Pakistan J. Zool., vol. 54(5), pp 2315-2327, 2022 DOI: https://dx.doi.org/10.17582/journal.pjz/20210713140738

Mining the SNPs of Human Low Density Lipoprotein (LDL) related Gene *APOB* through *in silico* Approaches

Muneeza Zafar[1,2,3], Fazli Rabbi Awan[2,*], Munazza Raza Mirza[3,*], Sumaira Nishat[2,4], Sajid Ali Rajput[5] and Imran Riaz Malik[1,*]

[1]*Department of Biotechnology, University of Sargodha, Sargodha*
[2]*Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Jhang Road, P.O. Box. 577, Faisalabad*
[3]*Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270*
[4]*Department of Computer Science, University of Agriculture, Faisalabad*
[5]*Institute of Biotechnology and Genetie Engineering, University of Sindh, Jamshoro*

Article Information
Received 13 July 2021
Revised 10 August 2021
Accepted 16 August 2021
Available online 07 December 2021 (early access)
Published 20 June 2022

Authors' Contribution
MZ, FRA and MRM conceptualized the study. MZ and SN performed data curation. MZ, FRA, IRM, MRM, SN and SAR performed the formal analysis. MZ, FRA, IRM, MRM, and SAR were involved in investigation. MZ and SN worked on methodology. MZ, FRA, IRM and MRM did the project administration and supervision. FRA, IRM, SN and MRM arranged the resources. MZ, MRM, FRA, IRM and SAR performed validation of results and data visualization. MZ wrote the original draft. MZ, MRM, FRA, IRM and SAR reviewed and edited the manuscript.

Key words
***APOB* gene, Low density lipoprotein, Single nucleotide polymorphism, SNPs, LDL, Apolipoproteins, VLDL.**

ABSTRACT

Apolipoprotein B (APOB) is the major part of low density lipoprotein (LDL), with two major isoforms: APOB100 and APOB48 found in the human body. Both isoforms are involved in the formation and transport of chylomicron and LDL-cholesterol. Point mutations in *APOB* may lead to change in protein stereochemistry, which may result in premature coronary artery disease, familial hypobetalipoproteinemia, hypocholesterolemia, mono-genic dyslipisimias and other atherogenic events in CVD. Here we evaluated the impact of all missense and non-coding single nucleotide polymorphisms (SNPs) of *APOB* retrieved from dbSNP using 17 different computational tools and further evaluated the structural impact of these convergent deleterious SNPs on APOB through HOPE. We found 9 missenses, 15 intronic or regulatory region SNPs and 2 were found in miRNA target sites of *APOB*. Out of these variant, the rs13306194 (Arg532Trp) was found in the conserved region of protein domain, which can potentially disrupt overall chemical structure and function of the APOB. Six missense SNPs in the coding, and 17 SNPs in non-coding regions are proposed as novel most deleterious variants of *APOB*. We also try to predict the structural model of APOB through protein docking. The results indicate the applicability of *in silico* approach to propose the most deleterious SNPs of *APOB* that should be prioritize for future genetic association studies in cohort of cardiovascular patients. While their structural impact on APOB may suggest these predicted nsSNPs possibly be a better drug target and contribute to the treatment and better understanding of human cardiovascular disease.

INTRODUCTION

Apolipoproteins (Apo) are the specific lipid binding proteins which act as lipoprotein or lipid transporters in the body and function as receptor ligand, enzyme cofactor and have core importance in lipid metabolism. Human body has several types of apolipoproteins that perform different functions which depend on the type of their attached lipoprotein particle (Liwen *et al*., 2019). These are classified as ApoA, B, C, D, and E. Both ApoA and ApoD compose the high density lipoprotein (HDL). ApoB plays a critical role in the low density lipoprotein (LDL) transport system. Whereas, ApoC has been described as a component of very low density lipoprotein (VLDL) along with ApoE, which is also the major apolipoprotein of chylomicrons.

Apolipoprotein B (ApoB) has a significant importance in human lipoprotein metabolism since it is a primary part of LDL and chylomicrons. ApoB is a large, non-exchangeable, amphipathic glycoprotein and its 43 kb gene *APOB* is located on the short arm of human chromosome 2 having 29 exons. The *APOB* has two discrete circulating isoforms, including apoB-48 (215 amino acids; 48% identity with amino terminal) and apoB-100 (4536 amino acids with 100% identity). Both of these isoforms are produced through post-transcriptional mRNA editing process using RNA-specific cytidine deaminase enzyme named as apoB mRNA editing enzyme catalytic complex 1 or C->U-editing enzyme apobec-1 (Nordestgaard *et al*., 2020). The synthesis and assembly

* Corresponding author: munazzaraza@iccs.com
0030-9923/2022/0005-2315 $ 9.00/0

of chylomicron or VLDL are influenced by the type of isoform produced through apobec-1 editing process. Furthermore, these apoB species are produced in different tissues and perform different functions in the body. ApoB-48 is produced in enterocytes of small intestine and required for chylomicron formation. The main function of chylomicrons is to transport triglycerides from the intestine to the liver, adipose, and muscle tissue. ApoB-100 is an essential structural component of VLDL and its metabolic products. VLDL is predominantly filled with triglycerides and its hydrolysis by lipoprotein lipase yields intermediate density lipoprotein (IDL). In the next step, IDL (VLDL remnant) is either cleared from the circulation through its hepatic remnant receptors or hydrolyze further by hepatic lipase and yields LDL. The resultant LDL is reduced in size as compared to its precursor VLDL particle and cleared from blood after binding with LDL receptor in the liver. Reduced secretion of apoB results in decreased production of chylomicron and VLDL, which ultimately leads to malabsorption of fats and fat-soluble vitamins. apoB containing lipoproteins are pivotal for lipid absorption and triglyceride homeostasis, their enhanced levels in plasma induce atherosclerosis. Subendothelial retention of ApoB containing lipoproteins is a critical event in the development of atherogenesis. High plasma levels of ApoB and LDL-cholesterol are risk factors for atherosclerosis, whereas low levels of ApoB may provide protection against the development of atherosclerosis (Navarese *et al*., 2018). Experimental studies suggest that 50–60% of the variation in plasma levels of ApoB is genetically determined (Wang *et al*., 2018).

