

RESEARCH

Open Access



A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human *MBL2* gene

Namarta Kalia¹, Aarti Sharma², Manpreet Kaur², Sukhdev Singh Kamboj¹ and Jatinder Singh^{1*}

Abstract

Mannose binding lectin (MBL) is a liver derived protein which plays an important role in innate immunity. Mannose binding lectin gene 2 (*MBL2*) polymorphisms are reported to be associated with various diseases. In spite of being exhaustively studied molecule, no attempt has been made till date to comprehensively and systematically analyze the SNPs of *MBL2* gene. The present study was carried out to identify and prioritize the SNPs of *MBL2* gene for further genotyping and functional studies. To predict the possible impact of SNPs on MBL structure and function SNP data obtained from dbSNP database were analyzed using various bioinformatics tools. Out of total 661 SNPs, only 37 validated SNPs having minor allele frequency ≥ 0.10 were considered for the present study. These 37 SNPs includes one in 3' near gene, nine in 3' UTR, one non-synonymous SNP (nsSNP), thirteen intronic SNPs and thirteen in 5' near gene. From these 37 SNPs, 11 non-coding SNPs were identified to be of functional significance and evolutionary conserved. Out of these, 4 SNPs from 3' UTR were found to play role in miRNA binding, 7 SNPs from 5' near and intronic region were predicted to involve in transcription factor binding and expression of *MBL2* gene. One nsSNP Gly54Asp (rs1800450) was found to be deleterious and damaging by both SIFT and Polyphen-2 servers and thus affecting *MBL2* protein stability and expression. Protein structural analysis with this amino acid variant was performed by using I-TASSER, RAMPAGE, Swiss-PdbViewer, Chimera and I-mutant. Information regarding solvent accessibility, molecular dynamics and energy minimization calculations showed that this variant causes clashes with neighboring amino acids residues that must interfere in the normal triple helix formation of trimeric subunit and further with the normal assembly of MBL oligomeric form, hence decrease in stability. Thus, findings of the present study indicated 12 SNPs of *MBL2* gene to be functionally important. Exploration of these variants may provide novel remedial markers for various diseases.

Keywords: Candidate gene studies, Computational analysis, Mannose binding lectin gene 2, Non-coding SNPs, Non-synonymous SNPs, Single nucleotide polymorphism (SNP)

Background

Mannose binding lectin (MBL) is a liver derived acute phase protein. It binds to carbohydrates on the surface of mannose-rich pathogens and mediates clearing by phagocytosis or complement activation (Nepomuceno et al. 1997; Selander et al. 2006). Mice have two homologous genes of MBL i.e. *MBL1* and *MBL2*, but in

human the first of them is pseudogene leaving one functional gene *MBL2* (Guo et al. 1998). The *MBL2* gene is composed of 7461 bases and lies between the regions 52765380 to 52772841 bp of chromosome no. 10 (NCBI reference sequence number NC_000010.11). The gene contains four exons and three introns. It encodes a 248 amino acid residue transmembrane protein, MBL (NCBI accession number XP_011538118.1) which is encoded via a 3570 bp long mRNA (NCBI accession number NM_000242.21). It belongs to a family of proteins called collectins, which consists of collagenous region and a carbohydrate recognition domain (Taylor et al. 1989).

*Correspondence: jatinderarora2009@gmail.com

¹ Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar 143005, India

Full list of author information is available at the end of the article

MBL consists of multimers of an identical polypeptide chain of 32 kDa.

A single-nucleotide polymorphism (SNP) is the most common type of genetic variation. Several studies have shown *MBL2* SNPs of promoter and exonic region regulate the *MBL2* serum levels in different autoimmune diseases and infectious diseases, including HIV infection, leishmaniasis, leprosy, malaria, schistosomiasis, tuberculosis and trypanosomiasis (Madsen et al. 1995; Summerfield et al. 1995; Garred et al. 1997; Kelly et al. 2000; Klabunde et al. 2000; Jack and Turner 2003; IP et al. 2005; Garred et al. 2006; Alonso et al. 2007). SNPs of *MBL2* gene cover both coding and non-coding regions. However, not all the coding region elements are functionally important. Only the non-synonymous SNPs (nsSNPs), also called as missense variants are particularly important as they result to changes in the translated amino acid residue sequence. nsSNPs may affect the protein function by reducing protein solubility or by destabilizing protein structure (Chasman and Adams 2001). Moreover, analyses on conserved non-coding region have shown that non-coding DNA is involved in biological functions (Alexander et al. 2010). These non-coding elements can have various regulatory functions within the genome, such as interacting with transcription factors (TFs), miRNA, creating splice sites and acting as exonic splicing enhancers (ESEs) (Birney et al. 2007). Despite their important regulatory role, much less effort has been invested in the functional analysis of non-coding SNPs for candidate gene studies as compared to the coding regions SNPs.

There are several publically available databases for SNPs, such as dbSNP, GWAS Central, SwissVar etc. dbSNP is the most extensive among all the databases (Sherry et al. 2001; Bhagwat 2010). It contained a total of 661 SNPs in the *MBL2* gene of *Homo sapiens* as of October, 2015. To date the functional significance has not been established for the majority of them. In the absence of any experimental information on their functional effects, the potential functional consequences of a SNP can be predicted using various bioinformatics tools (Bhatti et al. 2006; Johnson 2009; Li and Wei 2015). These tools predict the functional effects of SNPs at five main levels: splicing, transcriptional, translational, post-translational and protein stability. The majority of current bioinformatics tools examine the functional effects of SNPs only with respect to a single biological function. However, the others provide a comprehensive assessment of SNP function based on different algorithms, data and resources (Bhatti et al. 2006; Johnson 2009; Li and Wei 2015).

All the SNPs present in the *MBL2* gene were analysed using various composite and singleton tools to

verify their putative functional effects. The SNPs that were identified as having functional effects were then prioritized on the basis of number of criteria; i.e. the significance of the function identified, presence within an evolutionary conserved region, validation status of the SNP, and the minor allele frequency of the SNP. Thus, the present study involves filtering through a list of SNPs to identify causal variants. The study was further explored to view the effect of nsSNP on the stability of MBL protein. To the best of our knowledge, this is the first comprehensive computational study undertaken for in silico analysis of nsSNPs as well as regulatory SNPs in *MBL2* gene.

Methods

The SNPs and their related protein sequence for the *MBL2* gene were obtained from dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) and were subjected to various computational analyses. The strategy followed to select SNPs having structural and functional impact is shown in Fig. 1.

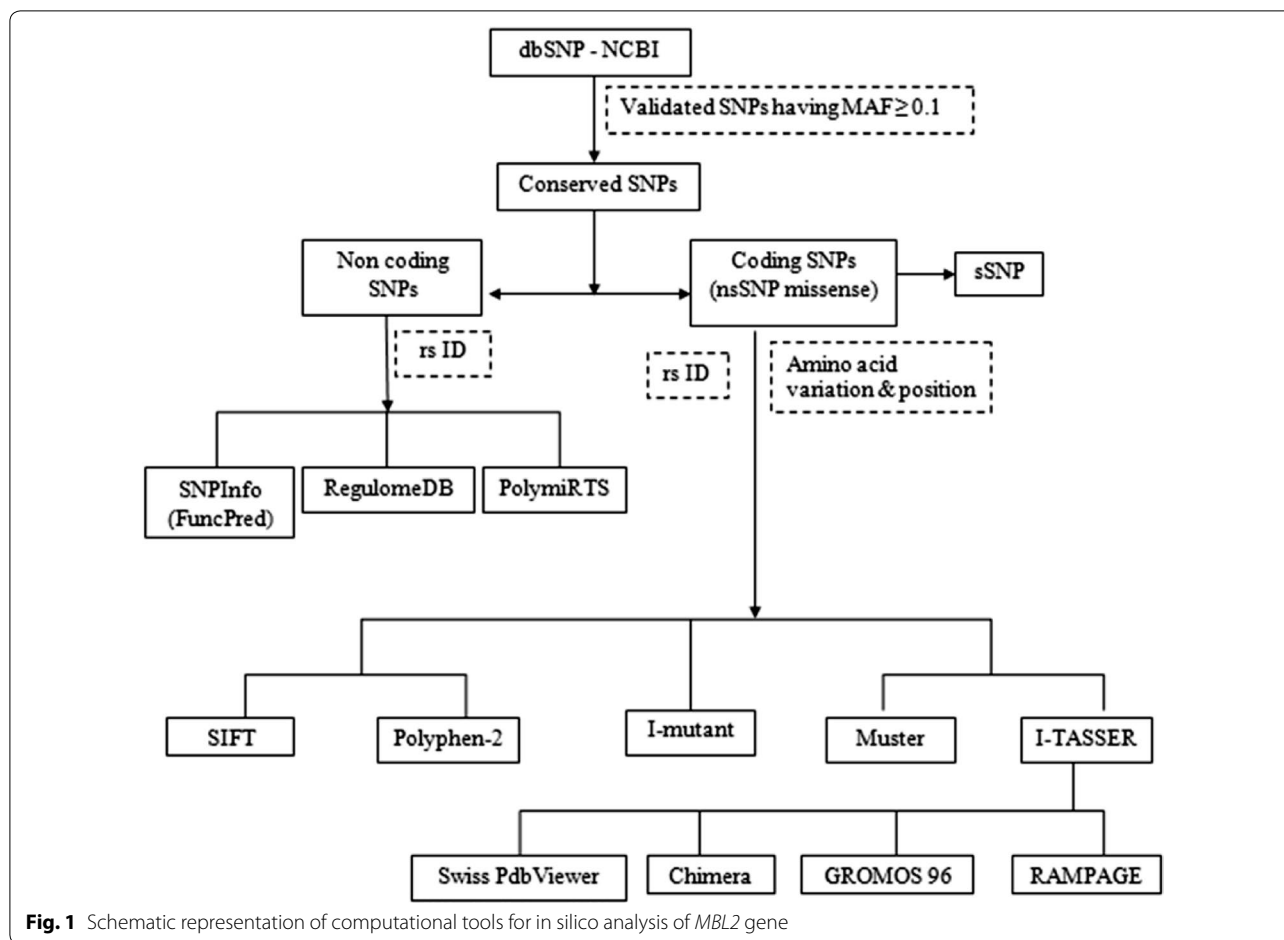
Identifying SNPs located in evolutionary conserved regions in the gene

