

Screening and structural evaluation of deleterious Non-Synonymous SNPs of ePHA2 gene involved in susceptibility to cataract formation

Tariq Ahmad Masoodi^{1*}, Sulaiman A Al Shammari¹, May N Al-Muammar¹, Turki M Almubrad² & Adel A Alhamdan¹

¹Health Care Development for Elderly Research Chair, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; ²Department of Optometry, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; Tariq Ahmad Masoodi – Email: tahamasoodi@gmail.com; Phone +966 (1) 4693705; Fax: +0966 (1) 4693698; *Corresponding author

Received May 21, 2012; Accepted May 24, 2012; Published June 28, 2012

Abstract:

Age-related cataract is clinically and genetically heterogeneous disorder affecting the ocular lens, and the leading cause of vision loss and blindness worldwide. Here we screened nonsynonymous single nucleotide polymorphisms (nsSNPs) of a novel gene, EPHA2 responsible for age related cataracts. The SNPs were retrieved from dbSNP. Using I-Mutant, protein stability change was calculated. The potentially functional nsSNPs and their effect on protein was predicted by PolyPhen and SIFT respectively. FASTSNP was used for functional analysis and estimation of risk score. The functional impact on the EPHA2 protein was evaluated by using SWISSPDB viewer and NOMAD-Ref server. Our analysis revealed 16 SNPs as nonsynonymous out of which 6 nsSNPs, namely rs11543934, rs2291806, rs1058371, rs1058370, rs79100278 and rs113882203 were found to be least stable by I-Mutant 2.0 with DDG value of > -1.0 . nsSNPs, namely rs35903225, rs2291806, rs1058372, rs1058370, rs79100278 and rs113882203 showed a highly deleterious tolerance index score of 0.00 by SIFT server. Four nsSNPs namely rs11543934, rs2291806, rs1058370 and rs113882203 were found to be probably damaging with PSIC score of ≥ 2.0 by Polyp hen server. Three nsSNPs namely, rs11543934, rs2291806 and rs1058370 were found to be highly polymorphic with a risk score of 3-4 with a possible effect of Non-conservative change and splicing regulation by FASTSNP. The total energy and RMSD value was higher for the mutant-type structure compared to the native type structure. We concluded that the nsSNP namely rs2291806 as the potential functional polymorphic that is likely to have functional impact on the EPHA2 gene.

Keywords: Computational analysis, single nucleotide polymorphism, EPHA2, cataract

Background:

Light scattering opacities of the crystalline lens, or cataracts, are a common ocular symptom of aging (>50 years), and despite advances in their surgical treatment age-related cataracts remain a leading cause of visual impairment, accounting for ~48% of world blindness [1], with approximately 37 million people affected [2]. Approximately 80% of all cataracts are age-related and idiopathic [3]. Age-related cataract has been defined as the appearance of the clinical sign of cataracts in one or both

eyes in a person older than 50 years [4]. Among Caucasian Americans aged 65 years or older, the reported prevalence of cataract is 24% [5]. Incidence of cataract surgery in the United States is highest amongst those over 70 years, with an annual increase of about 14% [6]. Age-related cataract has consistently been attributed to risk factors such as old age, female gender, diabetes, hypertension, smoking, UV light exposure, and heavy alcohol intake [7]. EPHA2 belongs to the tyrosine kinase family of proteins and is an epithelial cell kinase that has been