In addition to its structural role, apoB-100 is a ligand for receptor-mediated endocytosis of LDL. Essentially all circulating ApoB are associated with lipoproteins, and unlike most other apolipoproteins, ApoB cannot exchange freely among lipoprotein particles. Increased plasma concentrations of ApoB-containing lipoproteins have been demonstrated to be key risk factors for the development of atherosclerosis. Furthermore, missense mutations in the LDL-receptor binding domain of ApoB may cause familial ligand-defective ApoB-100 characterized by hypercholesterolemia and premature coronary artery disease. Other mutations in *APOB* can cause familial hypobetalipoproteinemia, characterized by hypocholesterolemia and resistance to atherosclerosis. These naturally occurring mutations reveal key domains in ApoB and demonstrate how monogenic dyslipidemia can provide insight into biologically important mechanisms that may lead to complex conditions, such as atherosclerosis.

SNPs are the simplest form of genetic variations and source of 90% of variations reported in human population. These can be of many types including synonymous SNPs, non-synonymous SNPs (nsSNPs) as well as 3′UTR, 5′UTR and intronic variants. It is likely that nsSNPs play important role in the functional diversity of encoded proteins and have been linked with many disease conditions (Burton *et al*., 2007; Joshi *et al*., 2015). These SNPs may affect protein function by reducing protein solubility or by destabilizing protein structure. The other variants in promoter or intronic regions may affect gene regulation by altering transcription and subsequently translation through altered transcription factor binding sites or splicing sites.

In large population-based studies, the analysis of all the genetic variants is a challenging task due to increased cost, complexity and time consumption. Recent studies have revealed that all reported genetic variants may or may not cause susceptibility to the disease. Some of these may be involved genotypically and/or phenotypically. Mining functional SNPs in the given plethora of SNPs is important for the structural and functional studies of genes and their products. Taking into account all these considerations, the present study was undertaken to extract and prioritize various *APOB* variants and study their effects on structure and function of ApoB100 using different computational/bioinformatics tools and algorithms and hence narrow down the functional SNPs strongly involved in the pathogenicity of cardiovascular disorders.

MATERIALS AND METHODS

Data retrieval

The data on human *APOB* was retrieved from Entrez Gene from National Center for Biological Information (NCBI) database. The SNP information (reference sequence ID) and protein sequence (accession number) of the *APOB* were retrieved from NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/) SwissProt (http://expasy.org/) databases, respectively. The criteria used for selection of SNPs was based on at least one of the following condition (i) It should be sequenced in the 1000 Genomes Project phase I (http://www.1000genomes.org); (ii) it has minor alleles observed in at least two chromosomes; and (iii) it has multiple, independent submissions to the refSNP cluster. The variation class used for SNPs selection were included the missense class, Intronic, 3′-UTR, and 5′-UTR. The cytogenetic location and the transcript details were obtained from Online Mendelian Inheritance in Man (OMIM) and Ensembl databases.

Pathogenicity testing of missense SNPs

After the data mining and extracting the desired missense SNPs information, functional analysis and pathogenicity testing was done through 16 different bioinformatics tools. These tools were divided into 4 categories based on sequence, supervised learning-based, structure and consensus-based methods. The retrieved missense SNPs were filtered through each method by

using the criteria of predicted deleterious by at least half numbers of tools in each group.

Sequence homology-based methods

This method used sequence homology principles to predict whether an amino acid substitution will affect protein function or not and included tools, such as SIFT (https://sift.bii.a-star.edu.sg/) (Reva *et al*., 2011), PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org/index.php) (Choi and Chan, 2015), Mutation assessor (http://mutationassessor.org/r3/) (Hepp *et al*., 2015), PON-P2 (http://structure.bmc.lu.se/PON-P2/) (Niroula *et al*., 2015), and PhD-SNP (Predictor of human deleterious SNPs) (https://snps.biofold.org/phd-snp/phd-snp.html) (Mah *et al*., 2011).

Supervised learning methods

These algorithms used for prediction of missense SNPs by using neural networks: SNAP (http://www.rostlab.org/services/) (Mah *et al*., 2011), and support vector machines: MutPred 2 (http://mutpred.mutdb.org/) (Pejaver *et al*., 2017) and SuSPect (http://www.sbg.bio.ic.ac.uk/suspect/download.html) (Pires *et al*., 2017).

Protein sequence and structure-based methods

The following methods used either combine information from protein sequence and structure or used only protein structural information to analyze missense variants. These included PolyPhen (Polymorphism phenotyping) (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al*., 2010) and I-Mutant3 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/IMutant3.0.cgi) (Capriotti *et al*., 2005).

Consensus based methods

This method provide the effect of mutation on the protein biological activity by using consensus score generated through various tools, it included Condel (CON comes from "consensus" and DEL for "deleterious") (https://sourceforge.net/projects/condel/) (González-Pérez and López-Bigas, 2011), MetaSNP (https://snps.biofold.org/meta-snp/) (Capriotti *et al*., 2013), and PredictSNP (https://loschmidt.chemi.muni.cz/predictsnp/) (Bendl *et al*., 2014).

Evolutionary conservation based analysis

PANTHER (Protein Analysis through Evolutionary Relationships) (http://www.pantherdb.org/tools/) is an online widely used tool for comprehensive evolutionary and functional classification of protein (Tang and Thomas, 2016). The classification of proteins is based on their molecular function, protein-protein interactions and evolutionary relationships with outcome score is presented as subPSEC (Substitution Position-Specific Evolutionary Conservation Score).

Functional analysis of noncoding region SNPs

PolymiRTs (Polymorphism in MicroRNAs and their Target Sites) (http://compbio.uthsc.edu/miRSNP/) was used to predict naturally present SNPs in microRNA seed regions and miRNA target sites (Chirumbolo, 2016).