SNPs located in evolutionary conserved regions (cSNPs) were identified using the *Ensembl* Genome browser release 48 (<http://www.ensembl.org/>). *MBL2* gene from seventeen eutherian mammals was selected for the comparative analysis with the human *MBL2* gene. These includes *Homo sapiens* (Human), *Pan troglodytes* (Chimpanzee), *Gorilla gorilla gorilla* (Gorilla), *Pongo abelii* (Orangutan), *Chlorocebus sabaeus* (Vervet-AGM), *Macaca mulatta* (Macaque), *Papio Anubis* (Olivebaboon), *Bos Taurus* (Cow), *Ovis aries* (Sheep), *Sus scrofa* (Pig), *Equus caballus* (Horse) *Felis catus* (Cat), *Canis lupus familiaris* (Dog), *Callithrix jacchus* (Marmoset), *Mus musculus* (Mouse), *Oryctolagus cuniculus* (Rabbit) and *Rattus norvegicus* (Rat). For a detailed comparison, base pair view was selected from the database and the SNPs were verified manually.

Non-coding SNP functional analysis

To identify the effect of SNPs in non-coding regions, tools predicting potential functional effect of SNPs at Transcription factor binding sites (TFBS), Intron/exon border consensus sequences (splice sites), Exonic splicing enhancers (ESEs), and miRNA binding were used.

SNPinfo (*FuncPred*) and *RegulomeDB* provide a collection of functional information using variety of tools and resources. The SNPs functionality was primarily determined by *SNPinfo* (*FuncPred*) (<http://snpinfonihs.nih.gov/snpinf/snpfunc.htm>) web server in Asian population (Xu and Taylor 2009). This webserver helps in



selecting SNPs for genetic association studies and consists of 3 pipelines for SNP selection and a composite tool for SNP function prediction named 'FuncPred'. Which is composite of Polyphen, SNP3D, MATCH, TRANSFAC 12.1, rescue ESE, ESEfinder, FAS-ESS, miRanda and miRBase. A user can query results for all the SNPs in a gene or chromosomal region, or for a list of input SNPs. For the present study, a list of SNP rsIDs was uploaded for batch analysis with default settings. The output was a list of SNPs with possible functional effect.

To complement SNP prioritization, SNPs were further analysed by RegulomeDB (<http://regulomedb.org/>). It is an online composite database and prediction tool to annotate as well as prioritize potential regulatory variants from human genome (Boyle et al. 2012). The database includes up-to date high quality datasets from Encyclopedia of DNA Elements transcription factor, chromatin immunoprecipitation sequencing (ChIP-seq), histone ChIP-seq, Formaldehyde-Assisted Isolation of Regulatory Elements, DNase I hypersensitive site data and other sources like large collection of Expression quantitative trait loci, dsQTL, ChIP-exo data to identify putative

regulatory variants. RegulomeDB classifies variants into six categories ranging from 1 to 6, where category 1 variants are 'likely to affect binding and linked to expression of a gene target', category 2 variants are 'likely to affect binding', Category 3 variants are 'less likely to affect binding', and Category 4, 5 and 6 variants have 'minimal binding evidence'. RegulomeDB also assigns a score of 7 for variants with no annotation data available. dbSNP rsIDs were used as input for the present study. A list of rsIDs allotted with different categories and also with no data available was obtained.

PolymiRTS (v 3.0) is an integrated database that systematically identifies DNA polymorphisms in miRNAs and miRNA target sites (PolymiRTS). It also elucidates the potential links of SNPs to molecular, physiological, behavioral and disease phenotypes (Bhattacharya et al. 2014). This computational server was used in order to determine the effect of 3' UTR SNPs of *MBL2* gene in creating and abolishing miRNA target sites resulting in diverse functional consequences. Gene symbol was selected, *MBL2* was entered as search information and submitted to PolymiRTS (v3.0) (<http://compbio.uthsc.edu/>).

[edu/miRSNP/](#)). A list of all 3' UTR SNPs of *MBL2* gene, creating or destructing miRNA site was obtained. These 3' UTR SNPs were further looked for possible miRNA IDs for which they create and abolish site for.

Non-synonymous SNP functional analysis

Deleterious and damaging effect of nsSNP was predicted using web based tools SIFT (Sorting Intolerant from Tolerant) and Polyphen-2 (Polymorphism Phenotyping v2).

SIFT (<http://sift.bii.a-star.edu.sg/>) performs analysis based on different algorithms and it interprets the homologous sequences using the Swiss-Prot (version 51.3) and TrEMBL (version 34.3) (Kumar et al. 2009). Results were expressed as SIFT scores which were classified as damaging (0.00–0.05), potentially damaging (0.051–0.10), borderline (0.101–0.20), or tolerant (0.201–1.00). nsSNP rsID were uploaded to get the possible results.

Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>) was used to predict the possible impact of an amino acid substitution on both structure and function of protein by analysis of multiple sequence alignment and protein 3D structure (Adzhubei et al. 2013). Protein sequence, database ID/accession number, amino acid position and amino acid variant details are the input options for PolyPhen-2. Protein sequence with NCBI accession number XP_011538118.1 was input for the present study. Prediction outcomes could be classified as probably damaging, possibly damaging or benign according to the score ranging from (0–1). “Score” is the probability of the substitution being damaging; “sensitivity” and “specificity” correspond to prediction confidence. The predicted damaging effect is also indicated by a vertical black marker inside a color gradient bar, where green is benign and red is damaging.

Protein structure prediction

Protein blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) was used to find the proteins related to the *MBL2*. To obtain a list of the closest matches it performs a FASTA search against every macromolecular structure deposited in protein data bank (PDB) (<http://www.rcsb.org/pdb/home/>).

Muster v1.0 (<http://zhanglab.ccmb.med.umich.edu/MUSTER/>) provides the Z-score and complete full length models by using Modeller v8.2 (Wu and Zhang 2008). The corresponding template is considered good if the calculated Z-score is greater than 7.5.

I-Tasser (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) server provides accurate structural and functional predictions using state-of-art algorithms (Roy et al. 2010). It reports up to five models which correspond to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score (confidence

score), calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score typically lies in the range of –5 to 2 where a higher value signifies a model with a higher confidence and vice versa. TM-score (template modeling score) >0.5 highlights a model of correct topology and a TM-score <0.17 indicates a random similarity. It also predict solvent accessibility with values range from 0 (buried residue) to 9 (highly exposed residue). Protein sequence with NCBI accession number XP_011538118.1 was input for the present study. Structural evaluation was carried out using *RAMPAGE* ramachandran plot analysis (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell et al. 2003).

Modeling nsSNP on protein structure

Swiss-PdbViewer (v4.10) was used to generate the mutated models of the selected protein structure for the corresponding amino acid substitutions (Guex and Peitsch 1997). It replaces the native amino acid with the variant. The.pdb files were saved for the model. This server also uses Gromacs as the default force field for energy minimization calculations.

Chimera (<http://www.cgl.ucsf.edu/chimera>) is a extensible molecular modeling system used for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories and conformational analysis (Pettersen et al. 2004).

Protein stability prediction

I-Mutant 2.0 (<http://folding.uib.es/i-mutant/i-mutant2.0.html>), a neural network based tool, predicts the change in the stability of the protein upon mutation (Capriotti et al. 2005). This tool automatically predicts protein stability changes upon single site mutations. Prediction can be performed using either protein structure or sequence. The FASTA sequence of protein with NCBI accession number XP_011538118.1 retrieved from NCBI is used as an input to predict the mutational effect on protein stability. Output obtained is in the form of protein stability change upon mutation and Gibbs-free energy change (DDG).