associated with autosomal dominant cataracts and recently it was implicated in age-related cortical cataracts in humans and mice [8-10]. EPHA2 interacts with membrane-bound ephrin ligands, which play an important role in morphogenesis and in numerous developmental processes [11]. Recently, several EPHA2 single nucleotide polymorphisms were found to be associated with age-related cortical cataracts in different worldwide Caucasian populations [8, 9, 12]. It is estimated that around 90% of human genetic variations are differences in single bases of DNA, known as single nucleotide polymorphisms (SNPs) [13, 14]. Among them, non-synonymous single nucleotide polymorphisms (nsSNPs), that cause amino acid changes in proteins have the potential to affect both protein structure and protein function [15, 18]. These nsSNPs can result in inherited diseases and drug sensitivity [19, 21]. These nsSNPs affect gene expression by modifying DNA and transcription factor binding [22, 23] and inactivate active sites of enzymes or change splice sites, thereby produce defective gene products [24, 25]. Some of the mutations in nsSNP sites are not associated with any changes in phenotype and are considered functional neutral, but others bringing deleterious effects to protein function and are responsible for many human genetic diseases [26, 27]. There are several databases with these variations of SNPs, such as the human genome variation database, HGVBBase [28], National Center for Biotechnology Information (NCBI) database, dbSNP [29], and SWISSPROT [30]. The large size of these databases presents a challenging hurdle for annotating the effects of all nsSNPs by experimental approaches. Therefore, computational methods that can quickly distinguish disease-causing nsSNPs from neutral nsSNPs are in urgent need.

Epidemiologic association studies focus a great amount of effort into identifying SNPs in genes that may have an association with disease risk. Often, the SNPs that have an association with disease are those that are known as nonsynonymous SNPs, which result in an amino acid substitution. Many molecular epidemiologic studies focus on studying SNPs found in coding regions in hopes of finding significant association between SNPs and disease susceptibility, but often find little or no association [31]. With the availability of high-throughput SNP detection techniques, the population of nsSNPs is increasing rapidly, providing a platform for studying the relationship between genotypes and phenotypes of human diseases. Our ability to better select a nsSNP for an association study can be enhanced by first examining the potential impact an amino acid variant may have on the function of the encoded protein with the use of different SNP detection programs like, I-Mutant, Sort Intolerant from Tolerant (SIFT) and Polymorphism Phenotype (PolyPhen) [31]. Discovering the deleterious nsSNPs out of a pool of all the SNPs will be very useful for epidemiological population based studies. So the main aim of this study is to identify deleterious nsSNPs associated with EPHA2 gene. We undertook this work mainly to perform computational analysis of the nsSNPs in EPHA2 gene and to identify the possible mutations and proposed modeled structures for the mutant proteins and compared them with the native protein. We report that, the mutation namely, E825K in the native protein of EPHA2 gene could be a candidate of major concern for cataract formation.

Methodology

Methodology used was the same as described earlier [31-33].

Screening of SNPs

We used dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) to identify SNPs and their related protein sequence for the EPHA2 gene [29].

Protein stability prediction

I-Mutant 2.0 software was used to predict nsSNP causing protein stability change [34]. I-Mutant 2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability change upon single amino acid substitution. The protein stability change was predicted from the EPHA2 protein sequence (NP_004422). The software computed the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus unfolding Gibbs free energy value of the native protein (kcal/mol). A positive DDG value indicates that the mutated protein possess high stability and vice versa.

Analysis of functional effect on protein

There are many web-based resources available that allow one to predict whether nonsynonymous coding SNPs may have functional effects on proteins. We chose SIFT [35] to perform protein conservation analysis and predict the phenotypic effect of amino acid substitutions. SIFT is based on the premise that protein evolution is correlated with protein function. Variants that occur at conserved alignment positions are expected to be tolerated less than those that occur at diverse positions. The algorithm uses a modified version of PSIBLAST [36] and Dirichlet mixture regularization [37] to construct a multiple sequence alignment of proteins that can be globally aligned to the query sequence and belong to the same clade. The underlying principle of this program is that it generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from zero to one. SIFT scores [38] were classified as intolerant (0.00-0.05), potentially intolerant (0.051-0.10), borderline (0.101-0.20), or tolerant (0.201-1.00). Higher the tolerance index of a particular amino acid substitution, lesser is its likely impact.