Regulome DB (https://regulomedb.org/regulome-search/) is a prediction tool to prioritize as well as annotate potential regulatory variants from human genome. The database includes 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, large collection of Expression quantitative trait loci, dsQTL, and ChIP-exo data to identify putative regulatory variants (Boyle *et al*., 2012).

SNPinfo (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html) SNPinfo server is a set of web-based various selection tools including Gene pipe, Genome pipe, Linkage pipe, Taq SNP, Func Pred, SNPseq, which were used to select functional coding and non-coding SNPs for genetic association studies (http://snpinfo.niehs.nih.gov/) (Xu and Taylor, 2009). The details of number of tools used in each method with their working principle and prediction score criteria are mentioned in Table I.

Table I.- Evolutionary analysis of all the retrieved deleterious missense nsSNPs.

SNP ID	PANTHER	
	Score (Million years)	Prediction
rs676210	750	Probably damaging
rs13306194	750	Probably damaging
rs533617	750	Probably damaging
rs41288783	910	Probably damaging
rs544542990	750	Probably damaging
rs72653074	750	Probably damaging
rs181737266	750	Probably damaging
rs536328155	750	Probably damaging
rs540387864	750	Probably damaging

Structural impact of deleterious SNPs

To analysis the effect of deleterious SNPs on protein structure HOPE (Have Your Protein Explained) (https://www3.cmbi.umcn.nl/hope/) was used. It acts as automatic mutant analysis server which can generate the both mutant and wild type models of the interested protein with their change residues. Furthermore, it collects structural information from 3D protein structure, UniProt sequence annotations and Reproof software prediction (Venselaar *et al*., 2010; Rost, 2001).

Docking simulation of APOB

For prediction of protein structure, I-TASSER (Iterative Threading ASSEmbly Refinement) and UCSF Chimera tools were used. I-TASSER predicts best model using TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library and RMSD value that are residues aligned by TM-align (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Grillo *et al*., 2010). UCSF Chimera tool allows the merging of different structures into a single model using copy/combine feature (https://www.cgl.ucsf.edu/chimera/about.html) (Kaur *et al*., 2017).

The details of number of tools used in each method with their working principle and prediction score criteria are mentioned in Supplementary Table I.

RESULTS

Prediction of pathogenic missense SNPs of APOB

A total of 473 SNPs of *APOB* were selected which fulfilled the selection criteria using the dbSNP of NCBI, UniprotKB, GeneCards and Ensembl databases (Supplementary Table II). Out of these, 63% (n = 297) SNPs belonged to missense class, 36% (n = 171) were from intronic region, 1% (n = 4) from 3′-UTR and 0% (n = 1) belong to the selection class of 5′-UTR, respectively (Fig. 1A).

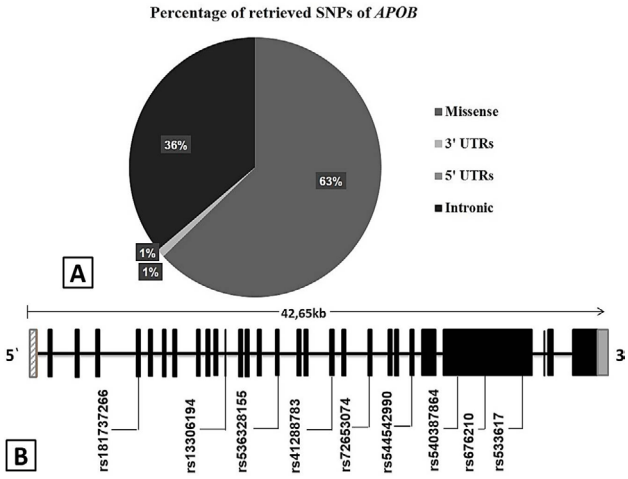

Fig. 1. **A**, Pie chart of retrieved validated SNPs of *APOB* from NCBI and Ensembl data bases. It includes 295 (62%) missense, 170 (36%) intronic, 4 (1%) 3′ UTR and 5 (1%) 5′-UTR. **B**, Convergent deleterious and functionally important SNPs are located in distinct exonic region of *APOB* gene. The 3’ and 5’ un-translated regions are represented by hatched bars and the exons are represented by filled bars. The *APOB* amino acid position is relative to Gene Bank Accession number NC_000002. 12.

We employed 16 different tools including SIFT, PROVEAN, Mutation Assessor, PON-P2, PhD-SNP, SNAP, SuSPect, PolyPhen, I-Mutant, MutPred, Condel, MetaSNP. PredictSNP, PANTHER for missense SNPs, while in case of 3′-UTR, 5′-UTR and coding synonymous SNPs PolymiRTs along with RegulomeDB were used. Furthermore, to analyze the effect of SNPs on protein structure, HOPE was used. In this study, on the basis of their working methodology, these tools were categorized in 5 groups; protein sequence-based, structure-based, supervised learning, evolutionary and consensus-based. All the retrieved missense SNPs were sequentially passed through these tools for pathogenicity testing and picked out on 50 % of selection criteria.

In the first category of sequence-based analysis, SIFT showed 46% (n =132) SNPs as damaging (DAM) having scored > 0.05 while all the remaining SNPs have scored < 0.05 and were in tolerable (TOL) range. Similarly, PROVEAN showed 36% (n=105) SNPs as “Deleterious” and 62% (n =180) as “Neutral”. Four SNPs showed both neutral and deleterious effect due to being multi-allelic in nature. In contrast to this, 162 SNPs were found to have effect as Med/High (deleterious) while all others were found to have low or neutral effect by Mutation - assessor. The pathogenicity testing of single amino acid substitution was also checked by PON-P2 and Phd-SNP. According to PON-P2 analysis, 152 variants were falling under the Pathogenic class while 120 SNPs have showed Neutral effect and the remaining were unknown to the software. Furthermore, the reliability index (RI) of Phd-SNP was ≤ 0.5 for 166 (58%) SNPs, and ≥ 0.5 for all the remaining variants. We shortlisted 92 SNPs as deleterious that were predicted by at least three of the above-mentioned tools in a sequence - based category, and were subjected to analysis by the next category. The detailed distribution of deleterious SNPs predicted by each tool is given in the Supplementary Table III.