Results

The dbSNP database contains both validated and non-validated polymorphisms. In spite of this drawback, dbSNP was availed because it is the most extensive SNP database (Sherry et al. 2001; Bhagwat 2010). *MBL2* gene contains 661 SNPs in dbSNP database. Out of these 661 SNPs, only 37 validated SNPs having MAF \geq 0.10 were considered for the present study. Those SNPs which either have MAF < 0.10 or are not validated were

excluded for further analysis. These 37 SNPs includes one 3' near gene, nine 3' UTR, thirteen intronic SNPs, thirteen 5' near gene and one synonymous SNP (sSNP). One sSNP was excluded from the study leaving 36. However no non-synonymous SNPs were found to have MAF even ≥ 0.05 . So, SNPs falling in the list of pathogenic significance were also checked for validation status and $MAF \geq 0.1$. Only one SNP Gly54Asp (rs1800450) have $MAF = 0.1220$ was found and hence included in the study. Thus, our investigation accounted for 37 SNPs including 3' UTR, intronic, 5'/3' near gene and nsSNP. A map of the *MBL2* gene region highlighting the positions of each of these SNPs is shown in Fig. 2.

A list of 37 SNPs was submitted to the FuncPred program and results obtained are shown in Table 1. Total 24 SNPs were predicted to have effect on function, of which nine 3' UTR SNPs were found to affect miRNA binding site and three intronic and twelve 5' SNPs were found to be affecting transcription binding site. RegulomeDB database has divided 37 SNPs included in the present study into six broad categories (Category 1 to Category 6)—where twenty nine SNPs had annotation scores in between 1 and 6 (Table 2) and the remaining eight SNPs had no data annotation (not shown in table). Of the 29 SNPs, only one SNP had RegulomeDB score of 1f three SNPs were less likely to be functional (Category 3a) while,

25 SNPs had minimum functional evidence (Category 4, 5 and 6). The top ranked SNP rs7096206 had annotation for eQTL + TF binding/DNase peak and thus very likely to have regulatory functions.

In sequence alignment, except 6 species including *F. catus*, *C. lupus familiaris*, *C. jacchus*, *M. musculus*, *O. cuniculus* and *R. norvegicus*, all were found to show alignment with the human *MBL2* gene. Furthermore, all the 37 SNPs involved in present study were found to lie in conserved region, hence are called conserved SNPs (cSNPs) (Fig. 3). 3' UTR SNPs submitted online to PolymiRTS server showed four variants affecting miRNA target sites, hence proposed to have obvious consequences on protein truncation (Table 3). These four SNPs were also predicted by FuncPred for miRNA binding. The 3' UTR SNPs identified as having putative functional effects, by the four tools were then screened as shown in Table 4. Two SNPs predicted by all the tools and two SNPs i.e. rs2099903, rs2165813, not predicted by regulomeDB but by other 3 tools, were selected for further analysis because prediction by polymiRTS cannot be ignored.

Similarly, intronic and 5' near gene SNPs identified as having putative functional effects, by the three tools were screened as shown in Table 5. From the SNPs identified by all the three tools only 7 SNPs which were reported and have regulomeDB score ≤ 5 were selected and those

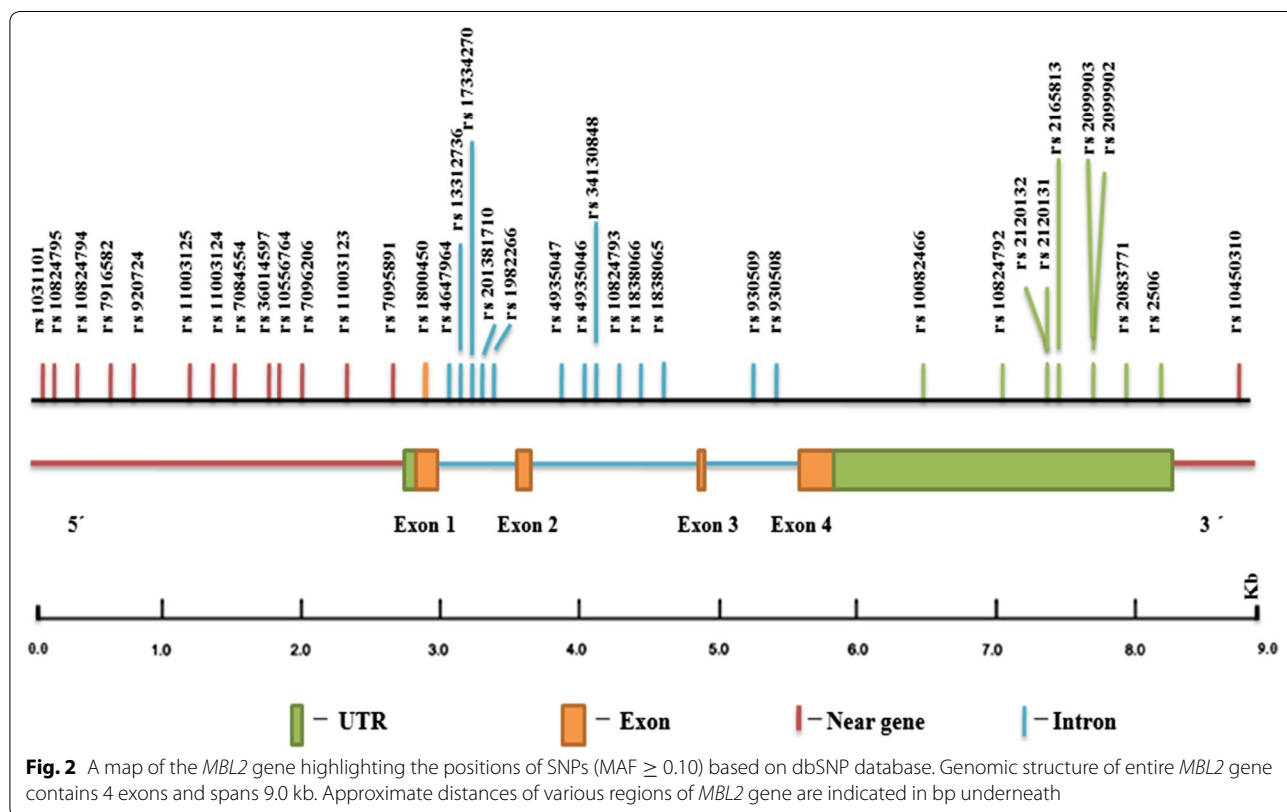


Table 1 List of non-coding SNPs predicted by FuncPred as having functional effect

Region	SNP ID	Chr. position	MAF	TFBS	miRNA
3' UTR	rs2506	52765749	0.3389	-	✓
3' UTR	rs2083771	52765918	0.3391	-	✓
3' UTR	rs2099902	52766089	0.3387	-	✓
3' UTR	rs2099903	52766097	0.2732	-	✓
3' UTR	rs2165813	52766224	0.2742	-	✓
3' UTR	rs2120131	52766258	0.2742	-	✓
3' UTR	rs2120132	52766280	0.2734	-	✓
3' UTR	rs10824792	52766446	0.4519	-	✓
3' UTR	rs10082466	52766862	0.2770	-	✓
Intron	rs1982266	52770876	0.4774	✓	-
Intron	rs17334270	52771083	0.2835	✓	-
Intron	rs13312736	52771221	0.2835	✓	-
5' near gene	rs7095891	52771701	0.2833	✓	-
5' near gene	rs11003123	52771774	0.2837	✓	-
5' near gene	rs7096206	52771925	0.1955	✓	-
5' near gene	rs36014597	52772040	0.2837	✓	-
5' near gene	rs7084554	52772053	0.2831	✓	-
5' near gene	rs11003124	52772131	0.2835	✓	-
5' near gene	rs11003125	52772254	0.3061	✓	-
5' near gene	rs920724	52773037	0.2953	✓	-
5' near gene	rs7916582	52773235	0.2841	✓	-
5' near gene	rs10824794	52773429	0.2841	✓	-
5' near gene	rs10824795	52773533	0.2841	✓	-
5' near gene	rs1031101	52773600	0.1222	✓	-

MAF minor allele frequency, TFBS transcription factor binding site, ✓ SNPs which affect function, - SNPs which does not affect function

which either were not reported (rs17334270, 13312736) or have regulomeDB score ≥ 6 (rs10824794, rs1031101, rs1982266) were not selected. Thus overall total 11 non-coding SNPs from 3' UTR and 5' near region of *MBL2* gene were filtered through a plethora of SNPs and identified as causal variants. Predictions of deleterious effect of rs1800450 (nsSNP) was performed by SIFT and Polyphen-2 softwares. It was predicted to be damaging by both the servers. The SIFT prediction was deleterious with a score of 0.006. This deleterious nsSNP was submitted to Polyphen-2 as query protein sequences in FASTA format with prediction outcomes to be probably damaging score closer to 1.