Simulation for functional change in coding nsSNPs

PolyPhen [39] is a computational tool for identification of potentially functional nsSNPs. Predictions are based on a combination of phylogenetic, structural and sequence annotation information characterizing a substitution and its position in the protein. For a given amino acid variation, PolyPhen performs several steps: (a) extraction of sequence-based features of the substitution site from the UniProt database, (b) calculation of profile scores for two amino acid variants, (c) calculation of structural parameters and contacts of a substituted residue. PolyPhen scores were classified as 'benign' or 'probably damaging' [38]. Input options for the PolyPhen server are protein sequence or accession number together with sequence position with two amino acid variants. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. PolyPhen searches for three-dimensional protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases. Then, it calculates position-specific independent counts (PSIC) scores for each of two variants, and computes the difference of the PSIC scores of the two variants. The higher a PSIC score difference, the higher functional impact a particular amino acid

substitution is likely to have. PolyPhen scores of >2.0 are expected to be "Probably damaging" to protein structure and function and PolyPhen scores of 1.99-1.50 are expected to be "Possibly damaging" to protein function [33].

Analysis of functional nsSNPs by FASTSNP

The Functional Analysis and Selection Tool for Single Nucleotide Polymorphism (FASTSNP) is a web server which connects many softwares and databases for processing analysis [40]. We used FASTSNP for the prediction of the functional effect of nsSNPs and an estimation of their risk score. The FASTSNP uses a decision tree for prioritizing the functional effect and estimating risk score. The nsSNPs were submitted for FASTSNP analysis and output files were displayed as a decision tree.

Total energy and RMSD calculations

Structure analysis was performed for evaluating the structural stability of native and mutant protein. We used the web resource SAAPdb [41] to identify the protein coded by EPHA2 gene (PDB id 1MQB). The mutation position and residue were in complete agreement with the results obtained with SIFT and PolyPhen programs. The mutation was performed by using SWISSPDB viewer and energy minimization for 3D structures was performed by NOMAD-Ref server [42]. This server use Gromacs as default forcefield for energy minimization based on the methods of steepest descent, conjugate gradient and L-BFGS methods [43]. We used conjugate gradient method for optimizing the 3D structures. Deviation between the two structures is evaluated by their RMSD values. Divergence in mutant structure with native structure is due to mutation, deletions, and insertions [44] and the deviation between the two structures is evaluated by their RMSD values which could affect stability and functional activity [45].

Results

SNP dataset

The EPHA2 gene investigated in this work contained a total of 181 SNPs, of which 16 were nsSNPs, 9 were synonymous SNPs, and none were in noncoding regions. The rest were in the intron region. We selected nonsynonymous coding SNPs for our investigation **Table 1 (see supplementary material)**.

Stability of nsSNPs

More negative is the DDG value, less stable the given point mutation is likely to be, as predicted by I-Mutant 2.0 server. We obtained 8 nsSNPs that were found to be less stable by this server as shown in Table 1. Out of 8 nsSNPs, 6 nsSNPs, namely rs11543934, rs2291806, rs1058371, rs1058370, rs79100278 and rs113882203 showed a DDG value of > -1.0 . The remaining nsSNPs showed a DDG value of < -1.0 , as depicted in **Table 1**. Since these six nsSNPs showed higher DDG value, we considered these nsSNPs to be least stable and deleterious.

Prediction of deleterious nsSNPs by SIFT

Sixteen nsSNPs retrieved from EPHA2 gene, submitted independently to the SIFT showed 8 nsSNPs to be deleterious, having the tolerance index score of ≤ 0.05 . The results are shown in **Table 1 (see supplementary material)**. We observed that, out of 8 deleterious nsSNPs, 6 nsSNPs showed a highly deleterious tolerance index score of 0.00 **Table 1 (see supplementary material)**. We found that among these 6 nsSNPs that are seen to

be deleterious according to SIFT, 4 nsSNPs namely rs2291806, rs1058370, rs79100278 and rs113882203 were also found less stable by I-Mutant 2.0 server.