The second category was supervised learning - based analysis, which was carried out by using tools including; Suspect, SNAP and Mutpred2. Out of 92 extracted missense variants from previously mentioned category, 37 (41%) were picked out as deleterious which were present in at least two tools out of three in this category, while remaining variants were found to be unaffected. Both Mutpred2 and Suspect showed maximum score of 0.92 (≥ 0.5) for rs372035579 and 90 (between+1 to +100) for rs676210, respectively. While SNAPS predicted 25 such variants having deleterious effect (EFF) and remaining were neutral (NEU). In the next step, these filtered deleterious variants were passed through from the third category of tools based on structure including, Polyphen and I-mutant 3.0. The PCSI score of Polyphen was 1

(possibly damaging) for 15 variants and ranges between 0 - 1 (probably damaging) for remaining 12 variants.

Prediction of missense SNPs on the base of protein stability

I-Mutant 3.0 was used to analyze the effect of missense SNPs on protein stability in terms of Gibbs free energy or ΔΔG values. According to the prediction via a Ternary classificator of I-Mutant 3.0, we found 27 (73 %) out of total 37 SNPs to be predicted as "Unstable" (ΔΔG < -0.5) with highest score of -2.31 was showed by rs561774487 which was highly unstable, while remaining 10 variants were predicted to be "Neutral" with $-0.5 \leq \Delta\Delta G \geq 0.5$. In our study, no SNP was predicted to have "Stable" effect (ΔΔG > 0.5) on protein. We extracted 20 SNPs with deleterious prediction by comparing the scores of both Polyphen and I-mutant 3.0. These selected 20 SNPs were then subjected to last consensus – based category including Condel, Meta-SNP and Predict-SNP. All of these SNPs were predicted a "Disease Causing" and found to have deleterious effect by all the above mentioned tools of this category.

Prediction of missense SNPs on the base of evolution

In terms of evolutionary analysis of these missense nsSNPs, we used PANTHER–PSCEP (position-specific evolutionary preservation) scoring method. In the present study, we found 9 such missense SNPs out of total 20 SNPs that were predicted as "Probably Damaging" with preservation time > 450 million years while maximum preservation score was found for rs41288783 of 910 million years in APOB lineage (Table I). Furthermore, we also used combinatorial approach and found all of these nine extracted missense SNPs *i.e.* rs676210, rs13306194, rs533617, rs41288783, rs544542990, rs72653074, rs181737266, rs536328155, and rs540387864 deleterious by maximum no. of tools used as shown in Table II. The details of all these extracted SNPs with validation status are given in Table III, while their position on *APOB* is presented in Figure 1B.

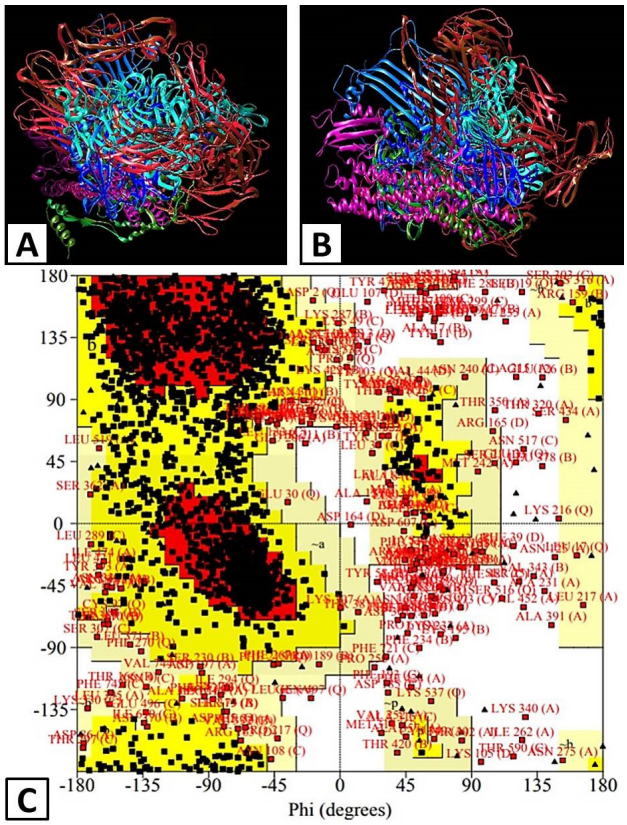

Fig. 2. Predicted Model of APOB protein by using UCSF Chimera. a) Front view of the predicted model. B) Lateral view of the predicted model. C) Ramachandran plot of predicted APOB model by using PROCHECH. Region of A, B and L were considered core regions. a, b, l and p showed allowed region, ~a, ~b, ~l and ~p represents additional allowed region and XX for disallowed region.

Table II.- Extracted deleterious missense nsSNPs and their predicted pathogenic scores by all the in silico tools used.