By using Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), the *MBL2* gene related protein structures were searched. Although different mutations in *MBL2* gene are reported, the native primary structure (with complete 248 amino acid sequence) of MBL monomeric protein is not yet available. Only partial protein structure, with PDB ID 1HUP was found (Sheriff et al. 1994). This partial structure of

Table 2 List of SNPs predicted by regulomeDB score

dbSNP ID	RDB score	Category	Description
rs7096206	1f	Likely to affect binding and linked to expression of a gene target	eQTL + TF binding/ DNase peak
rs930508	3a	Less likely to affect binding	TF binding + any motif + DNase peak
rs10556764			
rs7084554			
rs930507	4	Minimal binding evidence	TF binding + DNase peak
rs930509			
rs13312736			
rs7095891			
rs11003123			
rs36014597			
rs10450310	5	Minimal binding evidence	TF binding or DNase peak
rs10824792			
rs10082466			
rs35137523			
rs1838065			
rs1838066			
rs34130848			
rs4935046			
rs17287498			
rs17334270			
rs4647964			
rs11003124			
rs11003125			
rs34130848			
rs2099902	6	Minimal binding evidence	Motif hit
rs2120131			
rs1982266			
rs10824794			
rs1031101			

RDB RegulomeDB, TF transcription factor, eQTL expression quantitative trait loci

MBL2 is a 148-residue peptide, stretches from 107 to 248 aa residues consisting of the 'neck' and carbohydrate recognition domains forming trimers in solution and in crystals. MUSTER provided highest score for the same template (PDB ID 1HUP) with Z score 13.966 and alignment length 141 aa residues with coverage of 0.568. However, the rs1800450 nsSNP lies in collagenous domain of human MBL protein (Super et al. 1989; Summerfield et al. 1997; Wallis and Cheng 1999; Wallis 2002; Turner 2003; Larsen et al. 2004). Hence, protein modelling based on its structural information was necessary for absolute understanding of its functionality. I-TASSER was chosen

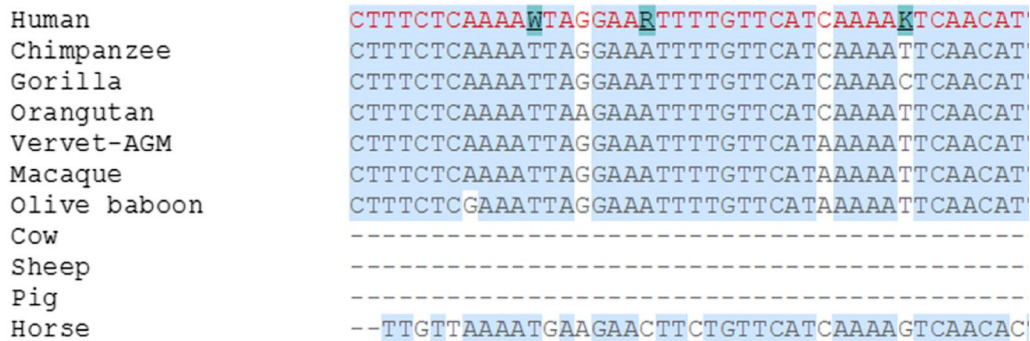


Fig. 3 A comparison of the human *MBL2* gene with eutherian mammals: the screenshot compares human chromosome no. 10: 52764780-52770775 bp region with homologous regions of other mammals. Human *MBL2* gene sequence indicated in red color. Conserved region are represented by blue highlighted region. The highlighted nucleotide indicates SNPs. The screen shot was taken from Ensembl Genome browser release 48. Only species showing alignment with human *MBL2* gene are shown in figure

Table 3 List of 3' UTR SNPs analyzed by PolymiRTS

dbSNP ID	Alleles	miRNA motif
rs2099902	G	hsa-miR-797
rs2099903	G	hsa-miR-4666a-5p/hsa-miR-8060
	T	hsa-miR-1252-3p/hsa-miR-3646
rs2165813	C	hsa-miR-1253/hsa-miR-6770-5
	T	hsa-miR-452-5p/hsa-miR-4676-3p/hsa-miR-4738-3p/hsa-miR-539-5p/hsa-miR-548 g-3p/hsa-miR-892c-3p
rs10082466	T	hsa-miR-4719/hsa-miR-513b-5p
	C	hsa-miR-27a-3p/hsa-miR-27b-3p/hsa-miR-3185/hsa-miR-513a-5p

Table 4 List of 3' UTR SNPs screened based on all the tools used

SNP ID	FuncPred	Regu-lomeDB score	cSNPs	PolymiRTS	Prioritize
rs2099902	✓	6	✓	✓	*
rs10082466	✓	5	✓	✓	*
rs2099903	✓	-	✓	✓	*
rs2165813	✓	-	✓	✓	*
rs2120131	✓	6	✓	-	-
rs10824792	✓	6	✓	-	-
rs2506	✓	-	✓	-	-
rs2083771	✓	-	✓	-	-
rs2120132	✓	-	✓	-	-

cSNPs conserved SNPs, ✓ SNPs which affect function, - SNPs which does not affect function, * prioritized for further studies

Table 5 List of non-coding SNPs except of 3' UTR screened based on all the tools used

Region	Snp ID	FuncPred	Regu-lomeDB score	cSNPs	Selected
5' near gene	rs7095891	✓	4	✓	*
	rs11003123	✓	4	✓	*
	rs7096206	✓	1f	✓	*
	rs10556764	-	3a	✓	-
	rs36014597	✓	4	✓	*
	rs7084554	✓	3a	✓	*
	rs11003124	✓	5	✓	*
	rs11003125	✓	5	✓	*
	rs920724	✓	-	✓	-
	rs7916582	✓	-	✓	-
	rs10824794	✓	6	✓	-
	rs10824795	✓	-	✓	-
	rs1031101	✓	6	✓	-
	Intronic	rs1982266	✓	6	✓
rs17334270		✓	5	✓	-
rs13312736		✓	4	✓	-
rs4647964		-	5	✓	-
rs1838065		-	5	✓	-
rs1838066		-	5	✓	-
rs34130848		-	5	✓	-
rs4935046	-	5	✓	-	
rs930508	-	3a	✓	-	
rs930509	-	4	✓	-	

cSNPs conserved SNPs, ✓ SNPs which affect function, - SNPs which does not affect function, * prioritized for further studies

to predict MBL monomeric protein secondary structure. Five models were obtained as an output, of these first model with highest C score of -2.80, TM score of

0.39 ± 0.13 and estimated RMSD 12.4 ± 4.3 Å was considered for the analysis of highly damaging nsSNPs on the structure and function of protein. The Ramachandran

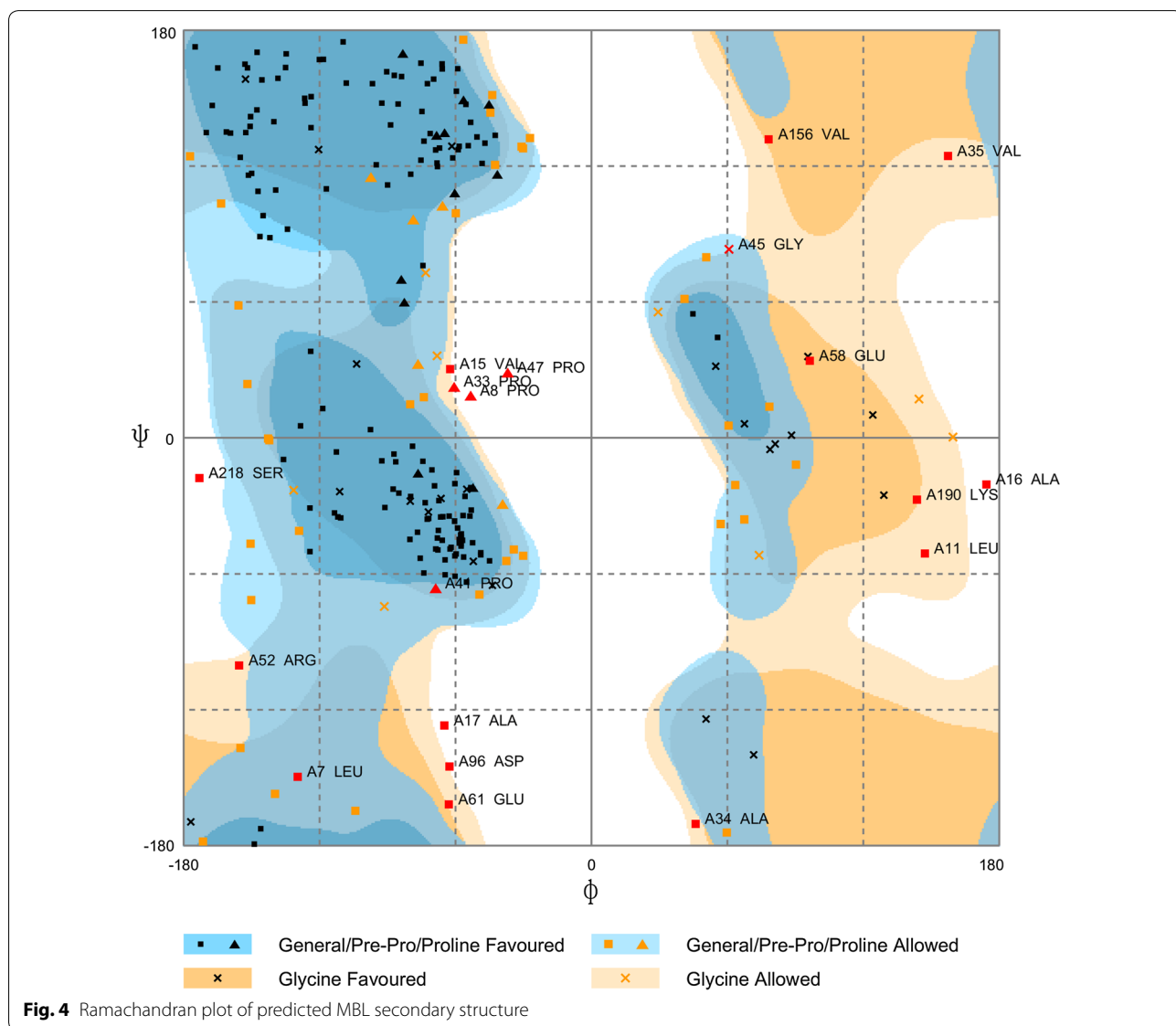
plot was also assessed to support quality of predicted MBL model using RAMPAGE server (Fig. 4). The model showed good proportions of residues in favored (72.4 %), allowed (19.9 %) and outlier regions (7.7 %). The results indicated that the 3D model was of fair quality. I-TASSER also predicted that native Gly residue with score of 3 as buried residue and variant Asp residue with score of 7 as highly exposed in terms of solvent accessibility.

