	A	98
660	N	PHE	A	99
662	C	PHE	A	99
663	O	PHE	A	99
671	N	ARG	A	100
672	CA	ARG	A	100
673	C	ARG	A	100
674	O	ARG	A	100
675	CB	ARG	A	100
677	CD	ARG	A	100
678	NE	ARG	A	100
679	CZ	ARG	A	100
680	NH1	ARG	A	100
681	NH2	ARG	A	100
682	N	GLY	A	101
683	CA	GLY	A	101
684	C	GLY	A	101
685	O	GLY	A	101
686	N	GLU	A	102
687	CA	GLU	A	102
688	C	GLU	A	102
689	O	GLU	A	102
690	CB	GLU	A	102
691	CG	GLU	A	102
692	CD	GLU	A	102
693	OE1	GLU	A	102
694	OE2	GLU	A	102
695	N	ALA	A	103
696	CA	ALA	A	103
697	C	ALA	A	103
699	CB	ALA	A	103
700	N	LEU	A	104
701	CA	LEU	A	104
702	C	LEU	A	104
704	CB	LEU	A	104
705	CG	LEU	A	104
706	CD1	LEU	A	104
707	CD2	LEU	A	104

DISCUSSION

From our analysis it can be concluded that a system's biology approach is essential for the interaction of genes/proteins/networks for understanding of the cellular processes, and there is a need to perform detailed analysis on repair pathways and associated human genetic disorders. The protein sequences involved in mismatch repair in different organisms have been found

to be evolutionary related as there is a common motif GFRGE[AG]L found in MLH1 protein of these species. Followed by the multiple sequence analysis using MAFFT and MUSCLE servers, the same pattern was found conserved among all species in this study. Phylogenetic tree generated based on MSA is also in agreement with standard phylogeny available for various biomarkers. Several other related proteins have also been identified through protein–protein interactions. All associated proteins are either mismatch repair proteins or associated with MLH1 in various pathways. Pathways information was also confirmed through MMR and other pathways in KEGG. Further studies from QSite Finder showed that the active site of MLH1 protein also involves these residues and this conserved pattern is involved in ligand–protein interactions as confirmed through a complex structure of MLH1. Information generated will definitely be an aid for further research and based on conserved residues of active sites and various ligand interaction cavities, new inhibitors can be designed.

Marker information is generated from sequence to structure level with conserved signature motif and active site residue within structural pockets, respectively. Besides that, evolutionary information has also been generated which suggests the selection of a specific and suitable molecular evolutionary model of substitution for MLH1 protein sequences among various organisms. Haplotype analysis revealed 24 (17+7) new alleles with significant statistical scores and confirmed the association of these alleles with various disorders. Two independent sites of action (two distinct but related blocks) have been identified for the same allele, which might be helpful in mapping various markers on genomic data. Overall this study provides a new direction towards repair proteins and their myriad analysis.