SNP ID	PhD-SNP	SNAP	Mutpred	Suspect	Condel	Meta-SNP	Pre-dict-SNP	SIFT	PON-P2	PROVEN	Mutation	PolyPhen-2	I-Mutant
rs676210	DIS	NEU	0.652	88, 90	0.308, 0.309	D	DIS	DAM	U	D	MED	PD	-1.18
rs13306194	DIS	EFF	0.828	83	0.528	D	DIS	DAM	U	D	MED	PD	0.03
rs533617	NEU	NEU	0.734	83	0.312	D	DIS	DAM	P	D	MED	PD	0.06
rs41288783	DIS	EFF	0.510	76	0.447	D	DIS	DAM	N	D	MED	PD	-0.77
rs544542990	NEU	NEU	0.803	75	0.687	D	DIS	DAM	U	D	MED	PD	-1.3
rs72653074	DIS	EFF	0.600	65	0.618	D	DIS	DAM	U	D	MED	PD	-0.69
rs181737266	DIS	NEU	0.713	80	0.514	D	DIS	DAM	U	D	MED	PD	-0.34
rs536328155	NEU	EFF	0.668	82	0.483	D	DIS	DAM	U	D	MED	PD	-1.23
rs540387864	NEU	EFF,	0.817	90	0.606, 0.607	D	DIS	DAM	P	D	HIGH	PD	-1.22

Convergent deleterious predicted SNPs analyzed by 13 prediction tools classified in four different groups. DIS, diseases causing; NEU, neutral; EFF, effect on protein structure; D, deleterious; U, unknown; N, neutral; P, pathogenic; MED, medium effect; PD, possibly damaging.

Table III.- Details of extracted nsSNPs of *APOB* proposed for prioritization.

SNP ID	Chromosome position	Nucleotide change	Annotation	AA change	MAF value	Population studied	References
rs676210	chr2:21008652	G>A / G>T	p.Pro2739Leu p.Pro2739Gln	P (Pro) > L (Leu) P (Pro) > Q (Gln)	A=0.29293(71949/245616, GnomA) A=0.29280 (35527/121336, ExAC) A=0.366 (1834/5008, 1000G)	Chinese Yugur population Chinese HAN population	Xiao *et al.* (2017) Gu *et al*. (2017)
rs13306194	chr2:21029662	G>A	p.Arg532Trp	R (Arg) > W (Trp)	A=0.01089 (2679/246014,GnomAD) A=0.01138 (1379/121230, ExAC) A=0.027 (137/5008, 1000G)	Chinese Population	Tang *et al*. (2015)
rs533617	chr2:21011100	T>C	p.His1923Arg	H (His) > R (Arg)	C=0.03179(7823/246094, GnomAD) C=0.03116 (3783/121388, ExAC) C=0.016 (81/5008, 1000G)	Finnish population	Meng *et al*. (1996) Ilmonen *et al*. (1995)
rs41288783	chr2:21019741	G>A	p.Pro994Leu	P (Pro) > L (Leu)	A=0.00048 (117/246060, GnomAD) A=0.00041 (49/120752, ExAC) A=0.001 (5/5008, 1000G)	-	-
rs544542990	chr2:21014461	G>A	p.Leu1277Phe	L (Leu) > F (Phe)	A=0.00013 (32/246166, GnomAD) A=0.00016 (20/121332, ExAC)	-	
rs72653074	chr2:21016551	C>T	p.Gly1074Arg	G (Gly) > R (Arg)	T=0.00002 (5/246236, GnomAD) T=0.00002 (3/121400, ExAC) T=0.0001 (1/13006, GO-ESP)	-	-
rs181737266	chr2:21038061	G>A	p.Pro145Leu	P (Pro) > L (Leu)	A=0.00009 (22/246270, GnomAD) A=0.00010 (12/125568, TOPMED) A=0.00008 (10/121396, ExAC) A=0.00040 (1000 G)	-	-
rs536328155	chr2:21025043	A>G	p.Tyr776His	Y (Tyr) > H (His)	G=0.00001 (3/246202, GnomAD) G=0.00002 (3/121206, ExAC) G=0.000 2(1/5008, 1000G)	-	-
rs540387864	chr2:21007339	A>G / A>T	p.Tyr3177His p.Tyr3177Asn	Y (Tyr) > H (His) Y (Tyr) > N[(Asn)	G=0.000004(1/246002, GnomAD) T=0.000004 (1/246002, GnomAD) G=0.0002 (1/5008, 1000G)	-	-

GnomAD, genome aggregation database; ExAC, exome aggregation consortium; TOPMed, trans-omics for precision medicine; 1000G, 1000 genomes project.

Mutation analysis on native ApoB structure

The exact 3D structure of APOB protein with 4536 amino acids was not available until the drafting of this manuscript. However, by using Yasara and WHAT IF Twinset in HOPE tool, first 46-672 residues - based homologous structure was found from RCSB Protein data bank with PDB ID 1LSH. To check the sequence identity near the position of interest, the template was aligned with query sequence using Protein BLAST and resultant sequence identity was 22.2 %. Out of 9 selected deleterious SNPs, only one SNP rs13306194 (located in exon 12) was harbouring in this homologus model of Lipovitellin with PDB ID ILSH. Both the native and mutant protein models are presented in Figure 2A. The residue change was R (Arginine) > W (Tryptophan) and their detailed structure of amino acid residue change showed bigger in size, neutrally charged, less hydrophobic properties of mutant residue as compared to wild type positively charged residue (Fig. 2B).

Prediction of intronic and UTRs SNPs effecting TFBS

The non-coding regions including intronic and UTR of *APOB* serve as putative binding sites for transcription factors as well as splicing. A single nucleotide change at these positions may alter the binding and subsequently affect transcription or splicing mechanisms. In the present study, using Regulome DB, we found 105 ncSNPs to affect TFBS on the criteria of having minor allele frequency <1% (Supplementary Table III). However, we selected only those ncSNPs which were having Regulome DB scores < 3 as listed in Table IV. Out of these selected 15 non-coding variants, one variant, rs12714268, predicted to have effect on "TF binding + matched TF motif + matched DNase Footprint + DNase peak", with score 2a, ten variants including, rs488329, rs145100968, rs570904180, rs548067874, rs12720840, rs191618417, rs142229577, rs12720797, rs531023775, rs12720762 were predicted to have effect on "TF binding + any motif + DNase Footprint + DNase peak" while remaining four variants rs572186909, rs139313355, rs201106138, rs377355276, rs143452815" were found to be effect on "TF binding + matched TF motif + DNase peak" with score 2c. Furthermore, only one variant rs12720762 out of the above mentioned 15 ncSNPs predicted by Regulome DB was found to have effect on transcription factor binding site (TFBS) with score Y by using SNPinfo tool.