Damaged nsSNPs by PolyPhen

To identify the EPHA2 nsSNPs that affected protein structure, the EPHA2 nsSNPs were analyzed for predicting a possible impact of amino acids on the structure and function of the protein using PolyPhen server. The EPHA2 protein sequence (NP_004422) with each nsSNP position and their 2 amino acid variants was submitted as input for analyzing the protein structural change due to amino acids. Our result showed, 4 nsSNPs namely rs11543934, rs2291806, rs1058370 and rs113882203 to be probably damaging with PSIC score of ≥ 2.0 . The nsSNPs namely, rs2291806, rs1058370 and rs113882203 which were observed to be the cause of protein less stability by I-Mutant 2.0 server and SIFT were also predicted to be probably damaging by PolyPhen server. In addition two nsSNPs were highly predicted as possibly damaging and the remaining as benign **Table 1 (see supplementary material)**.

Functional effect and estimated risk of nsSNPs

In order to identify nsSNP with a high possibility of having a functional effect, FASTSNP was applied for the detection of nsSNP influence on cellular and molecular biological function e.g. transcriptional and splicing regulation. In addition the estimation of risk score was also calculated by FASTSNP. The functional effect and estimated risk of EPHA2 nsSNPs are shown in **Table 2 (see supplementary material)**. Three nsSNPs exhibited medium-high risk score (risk score = 3-4). The functional nsSNPs were rs11543934, rs2291806 and rs1058370. The eight nsSNPs showed low-medium risk (risk score=2-3). The risk score of the remaining four nsSNPs was not predicted. The most important finding detected by FASTSNP were the two nsSNPs namely rs2291806 and rs1058370 to have high possible functional effect. These were also found polymorphic by I-Mutant 2.0; SIFT as well as by PolyPhen server.

Structural modeling of mutant protein

Mapping the deleterious nsSNPs into protein structure information was obtained from Single Amino Acid Polymorphism database (SAAPdb) [19]. The available structure for EPHA2 gene is reported to be having a PDB id 1MQB. According to this resource, the mutations mainly occurred for 1MQB at one SNP id, namely rs2291806. The mutation was at the residue position E825K. The mutation for 1MQB at the position 825 was performed by SWISS-PDB viewer to get modeled structures. Then, energy minimizations was performed by NOMAD-Ref server [32] for the native type protein (PDB 1MQB) and the mutant type protein (1MQB) E825K. It can be seen from Table 3 that total energy for the native type structure (PDB 1MQB) and the mutant type structure 1MQB E825K are found to be -7476.263 and -13238.058 kcal/mol, respectively. **Table 3 (see supplementary material)**, also shows that the RMSD values between the native type (1MQB) and the mutant type (1MQB) E825K is 2.36 \AA . Since the RMSD values and total energy after energy minimization are very high for the mutant type structure as compared to the native type structure (1MQB), we may presume that, this mutation causes a significant change in the mutant structure of the protein with respect to the native protein structure. This mutation was also predicted to be

functionally significant based on I-Mutant, SIFT, PolyPhen and FASTSNP results.

Discussion:

Our analysis revealed 16 SNPs as nonsynonymous out of which 6 nsSNPs, namely rs11543934, rs2291806, rs1058371, rs1058370, rs79100278 and rs113882203 were found to be least stable by I-Mutant 2.0 with DDG value of > -1.0 . nsSNPs, namely rs35903225, rs2291806, rs1058372, rs1058370, rs79100278 and rs113882203 showed a highly deleterious tolerance index score of 0.00 by SIFT server. Four nsSNPs namely rs11543934, rs2291806, rs1058370 and rs113882203 were found to be probably damaging with PSIC score of ≥ 2.0 by PolyPhen server. Three nsSNPs namely, rs11543934, rs2291806 and rs1058370 were found to be highly polymorphic with a risk score of 3-4 with a possible effect of Non-conservative change and splicing regulation by FASTSNP. The total energy and RMSD value was higher for the mutant-type structure compared to the native type structure. A major interest in human genetics is to distinguish mutations that are functionally neutral from those that contribute to disease. Amino acid substitutions currently account for approximately half of the known gene lesions responsible for human inherited disease. Therefore, the identification of nsSNPs that affect protein functions and relate to disease is an important task. The effect of many nsSNPs will probably be neutral as natural selection will have removed mutations on essential positions. Assessment of non-neutral SNPs is mainly based on phylogenetic information (i.e. correlation with residue conservation) extended to a certain degree with structural approaches. However, there is increasing evidence that many human disease genes are the result of exonic or noncoding mutations affecting regulatory regions [33, 46]. Much attention has been focused on modeling by different methods the possible phenotypic effect of SNPs that cause amino acid changes, and only recently has interest focused on functional SNPs affecting regulatory regions or the splicing process. Moreover, because of their widespread distribution on the species genome, SNPs become particularly important and valuable as genetic makers in the research for the diseases and corresponding drug [47]. Currently, millions of human SNPs have reported by high-throughput methods. The vast number of SNPs causes a challenge for biologists and bioinformaticians although they provide lot information about the relationships between individuals [47]. Besides numerous ongoing efforts to identify millions of these SNPs, there is now also a focus on studying associations between disease risk and these genetic variations using a molecular epidemiological approach. This plethora of SNPs points out a major difficulty faced by scientists in planning costly population-based genotyping, which is to choose target SNPs that are most likely to affect phenotypic functions and ultimately contribute to disease development [33, 47, 48].

Currently, most molecular studies are focusing on SNPs located in coding and regulatory regions, yet many of these studies have been unable to detect significant associations between SNPs and disease susceptibility. To develop a coherent approach for prioritizing SNP selection for genotyping in molecular studies, an evolutionary perspective to SNP screening is applied. The hypothesis is that, amino acids conserved across species are more likely to be functionally significant. Therefore, SNPs that change these amino acids

might be more likely to be associated with disease susceptibility. It is becoming clear that application of the molecular evolutionary approach may be a powerful tool for prioritizing SNPs to be genotyped in future molecular epidemiological studies [33, 48]. Therefore, our analysis will provide useful information in selecting SNPs of EPHA2 gene that is likely to have potential functional impact.

Conclusion:

In our analysis, we found out that nsSNP (rs2291806) as less stable, deleterious, probably damaging and have high risk score. The mutant protein structures of this nsSNP also showed very high energy and RMSD values compared to the native type structure. We therefore concluded that this nsSNP as potentially functional polymorphic. To those conducting large-scale population-based epidemiologic studies, the idea of prioritizing nsSNPs in the investigation of association of SNPs with disease risk is of great interest. The use of these servers to select potentially polymorphic nsSNPs for epidemiology studies can be an efficient way to explore the role of genetic variation in disease risk and to curtail cost. Furthermore, predicted impact of these nsSNPs can be tested with the use of animal models or cell lines to determine if functionality of the protein has indeed been altered.

References:

- [1] <http://www.who.int/blindness/Vision2020%20-report.pdf>
- [2] Foster A & Resnikoff S, *Eye*. 2005 **19**: 1133 [PMID: 16304595]
- [3] Zuercher J *et al. Invest Ophthalmol Vis Sci*. 2010 **51**: 3354 [PMID: 20181839]
- [4] Li T *et al. Biol Trace Elem Res*. 2009 **128**: 1 [PMID: 18972073]
- [5] Abraham AG *et al. Ophthalmol Clin North Am*. 2006 **19**: 415 [PMID: 17067897]
- [6] Erie JC, *J Cataract Refract Surg*. 2007 **33**: 1273 [PMID: 17586386]
- [7] McCarty CA & Taylor HR, *Ophthalmol Vis Sci*. 2001 **42**: 1677 [PMID: 11431427]
- [8] Shiels A *et al. Mol Vis*. 2008 **14**: 2042 [PMID: 19005574]
- [9] Jun G, *PLoS Genet*. 2009 **5**: e1000584 [PMID: 19649315]
- [10] Zhang T *et al. Hum Mutat*. 2009 **30**: E603 [PMID: 19306328]
- [11] Herath NI, *BMC Cancer*. 2006 **6**: 144. [PMID: 16737551]
- [12] Iyengar SK *et al. Proc Natl Acad Sci U S A*. 2004 **101**: 14485 [PMID: 15452352]
- [13] Collins FS, *Genome Res*. 1998 **8**: 1229 [PMID: 9872978]
- [14] Sachidanandam R *et al. Nature*. 2001 **409**: 928 [PMID: 11237013]
- [15] Stenson PD, *Human Mutat*. 2003 **21**: 577 [PMID: 12754702]
- [16] Chasman D & Adams RM, *J Mol Biol*. 2001 **307**: 683 [PMID: 11254390]
- [17] Dryja TP *et al. N Engl J Med*. 1990 **323**: 1302 [PMID: 2215617]
- [18] Smith EP *et al. N Engl J Med*. 1994 **331**: 1056 [PMID: 8090165]
- [19] Yue P & Moulton J, *J Mol Biol*. 2006 **356**: 1263 [PMID: 16412461]
- [20] Ramensky V *et al. Nucleic Acids Res*. 2002 **30**: 3894 [PMID: 12202775]
- [21] Wang Z & Moulton J, *Hum Mutat*. 2001 **17**: 263 [PMID: 11295823]
- [22] Barroso I *et al. Nature*. 1999 **402**: 880 [PMID: 10622252]
- [23] Thomas R, *Am J Hum Genet*. 1999 **65**: 39 [PMID: 10364515]

- [24] Yoshida A *et al.* *Proc Natl Acad Sci U S A.* 1984 **81**: 258 [PMID: 6582480]
- [25] Jaruzelska J, *J Biol Chem.* 1995 **270**: 20370 [PMID: 7657610]
- [26] Ng PC & Henikoff S, *Genome Res.* 2002 **12**: 436 [PMID: 11875032]
- [27] Bell J, *Nature.* 2004 **429**: 453 [PMID: 15164070]
- [28] Fredman D *et al.* *Nucleic Acids Res.* 2002 **30**: 387 [PMID: 11752345]
- [29] Sherry ST, *Nucl Acids Res.* 2001 **29**: 308 [PMID: 11125122]
- [30] Yip YL, *Human Mutat.* 2004 **23**: 464 [PMID: 15108278]
- [31] Johnson MM, *Cancer Epidemiol Biomarkers Prev.* 2005 **14**: 1326 [PMID: 15894696]
- [32] Rajasekaran R *et al.* *Genomics.* 2007 **90**: 447 [PMID: 17719744]
- [33] Doss CG & Sethumadhavan R, *J Biomed Sci.* 2009 **16**: 42 [PMID: 19389263]
- [34] Capriotti E, *Nucleic Acids Res.* 2005 **33**: W306 [PMID: 15980478]
- [35] Ng PC, *Genome Res.* 2001 **11**: 863 [PMID: 11337480]
- [36] Altschul SF *et al.* *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
- [37] Sjolander K, *Comput Appl Biosci.* 1996 **12**: 327 [PMID: 8902360]
- [38] Xi T, *Genomics.* 2004 **83**: 970 [PMID: 15177551]
- [39] Ramensky V *et al.* *Nucleic Acids Res.* 2002 **30**: 3894 [PMID: 12202775]
- [40] Yuan HY *et al.* *Nucleic Acids Res.* 2006 **34**: W635 [PMID: 16845089]
- [41] Cavallo A & Martin AC, *Bioinformatics.* 2005 **21**: 1443 [PMID: 15613399]
- [42] Lindahl E, *Nucleic Acids Res.* 2006 **34**: W52 [PMID: 16845062]
- [43] Delarue M & Dumas P, *Proc Natl Acad Sci U S A.* 2004 **101**: 6957 [PMID: 15096585]
- [44] Han JH, *Structure.* 2006 **14**: 935 [PMID: 16698554]
- [45] Varfolomeev SD *et al.* *Biochemistry (Mosc).* 2002 **67**: 1099 [PMID: 12460108]
- [46] Ng PC & Henikoff S, *Genome Res.* 2002 **12**: 436 [PMID: 11875032]
- [47] George Priya Doss C *et al.* *Genomic Med.* 2008 **2**: 23 [PMID: 18716917]
- [48] Zhu Y, *Cancer Res.* 2004 **64**: 2251 [PMID: 15026370]