To get variant modeled structure at position 54 i.e. Gly → Asp conversion Swiss-PdbViewer and chimera was used (Fig. 5). This variation leads to slight deviation in various parameters including total energy after minimization, total electrostatic constraint, total bond energy, torsion and non-bonded energy (Table 6). It was found that Gly → Asp conversion at position 54 showed a network of clashes with nearby residues that is Asp at position 53 and Gly at position 48 as shown in Fig. 6 while

this network is lacking when native residue glycine was there. Furthermore, analysis by I-mutant indicated that Gly → Asp conversion at position 54 decreases the stability of protein with DDG value of −1.37 kcal/mol.

Discussion

Mannose binding lectin (MBL) is a pattern recognition molecule of the innate immune system. It plays an important role in the first line of defense and provides primary immune response to pathogens and thus is important when adaptive immune response will not operate due to some reasons (Ip et al. 2009). Studies have indicated association of some SNPs of *MBL2* gene with susceptibility to various diseases as well as were shown to affect serum MBL levels (Madsen et al. 1995; Summerfield et al. 1995; Garred et al. 1997; Kelly et al. 2000; Klabunde et al. 2000; Jack and Turner 2003; IP et al. 2005; Garred et al. 2006;



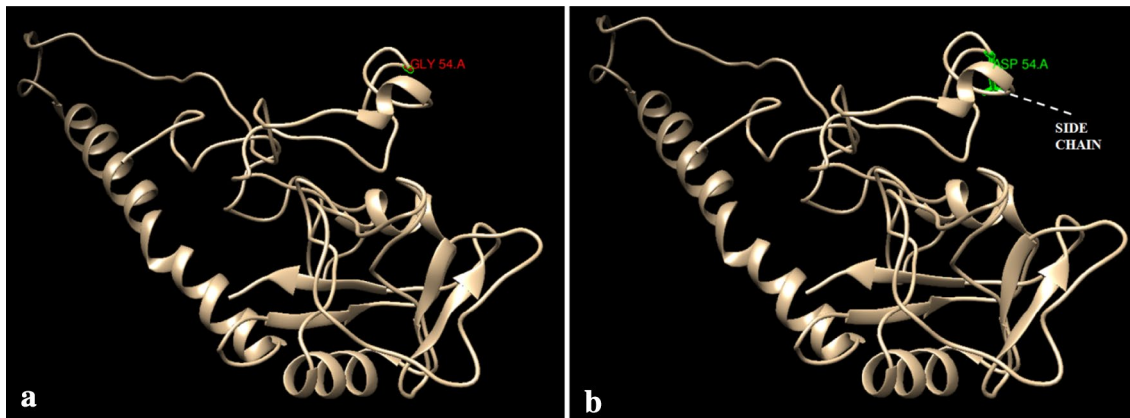


Fig. 5 Full length models of MBL protein monomer with **a** native residue glycine (red colored and labeled) at position 54, **b** variant residue aspartic acid (green colored and labeled) with side chain indicated by dotted white lines

Table 6 Parameters of native and mutant structure after energy minimization

Parameters	Native amino acid structure	Variant amino acid structure
Total energy (KJ/mol)	-5854.403	-5852.157
Total electrostatic constraint	-6399.19	-4496.11
Total bond energy (KJ/mol)	1833.712	1834.508
Torsion (KJ/mol)	2475.780	2485.522
Non-bonded energy (KJ/mol)	-6399.19	-6385.21

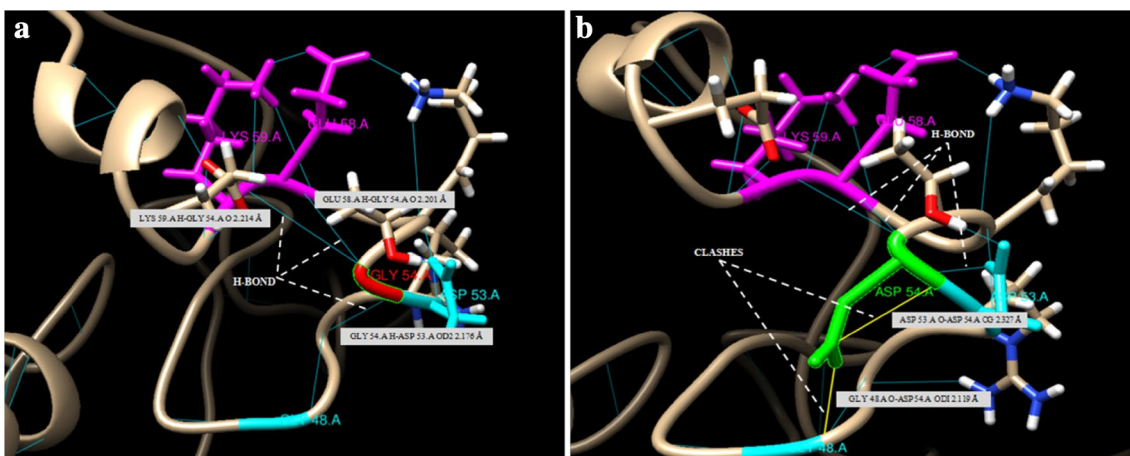


Fig. 6 3D analysis of wild and variant residues at position 54 **a** Gly (red color) at 54 position, forms three H-bonds (blue lines, indicated with white dotted lines) with nearby residues Asp 53 (sky-blue color), Glu 58 and Lys 59 (purple color). **b** Network of clashes (yellow lines, indicated with white dotted lines) appeared between variant residue Asp 54 (green color) with Gly 48 and Asp 53 (sky-blue color)

Alonso et al. 2007). However, there are large numbers of SNPs present in the human *MBL2* gene that have been identified. But it is difficult to study all the SNPs present in a gene for candidate gene approach as some SNPs may have no functional effect or of very low frequency. As

the interpretation of clinically important novel variants often remain challenging, many bioinformatics tools have been developed that predict biological consequences of these polymorphisms. The present study has hauled out some novel SNPs of MBL protein along with prediction

of biological effect of some already studied SNPs using these tools. Thus focus can be put on studying those SNPs which probably may have important role in various diseases. Variants of non-coding and coding regions are equally important and hence are the focus of the present study (Alexander et al. 2010; Birney et al. 2007).

To identify most likely functional variants of *MBL2* gene, of which many are still unknown, in silico analyses was used. However, no single bioinformatics tool can be used to obtain a complete picture of the functional significance of allelic variants. Hence the current analysis was conducted using a number of complementary bioinformatic tools. As expected the results obtained from different tools were not directly overlapping for all the studied SNPs. Four composite tools SNPinfo, Genomic alignments by Ensembl, regulomeDB and poylmiRTS were used for the analysis of non-coding SNPs. The entire four tools played equal important role in prioritizing functional SNPs leading to selection of 11 non-coding SNPs for disease association studies. Two of the most widely accepted tools SIFT and PolyPhen-2 was used to predict the phenotypic effect of nsSNP on the physico-chemical properties of the concerned proteins and thus expression. Both the tools predicted rs1800450, a deleterious and damaging single nucleotide variant.