Prediction of putative miRNA target sites

3` UTR serve as putative target sites for miRNA, an important regulator of gene expression. In the current study, we also planned to predict the 3′-UTR of APOB, as a result, we found two such variants rs72654430 and rs142151703 using PolymiRNA. SNP rs72654430 was predicted to mutate into two functional classes as D (disturb the conserved site of the miRNA) with context score of -0.138 and C (create new miRNA site) having context score of -0.242. While in case of rs142151703, it was predicted to disturb the conserved site of miRNA with context score of -0.16 as mentioned in Table IV.

Table IV.- 3′-UTR SNPs in miRNA binding sites of *APOB* analyzed through PolymiRNA.

CHR. location	dbSNP ID	Allele	MIR ID	Cons. score	MIR Site	Functional class	Context score
chr2: 21224373	rs72654430 (T/C)	T	hsa-miR-29a-5p	7	aagaAAATCAGga	D	-0.138
			hsa-miR-378a-5p	5	aagaaaGTCAGGA	C	-0.242
chr2: 21224423	rs142151703 (G/A)	G	hsa-miR-1233-3p	2	AGGGCTCggaagg	D	-0.161

D, disturb the conserved site of the miRNA; C, create new miRNA site; MIR, microRNA.

Table V.- Models predicted by I-Tasser tool along with TM and SC scores.

Predicted models	PDB IDs	TM-score	SC-score	RMSD (Å)
Model 1	1LSH (A chain)	0.904	0.911	0.75
Model 2	4RU5 (A chain)	0.935	0.960	1.65
Model 3	509Z (L chain)	0.884	0.911	1.60
Model 4	4ACQ (A chain)	0.926	0.961	2.18
Model 5	5XBJ (A chain)	0.686	0.958	2.60

TM, template modeling; SC, sequence coverage; RMSD, root mean square deviation.

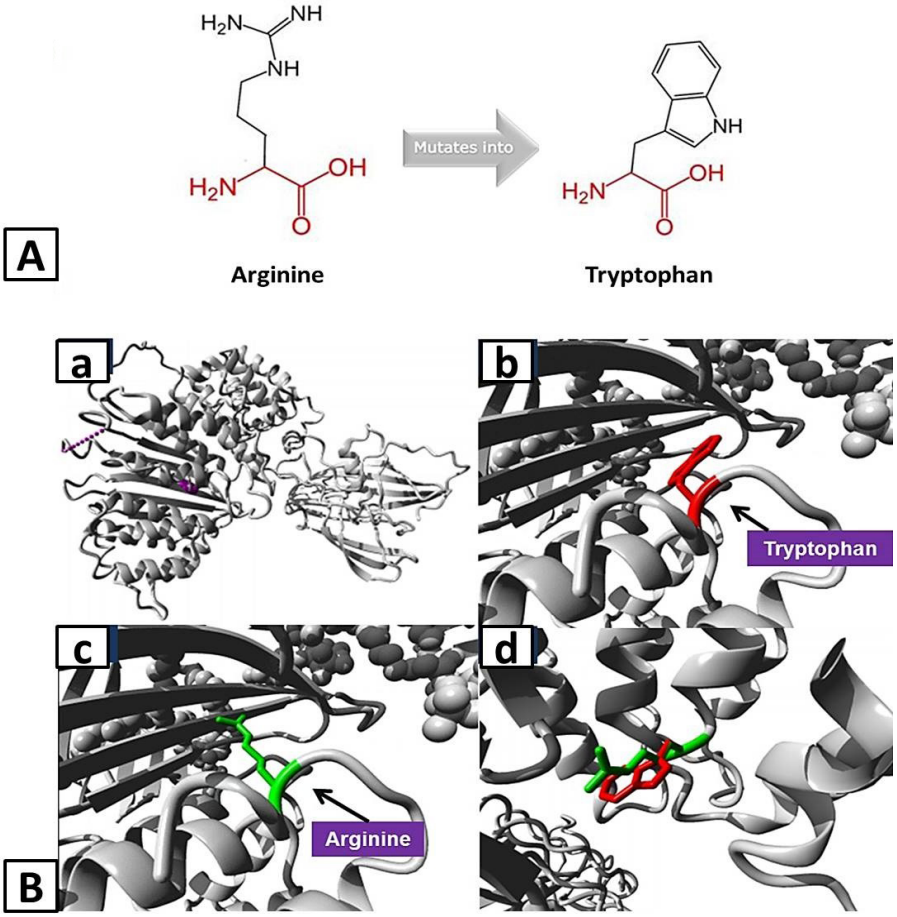

Fig. 3. **A**, Schematic structures of the original (left) Arginine and the mutant (right) amino acid tryptophan for variant Arg532Trp of *APOB* created by HOPE tool. The backbone, which is the same for each amino acid, is colored red and the side chain, unique for each amino acid, is colored black. **B**, Homology models of ApoB representing structural impact of variant Arg532Trp: a, Overview of the protein in ribbon-presentation with protein is colored grey, and the side chain of the mutated residue is colored magenta shown as small balls; b, Close-up of the mutation with protein is colored grey and red represent side chain of mutant residue; c, Close-up of the mutation with protein is colored grey and green represent side chain of wild-type residue; d, Close-up of the mutation with both wild-type and mutant residues side chain on the protein.