Edited by P Kanguane

Citation: Masoodi *et al.* *Bioinformation* 8(12): 562-567 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: I-Mutant, SIFT and PolyPhen results of EPHA2

SNP ids	Alleles	Amino acid Change	DDG	SIFT		PolyPhen	
				Tolerance Index	Predicted Impact	PSIC Score	Predicted Impact
rs116506614	C/T	R721Q	0.43	0.09	Potentially Intolerant	0.965	Benign
rs56198600	C/T	R568H	0.72	0.10	Potentially Intolerant	0.787	Benign
rs55747232	A/G	T511M	0.68	0.10	Potentially Intolerant	0.943	Benign
rs35903225	C/T	R876H	-0.97	0.00	Intolerant	1.081	Borderline
rs35484156	A/G	S277L	0.65	0.08	Potentially Intolerant	0.635	Benign
rs34192549	C/T	G391R	0.40	0.09	Potentially Intolerant	0.866	Benign
rs34021505	A/G	M631T	0.10	0.30	Tolerant	0.191	Benign
rs11543934	A/C	P350T	-1.22	0.01	Intolerant	2.209	Probably damaging
rs2291806	A/G	E825K	-2.68	0.00	Intolerant	2.666	Possibly damaging
rs1058372	C/G	K99N	-0.46	0.00	Intolerant	1.855	Possibly damaging
rs1058371	A/T	I96F	-1.59	0.01	Intolerant	1.028	Borderline
rs1058370	A/T	I94N	-2.57	0.00	Intolerant	2.535	Probably damaging
rs79100278	A/C	V255G	-1.65	0.00	Intolerant	1.706	Possibly damaging
rs113882203	C/T	R159H	-1.09	0.00	Intolerant	2.309	
rs115171763	C/T	D496N	0.80	0.29	Tolerant	0.143	Benign
rs116976705	A/G	I458T	0.57	0.20	Tolerant	0.047	Benign

Table 2: Functional Effect and Estimated Risk (FASTSNP).

SNP ids	Alleles	Amino acid	Possible Effect	Risk Score
rs35903225	C/T	R721Q	-NP-	
rs56198600	C/T	R568H	Conservative change	2-3
rs55747232	A/G	T511M	Conservative change, splicing regulation	2-3
rs35903225	C/T	R876H	Conservative change	2-3
rs35484156	A/G	S277L	Conservative change, splicing regulation	2-3
rs34192549	C/T	G391R	Conservative change	2-3
rs34021505	A/G	M631T	Conservative change	2-3
rs11543934	A/C	P350T	Non-conservative change, splicing regulation	3-4
rs2291806	A/G	E825K	Splicing site	3-4
rs1058372	C/G	K99N	Conservative change, splicing regulation	2-3
rs1058371	A/T	I96F	Conservative change, splicing regulation	2-3
rs1058370	A/T	I94N	Non-conservative change, splicing regulation	3-4
rs79100278	A/C	V255G	-NP-	
rs113882203	C/T	R159H	-NP-	
rs115171763	C/T	D496N	-NP-	
rs116976705	A/G	I458T	-NP-	

NP- No Prediction; nsSNPs which show high risk score are highlighted as bold

Table 3: RMSD and total energy of native-structure (1mqb) and its mutant model

	1mqb native-type structure	1mqb 825 mutant (rs229 1806) with 1mqb native-type structure
RMSD of entire structure	-	2.36 Å
Total energy after energy minimization	-7476.263 Kcal/mol	-13238.058 Kcal/mol