Out of 12 SNPs, 4 SNPs with rs numbers rs11003125, rs7095891, rs7096206 and rs1800450 are very well known and a plethora of literature is available regarding their functional and deleterious effect and their association with various diseases including filariasis, malaria, leishmaniasis, leprosy, tuberculosis, trypanosomiasis, HIV infection, systemic lupus erythematosus, rheumatoid arthritis (Choi et al. 2001; Boldt et al. 2006; Garred et al. 2006; Asgharzadeh et al. 2007; de Messias-Reason et al. 2007; Ip et al. 2009; Meyrowitsch et al. 2010; Panda et al. 2013; Martiny et al. 2012; Singla et al. 2012; Weitzel et al. 2012; Li et al. 2013; Chen et al. 2014; Jha et al. 2014). rs1800450 (Gly 54 Asp) variation leads to the production of nonfunctional monomers that further interfere with the formation of higher MBP oligomers, leading to alterations in the functional activity of the protein and their circulating levels (Sumiya et al. 1991; Lipscombe et al. 1992; Madsen et al. 1994; Terai et al. 2003; Larsen et al. 2004). Three SNPs in Promoter region rs7095891, rs7096206 and rs11003125 are functionally characterized to alter the transcriptional levels significantly contributing the large variation in the promoter activity of the *MBL2* gene elucidating their biological significance with respect to gene expression and hence the serum MBL levels (Naito et al. 1999b; Juliger et al. 2000). The combined effect of these polymorphisms on MBL basal expression and serum levels was also studied (Madsen et al. 1995; Garred et al. 2006).

Furthermore, a single study has indicated that rs2099902, rs10082466 variants of 3'-UTR and functional secretor haplotypes in *MBL2* are associated with increased colon cancer risk in African Americans. The C allele of rs10082466 exhibited a binding affinity of miR-27a and this allele was associated with both lower MBL plasma levels and activity (Zanetti et al. 2012). So validation of these SNPs by these reports complements the finding of the present study. However, very scanty data is available regarding role of rs11003123, rs36014597, rs7084554, rs11003124 SNPs of *MBL2* gene only a single study had shown the association of rs11003124 with leprosy in Han Chinese population (Boldt et al. 2009; Swale et al. 2014; Wong et al. 2012; Zhang et al. 2013). Furthermore, no clinical relevance of rs2165813 and rs2099903 SNP is documented till date. Thus, more epidemiological and clinical studies are needed to validate the SNPs reported in the present study.

MBL is a complex of six sets of homotrimers of a monomer containing 228 amino acids (Ezekowitz et al. 1988; Sastry et al. 1989; Taylor et al. 1989; Kurata et al. 1994). This monomer consists of four domains a 20-amino acid N-terminal cysteine-rich domain, a collagen-like domain consisting of 18–20 tandem repeats of Gly-Xaa-Yaa, an alpha helical coiled-coil neck region, and a carbohydrate recognition domain. The neck region initiates the folding and the collagen-like region zips toward the N terminus, creating trimeric subunits. Interchain disulphide bonds link the N terminal domains of polypeptides together both within and between subunits to form and stabilize higher oligomers (Wallis and Drickamer 1999). Human MBL comprises dimers to hexamers of trimeric subunits of which trimers and tetramers are probably the predominant form in circulation (Lu et al. 1990; Teillet et al. 2005). However, a variant monomer may result into the formation of nonfunctional, low-order oligomers with shortened half life (Lipscombe et al. 1995; Naito et al. 1999a; Petersen et al. 2001).

Although *MBL2* gene is widely known because of its important role and association with many diseases, still the full native protein structure is not yet available in protein data bank (PDB). Native protein structure is important for the in silico analysis of functional polymorphisms. As already explained in results, only partial structure, with PDB ID 1HUP corresponding to neck and carbohydrate recognition domains of human MBL is available (Sheriff et al. 1994). Gly54Asp variation lies in the collagenous domain of human MBL, therefore, I-TASSER was used for predicting 3D secondary structure of native MBL monomeric protein that includes all the four domains along with the signal peptide and analyzed for damaging mutation predicted by SIFT and polyphen-2. The TM-score of the present study predicted

structure (>0.5) which represents correct topology for protein while higher value of C-score signifies model with a higher confidence represents good quality protein. Ramachandran plot also supported the quality of monomeric MBL protein model.

The comparison of specific properties of native and variant structures revealed difference in stability. This variation leads to slight deviation in energy and decrease in stability. This change in stability can be explained by the reason that the variant residue has bulkier R group than the wild type and cannot fit within the available space. The free carboxylic acid group on Asp can disturb the ionic interaction of nearby residues and hence causing vulnerable effects. The wild-type residue is non-polar while the variant residue is negatively charged and hydrophilic. Furthermore, there is difference in solvent accessibility of both residues with native being buried and variant being fully exposed, hence slight deviation in energy and decrease in stability.

This variation which lies in the fifth Gly-Xaa-Yaa repeat distort the collagen-like region had been shown to interfere with normal assembly of the MBL oligomeric form and its interactions with MBL associated serine proteases (Wallis and Cheng 1999; Wallis and Dodd 2000; Turner 2003; Larsen et al. 2004; Wallis et al. 2004). Some studies have suggested that the MBL triple helix is formed when the variant state has an enthalpy similar to the native state but with reduced stability (Wallis and Cheng 1999; Wallis and Dodd 2000; Wallis et al. 2004). The structural and energetic consequences of Gly \rightarrow Asp conversion were also investigated by synthesizing two triple-helical peptides containing Gly/Asp at position 54 (Mohs et al. 2005). The study indicated loss of triple helix content by Gly \rightarrow Asp replacement accompanying with small decrease in stability of triple helix. However, no enthalpy change was observed. Furthermore NMR studies indicated the complete distortion of triple helix in peptide containing variant residue. Hence, providing evidence in support of predictions made by SIFT, Polyphen-2, I-mutant and clashes that was observed during Gly \rightarrow Asp conversion in the theoretical modeling of MBL monomeric peptide.

In the present study, we have used the relatively simple structure i.e. monomeric polypeptide of MBL to examine and understand the effect of this variation on its more complex homolog using theoretical modeling. Extensive studies had been carried out to investigate the effect of Gly substitution in the collagen like domain using synthetic peptides which offer an approach to characterizing the effect of mutations on the triple-helix (Baum and Brodsky 1999). Bella with other colleagues reported the high-resolution structure of a (Pro-Hyp-Gly)₁₀ peptide with a Gly to Ala substitution near the center

indicated a local loss of direct hydrogen bonding and a slight untwisting at the mutation site when the three collagen like peptide chains form a triple helix in solution (Bella et al. 1994). Similar effort was made in the present study using in silico tools where we used the theoretical 3D model of MBL monomer polypeptide to visualize the effect of Gly \rightarrow Asp variation at position 54 and the sequence surrounding the replacement and it was found that Asp at position 54 showed a network of clashes with nearby residues while this network was lacking when the native residue glycine was there. Thus the clashes observed must interfere in the normal triple helix formation of trimeric subunit and further with the normal assembly of MBL oligomeric form.

There are some essential features that are used in this study to avoid the common pitfalls while conducting an in silico analysis. Feature one, looking for validation status to check the reliability of a given SNP as all SNPs in dbSNP are not real. Some polymorphisms might have arisen solely due to sequencing errors and others may be sole to the individuals (Fredman et al. 2002). Second important feature is prioritizing these SNPs that lie within an evolutionary conserved region to remove false-positive predictions that encountered due to in silico tools (Fairbrother et al. 2002; Cartegni et al. 2003; Yeo and Burge 2004). The high rate of false positive findings produced by in silico prediction tools can be due to the short length of sequences (typically 6–8-mer) used in computer simulations (Yeo and Burge 2004). Thus, this strategy was used in the current study to validate the in silico predictions.

Third and very important feature to prioritize functional SNPs identified with in silico tools is the use of MAF. Because MAF is linked to statistical power of the study i.e. MAF and sample size to detect variant allele of a SNP in a given sample population have inverse relationship (Grover et al. 2007). Therefore, SNPs with a MAF of 0.05 or more are generally targeted in the majority of large scale genome studies for instance the international HapMap project. Keeping this in view, MAF of 0.10 or more were used for the present analysis. Thus by keeping in mind these features and based on the results of tools 12 SNPs out of 661 SNPs of *MBL2* were found to be functionally important for candidate gene studies.

Conclusion

The present study demonstrated that 12 SNPs in the *MBL2* gene are functionally important as well as deleterious to its structure and expression. These conserved SNPs may broaden our understanding of genotype phenotype relationship. Out of these 12 SNPs only 4 SNPs i.e. rs11003125, rs7095891, rs7096206 and rs1800450 are widely studied. But the rest of 8 functional SNPs

rs2099902, rs10082466, rs11003123, rs36014597, rs7084554, rs11003124, rs2165813, rs2099903 reported by the present study haven't been explored much till date. Thus, the findings of the present study provide a guideline to the fellow researchers to know important role of these SNPs in the etiology of complex diseases.