Prediction of APOB structure by docking simulation

In present study, we also attempt to predict the 3D structure of APOB by using I-Tasser and Chimera tools. As the APOB is 4563 amino acid long so, the sequence of our protein of interest was splited on assumption with every chain start with methionine. All the sequence chains were submitted and top 5 best match models were predicted by I-Tasser. The details of their template modeling score (TM), Sequence coverage score (SC), root mean square deviation (RMSD) and PDB IDs are mention in Table V. The predicted model 1 have best alignment similarity found with chain A of 1LSH in PDB library with TM score of 0.904 and SC score was 0.911 with RMSD value of 0.75 Å. The predicted model 2 have best alignment similarity found with chain A of 4RU5 in PDB library with TM score 0f 0.935 and CS was 0.960 with RMSD value of 1.65 Å. The predicted model 3 have best alignment similarity found with chain L of 509Z with TM score of 0.884 and SC was 0.911 with RMSD value of 1.60 Å. The predicted model4 have best alignment similarity with chain A of 4ACQ. The TM score was 0.926 that means it cover a good length with its template and SC score was 0.961 with RMSD value of 2.18 Å. The last predicted model 5 have found to best alignment similarity with chain A of 5XBJ in PDB library. Its TM score was 0.686 with SC score was 0.958 with RMSD value of 2.60 Å. All are the predicted model sequences were combined by using Chimera tool with copy/combine feature. The resultant model of APOB protein is shown in Figure 3A and B.

The quality of the predicted model was further

analyzed by PROCHECK. The resultant Ramachandran plot showed 61.1% residues in most favored region representing by A, B and L, 29.3% in additional allowed regions, 6.1% generally allowed region and 3.5% in disallowed regions based on resolution of 2.0 Å and R-factor < 20% as depicted in Figure 3C.

DISCUSSION

To date, the complete mechanisms by which a nucleotide variant may result in a phenotypic change are for the most part unknown. Many human SNPs that are now recognized (in excess of 4-million unique SNPs) (http://www.ncbi.nlm.h.gov/SNP/index.html) along with the genome sequence and other proteome information, provide an opportunity for a much broader understanding of the genotypic-phenotypic associations. Studying such a large number of SNPs in case-control association studies offers a great challenge for scientists. In silico analysis using powerful software tools can facilitate predicting the phenotypic effect of ns-coding or non-coding (intronic) SNPs on the physicochemical properties of the concerned proteins and can preferentially act as genetic markers (Vignal *et al*., 2002).

Several studies showed that to increase the prediction accuracy in terms of sensitivity and specificity for selection of most deleterious functional mutation, the well documented approach to retrieve them from multiple tools and algorithms rather than selecting a single one (Grillo *et al*., 2010; Kaur *et al*., 2017). Keeping track of this approach, we employed 16 different tools divided into five groups including sequence-based, structure-based, consensus-based, supervised learn-based, and evolutionary-based methods while in case of 3′-UTR, 5′-UTR and non-coding SNPs PolymiRTs, Regulome DB, and SNPinfo were used, respectively (Reumers *et al*., 2006; Wang *et al*., 2006; Yue *et al*., 2006). Sequence based approach often have an advantage that it is suitable for proteins having closely related members but to study genotype-phenotype relationships, structure based methods are mandatory and should be used in combination to sequence based and other approaches. Furthermore, structure based approach is used to predict the effect of variations on secondary structure, binding properties and surface accessibility of proteins (Kaur *et al*., 2017).

In the present study, we found three missense variants including Pro2739Leu (rs676210), Arg532Trp (rs13306194) and His1923Arg (rs533617) that were previously reported in literature. Pro2739Leu (rs676210) variant was located in exon 26 (Fig. 1A). Xiao *et al*. (2017) had concluded that variant Pro2739Leu was associated with increased risk of Ischemic stroke in their haplotype analysis on Chinese Han population. Moreover, it was also found to be associated with increased risk of hyperlipidemia (HL) and CVD events (Buroker, 2014; Mäkelä *et al*., 2014; Gu *et al*., 2017). Similarly, the second variant Arg532Trp (rs13306194) was located in *Vitellogenin_N* (exon 12) domain (also known as N- terminal lipid transport domain), which is a conserved region of APOB protein and is mainly involved in lipid transport (Anderson *et al*., 1989). Tang *et al*. (2015) already reported it to be independently associated with blood lipid traits including total cholesterol and LDL-cholesterol levels in Chinese population that were linked with coronary artery disease and familial hypercholesterolemia. The third variant His1923Arg (rs533617) was also located in exon 26 of *APOB*. It was found to be associated with serum LDL-cholesterol levels in men (Ilmonen *et al.,* 1995). Limited evidence was found on their previous validation. The remaining six out of nine missense variants; rs41288783, rs544542990, rs72653074, rs181737266, rs536328155 and rs540387864 have not been previously reported as no validation study about their functional and structural analysis was available till date to the best of our knowledge (Table III). Hence, these variants of *APOB* are proposed as novel most deleterious variants of current study for further genetic association and linkages studies in future.

To analyze the effect of SNPs on protein structure, HOPE tool was used. The 3D homologues model of PDB ID ILSH of N-terminal region domain of APOB was collected and found to harboring only one variant rs13306194 in this domain. Due to its position, harboring conserved region of domain might be important for the main activity of the protein and hence can abolish domain function. While its amino acid properties represent that it is bigger in size which might lead to bumps, and charge neutral, which can cause loss of interactions with other molecules or residues. Furthermore, this mutant residue is more hydrophobic as compared to the wild type positively charged residue, which resulted in the loss of hydrogen bonds and/or disturb correct folding of protein APOB, hence disrupt the LDL-cholesterol metabolism (Fig. 2B). Several studies described a 670 amino acid homology sequence in the N-terminal of apolipoproteinB (APOB), apolipovitellin, and microsomal triglyceride transfer protein (MTP) (Baker, 1988; Shoulders *et al*., 1993, 1994), which is involved in lipid transport from liver to different tissues in the body.