REFERENCES

1. Priebe SD, Hadi SM, Greenberg B, Lacks SA. Nucleotide sequence of the hexA4 gene for DNA mismatch repair in Streptococcuspneumoniae and homology of hexA to mutS of Escherichia coli and Salmonella typhimurium. J Bacteriol 1988;170:190-6.
2. Prudhomme M, Martin B, Mejean V, Claverys J. Nucleotide sequence of the Streptococcus pneumoniae hexB mismatch repair gene: Homology of HexB to MutL of Salmonella typhimurium and to PMS1 of Saccharomyces cerevisiae. J Bacteriol 1989;171:5332-8.
3. Hamilton SR, Liu B, Parsons RE, Papadopoulos N, Jen J, Powell SM, *et al*. The molecular basis of Turcot's syndrome. N Engl J Med 1995;332:839-47.
4. Shin BY, Chen H, Rozek LS, Paxton L, Peel DJ, Anton-Culver H, *et al*. Low allele frequency of MLH1 D132H in American colorectal and endometrial cancer patients. Dis Colon Rectum 2005;48:1723-7.
5. Kobayashi K, Matsushima M, Koi S, Saito H, Sagae S, Kudo R, *et al*. Mutational analysis of mismatch repair genes, hMLH1 and hMSH2, in sporadic endometrial carcinomas with microsatellite instability. Jpn J Cancer Res 1996;87:141-5.
6. Kruger S, Plaschke J, Jeske B, Gorgens H, Pistorius SR, Bier A, *et al*. Identification of six novel MSH2 and MLH1 germline mutations in HNPCC. Hum Mutat 2003;21:445-46.
7. Vasen HF, Moslein G, Alonso A, Bernstein I, Bertario L, Blanco I, *et al*. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). J Med Genet 2007;44:353-62.
8. Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res 2008;18:85-98.
9. Jiricny J. MutLalpha: At the cutting edge of mismatch repair. Cell 2006;126:239-41.
10. Flores-Rozas H, Clark D, Kolodner RD. Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. Nat Genet 2000;26:375-8.
11. Iyer RR, Pohlhaus TJ, Chen S, Hura GL, Dzantiev L, Beese LS, *et al*. The MutSalpha-proliferating cell nuclear antigen interaction in human DNA mismatch repair. J Biol Chem 2008;283:13310-9.
12. Manolio TA. Genomewide Association Studies and Assessment of the Risk of Disease. N Engl J Med 2010;363:166-76.
13. Suter CM, Martin IK, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 2004;36:497-501.
14. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. Annu Rev Med 2005;56:303-20.
15. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acid Res 2002;30:3059-66.
16. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792-7.
17. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, *et al*. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 2010;38:161-6.
18. Jobb G, von Haeseler A, Strimmer K. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 2004;4:18.
19. Lin CY, Lin FK, Lin CH, Lai LW, Hsu HJ, Chin CH, *et al*. POWER: PhylOgenetic WEb Repeater—an integrated and user-optimized framework for biomolecular phylogenetic analysis. Nucleic Acids Res 2005;33:W553-6.
20. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, *et al*. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39:D561-8.
21. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson T, Hogue CW. BIND-The Biomolecular Interaction Network Database. Nucleic Acids Res 2001;29:242-5.
22. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, *et al*. The IntAct molecular interaction database in 2010. Nucleic Acids Res 2010;38:D525-31.
23. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L. MINT: The Molecular INTeraction database. Nucleic Acids Res. 2007;35:D572-4.
24. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ. WoLF PSORT: Protein localization predictor. Nucleic Acids Res 2007;35:W585-7.
25. Fink JL, Aturaliya RN, Davis MJ, Zhang F, Hanson K, Teasdale MS, *et al*. LOCATE: A mouse protein subcellular localization database. Nucleic Acids Res 2006;34:D213-7.
26. Pierleoni A, Martelli PL, Fariselli P, Casadio R. BaCelLo: A Balanced subCellular Localization predictor. Bioinformatics 2007;22:e408-16.
27. Hoeglund A, Doennes P, Blum T, Adolph HW, Kohlbacher O. MultiLoc: Prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs, and amino acid composition. Bioinformatics 2006;22:1158-65.
28. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl Acids Res 2000;28:27-30.
29. Laurie AT, Jackson RM. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 2005;21:1908-16.
30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 2005;

21:263-5.
31. Purcell S, Daly MJ, Sham PC. WHAP: Haplotype-based association analysis. Bioinformatics 2007;23:255-6.
32. Vilkki S, Tsao JL, Loukola A, Poyhonen M, Vierimaa O, Herva R, *et al*. Extensive somatic microsatellite mutations in normal human tissue. Cancer Res 2001;61:4541-4.
33. Sub Y, Vijg J. SNP discovering in associating genetic variation with human disease phenotypes. Mutat Res 2005;573:41-53.
34. The International HapMap Consortium. The International HapMap Project. Nature 2003;426:789-96.

How to cite this article: Sehgal M, Singh TR. Identification and analysis of biomarkers for mismatch repair proteins: A bioinformatic approach. J Nat Sc Biol Med 2012;3:139-46.

Source of Support: Nil. **Conflict of Interest:** None declared.