Abbreviations

ChIPseq: chromatin immunoprecipitation sequencing; cSNPs: conserved SNPs; Eqt1: expression quantitative trait loci; ESEs: exonic splicing enhancers; MBL: mannose binding lectin; *MBL2*: mannose binding lectin gene; nsSNP: non-synonymous SNPs; PDB: protein data bank; Polyphen-2: polymorphism phenotyping v2; RAMPAGE: ramachandran plot analysis; SIFT: sorting intolerant from tolerant; SNP: single nucleotide polymorphism; TFBS: transcription factor binding sites; TFs: transcription factors; UTR: untranslated regions.

Authors' contributions

NK interpreted the data, reviewed the literature and drafted the manuscript. AS reviewed the literature. MK, SSK, JS participated in the data analysis, manuscript editing and supervision. All authors read and approved the final manuscript.

Author details

¹ Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar 143005, India. ² Department of Human Genetics, Guru Nanak Dev University, Amritsar, India.

Acknowledgements

NK is thankful to University Grants Commission, New Delhi, for Ph.D. fellowship.

Competing interests

The authors declare that they have no competing interests.

Received: 16 March 2016 Accepted: 8 June 2016

Published online: 21 June 2016

References

- Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. doi:10.1002/0471142905.hg0720s76
- Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB (2010) Annotating non-coding regions of the genome. *Nat Rev Genet* 11:559–571. doi:10.1038/nrg2814
- Alonso DP, Ferreira AF, Ribolla PE et al (2007) Genotypes of the mannan-binding lectin gene and susceptibility to visceral leishmaniasis and clinical complications. *J Infect Dis* 195:1212–1217
- Asgharzadeh M, Mazloumi A, Kafili HS, Ghazanchaei A (2007) Mannose-binding lectin gene and promoter polymorphism in visceral leishmaniasis caused by *Leishmania infantum*. *Pak J Biol Sci* 10:1850–1854. doi:10.3923/jpjs.2007.1850.1854
- Baum J, Brodsky B (1999) Folding of peptide models of collagen and misfolding in disease. *Curr Opin Struct Biol* 9:122–128
- Bella J, Eaton M, Brodsky B, Berman HM (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 266:75–81
- Bhagwat M (2010) Searching NCBI's dbSNP database. *Curr Protoc Bioinform* 1:1–19. doi:10.1002/0471250953.bi0119s32
- Bhattacharya A, Ziebarth JD, Cui Y (2014) PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res* 42:D86–D91. doi:10.1093/nar/gkt1028
- Bhatti P, Church DM, Rutter JL, Struwing JP, Sigurdson AJ (2006) Candidate single nucleotide polymorphism selection using publicly available tools: a guide for epidemiologists. *Am J Epidemiol* 164:794–804
- Birney E, Stamatoyanopoulos JA, Dutta A et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816
- Boldt AB, Luty A, Grobusch MP, Dietz K, Dzeing A, Kombila M, Kreamsner PG, Kun JF (2006) Association of a new mannose-binding lectin variant with severe malaria in Gabonese children. *Genes Immun* 7:393–400. doi:10.1038/sj.gene.6364312
- Boldt AB, Messias-Reason IJ, Lell B, Issifou S, Pedroso ML, Kreamsner PG, Kun JF (2009) Haplotype specific-sequencing reveals MBL2 association with asymptomatic *Plasmodium falciparum* infection. *Malar J* 8:97. doi:10.1186/1475-2875-8-97
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M et al (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22:1790–1797. doi:10.1101/gr.137323.112
- Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 33:306–310
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31:3568–3571
- Chasman D, Adams RM (2001) Predicting the functional consequences of nonsynonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 307:683–706
- Chen M, Deng J, Su C, Li J, Wang M, Abuaku BK, Hu S, Tan H, Wen SW (2014) Impact of passive smoking, cooking with solid fuel exposure, and MBL/MASP-2 gene polymorphism upon susceptibility to tuberculosis. *Int J Infect Dis* 29:1–6. doi:10.1016/j.ijid.2014.08.010
- Choi EH, Zimmerman PA, Foster CB, Zhu S, Kumaraswami V, Nutman TB, Chanoock SJ (2001) Genetic polymorphisms in molecules of innate immunity and susceptibility to infection with *Wuchereria bancrofti* in South India. *Genes Immun* 2:248–253
- de Messias-Reason IJ, Boldt AB, Braga AC, Stahlke EV, Dornelles L, Pereira-Ferrari L, Kreamsner PG, Kun JF (2007) The association between mannan-binding lectin gene polymorphism and clinical leprosy: new insight into an old paradigm. *J Infect Dis* 196:1379–1385. doi:10.1086/521627
- Ezekowitz RAB, Day LE, Herman GA (1988) A human mannose binding protein is an acute-phase reactant that shares sequence homology with other vertebrate lectins. *J Exp Med* 167:1034–1046
- Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297:1007–1013
- Fredman D, Siegfried M, Yuan YP, Bork P, Lehvaslaiho H, Brookes AJ (2002) HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Res* 30:387–391
- Garred P, Richter C, Andersen AB, Madsen HO, Mtoni I, Sveigaard A, Shao J (1997) Mannan-binding lectin in the sub-Saharan HIV and tuberculosis epidemics. *Scand J Immunol* 46:204–208
- Garred P, Larsen F, Seyfarth J, Fujita R, Madsen HO (2006) Mannose binding lectin and its genetic variants. *Genes Immun* 7:85–94. doi:10.1038/sj.gene.6364283
- Grover D, Woodfield AS, Verma R et al (2007) QuickSNP: an automated web server for selection of tag SNPs. *Nucleic Acids Res* 35:15–20
- Guex N, Peitsch MC (1997) Swiss-Model and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
- Guo N, Mogue T, Weremowicz S, Morton CC, Sastry KN (1998) The human ortholog of rhesus mannose-binding protein-A gene is an expressed pseudogene that localizes to chromosome 10. *Mamm Genome* 9:246–249
- Ip WK, Chan KH, Law HK, Tso GH, Kong EK, Wong WH, To YF, Yung RW, Chow EY, Au KL, Chan EY, Lim W, Jensenius JC, Turner MW, Peiris JS, Lau Y (2005) Mannose binding lectin in severe acute respiratory syndrome coronavirus infection. *J Infect Dis* 191:1697–1704
- Ip WK, Takahashi K, Alan Ezekowitz R, Stuart LM (2009) Mannose binding lectin and innate immunity. *Immunol Rev* 230:9–21
- Jack DL, Turner MW (2003) Antimicrobial activities of mannose binding lectin. *Biochem Soc Trans* 31:753–757
- Jha AN, Sundaravadevel P, Singh VK, Pati SS, Patra PK, Kreamsner PG, Velavan TP, Singh L, Thangaraj K (2014) MBL2 variations and malaria susceptibility in Indian populations. *Infect Immun* 82:52–61. doi:10.1128/IAI.01041-13
- Johnson AD (2009) Single-nucleotide polymorphism bioinformatics: a comprehensive review of resources. *Circ Cardiovasc Genet* 2:530–536. doi:10.1161/CIRCGENETICS.109.872010

- Juliger S, Luckner D, Mordmuller B, May J, Weierich A, Lell B, Luty A, Kreamsner PG, Kun JF (2000) Promoter variants of the human mannose-binding lectin gene show different binding. *Biochem Biophys Res Commun* 275:617–622. doi:10.1006/bbrc.2000.3343
- Kelly P, Jack DL, Naeem A, Mandanda B, Pollok RC, Klein NJ, Turner MW, Farthing MJ (2000) Mannose binding lectin is a component of innate mucosal defense against *Cryptosporidium parvum* in AIDS. *Gastroenterology* 119:1236–1242
- Klabunde J, Berger J, Jensenius JC, Klunkert MQ, Zelck UE, Kreamsner PG, Kun JF (2000) *Schistosoma mansoni*: adhesion of mannan-binding lectin to surface glycoproteins of cercariae and adult worms. *Exp Parasitol* 95:231–239
- Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073–1108. doi:10.1038/nprot.2009.86
- Kurata H, Sannoh T, Kozutsumi Y, Yokota Y, Kawasaki T (1994) Structure and function of mannan-binding proteins isolated from human liver and serum. *J Biochem* 115:1148–1154
- Larsen F, Madsen HO, Sim RB, Koch C, Garred P (2004) Disease associated mutations in human mannose-binding lectin compromise oligomerization and activity of the final protein. *J Biol Chem* 279:21302–21311. doi:10.1074/jbc.M400520200
- Li L, Wei D (2015) Bioinformatics tools for discovery and functional analysis of single nucleotide polymorphisms. *Advance in structural bioinformatics*. Springer, Dordrecht, pp 287–310
- Li H, Fu WP, Hong ZH (2013) Replication study in Chinese Han population and meta-analysis supports association between the MBL2 gene polymorphism and HIV-1 infection. *Infect Genet Evol* 20:163–170. doi:10.1016/j.meegid.2013.09.010
- Lipscombe RJ, Sumiya M, Hill AV, Lau YL, Levinsky RJ, Summerfield JA, Turner MW (1992) High frequencies in African and non-African populations of independent mutations in the mannose-binding protein gene. *Hum Mol Genet* 1:709–715. doi:10.1093/hmg/1.9.709
- Lipscombe RJ, Sumiya M, Summerfield JA, Turner MW (1995) Distinct physico-chemical characteristics of human mannose binding protein expressed by individuals of differing genotype. *Immunology* 85:660–667
- Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson DC (2003) Structure validation by Ca geometry: ϕ , ψ and $C\beta$ deviation. *Proteins* 50:437–450
- Lu J, Thiel S, Wiedemann H, Timpl R, Reid KBM (1990) Binding of the pentamer/hexameric forms of a mannan-binding protein to zymosan activates the proenzyme C1r2C1s2 complex of the classical pathway of complement, without involvement of C1q. *J Immunol* 144:2287–2294
- Madsen HO, Garred P, Kurtzhals JA, Lamm LU, Ryder LP, Thiel S, Svejgaard A (1994) A new frequent allele is the missing link in the structural polymorphism of the human mannan-binding protein. *Immunogenetics* 40:37–44. doi:10.1007/BF00163962
- Madsen HO, Garred P, Thiel S, Kurtzhals JA, Lamm LU, Ryder LP, Svejgaard A (1995) Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J Immunol* 155:3013–3020
- Martiny FL, Veit TD, Brenol CV, Brenol JC, Xavier RM, Bogo MR, Chies JA (2012) Mannose-binding lectin gene polymorphisms in Brazilian patients with rheumatoid arthritis. *J Rheumatol* 39:6–9. doi:10.3899/jrheum.110052
- Meyrowitsch DW, Simonsen PE, Garred P, Dalgaard M, Magesa SM, Alifrangis M (2010) Association between mannose-binding lectin polymorphisms and *Wuchereria bancrofti* infection in two communities in North-Eastern Tanzania. *Am J Trop Med Hyg* 82:115–120. doi:10.4269/ajtmh.2010.09-0342
- Mohs A, Li Y, Doss-Pepe E, Baum J, Brodsky B (2005) Stability junction at a common mutation site in the collagenous domain of the mannose binding lectin. *Biochemistry* 44:1793–1799
- Naito H, Ma Y, Uemura K, Arano Y, Kawasaki T (1999a) Metabolic properties of normal and mutant mannan-binding proteins in mouse plasma. *Biochem Biophys Res Commun* 256:231–234
- Naito H, Ikeda A, Hasegawa K, Oka S, Uemura K, Kawasaki N, Kawasaki T (1999b) Characterization of human serum mannan-binding protein promoter. *J Biochem* 126:1004–1012. doi:10.1093/oxfordjournals/jbchem.a022543
- Nepomuceno RR, Henschen-Edman AH, Burgess WH, Tenner AJ (1997) cDNA cloning and primary structure analysis of C1qR(P) the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro. *Immunity* 6:119–129
- Panda AK, Parida JR, Tripathy R, Pattanaik SS, Ravindran B, Das BK (2013) Low producer MBL genotypes are associated with susceptibility to systemic lupus erythematosus in Odisha, India. *Hum Immunol* 74:114–119. doi:10.1016/j.humimm.2012.09.003
- Petersen SV, Thiel S, Jensenius JC (2001) The mannan-binding lectin pathway of complement activation: biology and disease association. *Mol Immunol* 38:133–149
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738. doi:10.1038/nprot.2010.5
- Sastry K, Herman GA, Day L, Deignan E, Bruns G, Morton CC, Ezekowitz RA (1989) The human mannose-binding protein gene. Exon structure reveals its evolutionary relationship to a human pulmonary surfactant gene and localization to chromosome 10. *J Exp Med* 170:1175–1189
- Selander B, Martensson U, Weintraub A et al (2006) Mannan binding lectin activates C3 and the alternative complement pathway without involvement of C2. *J Clin Invest* 116:1425–1434
- Sheriff S, Chang CY, Ezekowitz RAB (1994) Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α -helical coiled-coil. *Nat Struct Mol Biol* 1:789–794
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski M, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311
- Singla N, Gupta D, Joshi A, Batra N, Singh J, Birbian N (2012) Association of mannose-binding lectin gene polymorphism with tuberculosis susceptibility and sputum conversion time. *Int J Immunogenet* 39:10–14. doi:10.1111/j.1744-313X.2011.01047.x
- Sumiya M, Super M, Tabona P, Levinsky RJ, Arai T, Turner MW, Summerfield JA (1991) Molecular basis of opsonic defect in immunodeficient children. *Lancet* 337:1569–1570. doi:10.1016/0140-6736(91)93263-9
- Summerfield JA, Ryder S, Sumiya M, Thursz M, Gorchein A, Monteil MA, Turner MW (1995) Mannose-binding protein gene mutations associated with unusual and severe infections in adults. *Lancet* 345:886–889
- Summerfield JA, Sumiya M, Levin M, Turner MW (1997) Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series. *Br Med J* 314:1229–1232
- Super M, Thiel S, Lu J, Levinsky RJ, Turner MW (1989) Association of low levels of mannan-binding protein with a common defect of opsonization. *Lancet* 334:1236–1239
- Swale A, Miyajima F, Kolamunnage-Dona R, Roberts P, Little M, Beeching NJ, Pirmohamed M (2014) Serum mannose binding lectin concentration, but not genotype, is associated with *Clostridium difficile* infection recurrence: a prospective cohort study. *Clin Infect Dis* 59:1429–1436
- Taylor ME, Brickell PM, Craig RK, Summerfield JA (1989) Structure and evolutionary origin of the gene encoding a human serum mannose-binding protein. *Biochem J* 262:763–771
- Teillet F, Dublet B, Andrieu JP, Gaboriau C, Arlaud GJ, Thielens NM (2005) The two major oligomeric forms of human mannan-binding lectin: chemical characterization, carbohydrate-binding properties, and interaction with MBL-associated serine proteases. *J Immunol* 174:2870–2877
- Terai I, Kobayashi K, Matsushita M, Miyakawa H, Mafune N, Kikuta H (2003) Relationship between gene polymorphisms of mannose-binding lectin (MBL) and two molecular forms of MBL. *Eur J Immunol* 33:2755–2763. doi:10.1002/eji.200323955
- Turner MW (2003) The role of mannose-binding lectin in health and disease. *Mol Immunol* 40:423–429
- Wallis R (2002) Structural and functional aspects of complement activation by mannose-binding protein. *Immunobiology* 205:433–445
- Wallis R, Cheng JY (1999) Molecular defects in variant forms of mannose-binding protein associated with immunodeficiency. *J Immunol* 163:4953–4959
- Wallis R, Dodd RB (2000) Interaction of mannose-binding protein with associated serine proteases. Effects of naturally occurring mutations. *J Biol Chem* 275:30962–30969
- Wallis R, Drickamer K (1999) Molecular determinants of oligomer formation and complement fixation in mannose-binding proteins. *J Biol Chem* 274:3580–3589

- Wallis R, Shaw JM, Uitdehaag J, Chen CB, Torgersen D, Drickamer K (2004) Localization of the serine protease binding sites in the collagen-like domain of mannose-binding protein: indirect effects of naturally occurring mutations on protease binding and activation. *J Biol Chem* 279:14065–14073
- Weitzel T, Zulantay I, Danquah I, Hamann L, Schumann RR, Apt W, Mockenhaupt FP (2012) Mannose-binding lectin and Toll-like receptor polymorphisms and Chagas disease in Chile. *Am J Trop Med Hyg* 86:229–232. doi:10.4269/ajtmh.2012.11-0539
- Wong M, Ohrmalm L, Broliden K, Aust C, Hibberd M, Tolfvenstam T (2012) Mannose binding lectin 2 polymorphisms do not influence frequency or type of infection in adults with chemotherapy induced neutropaenia. *PLoS ONE* 7:e30819. doi:10.1371/journal.pone.0030819
- Wu S, Zhang Y (2008) MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. *Proteins* 72:547–556. doi:10.1002/prot.21945
- Xu Z, Taylor JA (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 37:600–605. doi:10.1093/nar/gkp290
- Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377–394
- Zanetti KA, Haznadar M, Welsh JA, Robles AI, Ryan BM, McClary AC, Harris CC (2012) 3'-UTR and functional secretor haplotypes in mannose binding lectin 2 are associated with increased colon cancer risk in African Americans. *Cancer Res* 72:1467–1477. doi:10.1158/0008-5472.CAN-11-3073
- Zhang DF, Huang XQ, Wang D, Li YY, Yao YG (2013) Genetic variants of complement genes ficolin-2, mannose binding lectin and complement factor H are associated with leprosy in Han Chinese from Southwest China. *Hum Genet* 132:629–640. doi:10.1007/s00439-013-1273-8

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