The variants present in non-coding regions *i.e.* intronic, promoter regions or UTRs may also lead to several pathological conditions and could increase disease susceptibility. Several regulatory region SNPs of *VEGFA, ATF3, AKT3* genes have been described to play important role in susceptibility towards cancer development (Buroker,

2014; Kaur *et al*., 2017). In the current study using RegulomeDB and SNPinfo, we also found 15 non-coding region variants that were likely to affect transcription factor binding site. One variant rs12720762 was found to influence the gene expression, splicing and gene regulation by affecting transcription binding sites (TFBS) function, by applying both the both tools. No previous evidence or data was found about their clinical significance in dbSNP ClinVar. The 3′UTR also have vital role in gene expression as they provide the putative target site for miRNA binding. Any change in these regions by SNPs may either disrupt or create new target sites for miRNA and ultimately make susceptible to disease through affecting gene regulation. Several studies show that SNPs in miRNA target sites of *BRCA1*, *TGF-b* genes have been experimentally proved to increase the likelihood of lethal diseases, such as cancer (Nicoloso *et al*., 2010; Quann *et al*., 2015). Hence, in present study, by using polymiRNA tool we found two SNPs as rs72654430 and rs142151703 that could disturb the conserved site of miRNA or might create a new site for miRNA. Both of these variants were already reported with uncertain significance in Hypobetalipoproteinemia familial 1 and Familial hypercholesterolemia 2 by ClinVar in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). So they are also proposed as novel ncSNPs of 3′UTR of *APOB* in present study. The structure prediction of ApoB using I-Tasser and Chimera tools was also performed which yield the predicted model of 61.1% in most favorable region. Although a good predicted model has high score in most favorable region of Ramachandran plot (≥90%) but as the APOB is very large size protein so its exact structure prediction was difficult. Furthermore, energy minimization measurement of the predicted model and other computational tools should be used for the better model prediction of APOB in future.

CONCLUSION

The present study reports nine most deleterious missense coding SNPs including rs676210, rs13306194, rs533617, rs41288783, rs544542990, rs72653074, rs181737266, rs536328155, and rs540387864 which were extracted using 18 different computational tools. Three of them including rs676210, rs13306194, and rs533617 were already reported and validated in association with LDL-cholesterol while remaining six are proposed as novel missense variants of *APOB* that should be prioritized and investigated for further validation by *in vitro* or *in vivo* genetic association studies and clinical trials. Furthermore, in the context of protein structural and functional impact, the homology modeling of Arg532Trp variant constitute unique resource of genetic marker that may considerably increase the power of *APOB* mutation-screening in disease epidemiological studies. Interestingly, two variants of 3′UTR *i.e.* rs72654430 and rs142151703 were also proposed as novel variants of *APOB*. Thus, in a nutshell, we can say that the computational study carried out here was cost-effective, easy to analyze and monitor the predicted most deleterious coding nsSNPs and non-coding SNPs of *APOB* that should be prioritize in future genetic association studies of CVDs. Furthermore, their structural impact on APOB may suggest these predicted nsSNPs possibly be a better drug target and contribute to the treatment and better understanding of human cardiovascular disease.

ACKNOWLEDGEMENT

We would like to greatly acknowledge Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Jhang Road, Faisalabad, Pakistan, University of Sargodha and Dr. Panjwani Center for Molecular Medicine and Drug Research (ICCBS), Karachi Pakistan for supporting and facilitating this work.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the Pakistan.

Ethics approval

This article does not contain any studies with human participants performed by any of the authors.

Supplementary material

There is supplementary material associated with this article. Access the material online at: https://dx.doi.org/10.17582/journal.pjz/20210713140738

Statement of conflict of interest

The authors have declared no conflict of interests.

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A. and Bork, P., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods,* **7**: 248-249. https://doi.

org/10.1038/nmeth0410-248

Anderson, T.A., Levitt, D.G. and Banaszak, L.J., 1998. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure,* **6**: 895–909. https://doi.org/10.1016/S0969-2126(98)00091-4

Baker, M.E., 1988. Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem. J.,* **255**: 1057-1060. https://doi.org/10.1042/bj2551057

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D. and Zendulka, J., 2014. PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol.,* **10**: e1003440. https://doi.org/10.1371/journal.pcbi.1003440

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A. and Kasowski, M., 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Gen. Res.,* **22**: 1790-1797. https://doi.org/10.1101/gr.137323.112

Buroker, N.E., 2014. Regulatory SNPs and transcriptional factor binding sites in ADRBK1, AKT3, ATF3, DIO2, TBXA2R and VEGFA. *Transcription,* **5**: e964559. https://doi.org/10.4161/21541264.2014.964559

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P. and Duncanson, A., 2007. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.,* **39**: 1329-1337. https://doi.org/10.1038/ng.2007.17

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P.L. and Altman, R.B., 2013. WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genom.,* **14**: 1-7. https://doi.org/10.1186/1471-2164-14-S3-S6

Chirumbolo, S., 2016. Single nucleotide polymorphism (SNP) in the adiponectin gene and cardiovascular disease. *Iran. Biomed. J.,* **20**: 187-188.

Choi, Y. and Chan, A.P., 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics,* **31**: 2745-2747. https://doi.org/10.1093/bioinformatics/btv195

Feng, Z., Tie, G., Lv, Z., Yanhui, Z. and Dan, Y., 2017. Variants in the APOB gene was associated with ischemic stroke susceptibility in Chinese Han male population. *Oncotarget,* **9**: 2249–2254. https://doi.org/10.18632/oncotarget.23369

González-Pérez, A. and López-Bigas, N., 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.,* **88**: 440-449. https://doi.org/10.1016/j.ajhg.2011.03.004

Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S. and Banfi, S., 2010. UTRdb and UTRsite (RELEASE 2010): A collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucl. Acids Res.,* **38**(Suppl_1): D75-D80. https://doi.org/10.1093/nar/gkp902

Gu, Q., Han, Y., Lan, Y., Li, Y., Kou, W. and Zhou, Y., 2017. Association between polymorphisms in the *APOB* gene and hyperlipidemia in the Chinese Yugur population. *Braz. J. med. biol. Res.,* **50**: https://doi.org/10.1590/1414-431x20176613

Hepp, D., Gonçalves, G.L. and de Freitas, T.R.O., 2015. Prediction of the damage-associated non-synonymous single nucleotide polymorphisms in the human MC1R gene. *PLoS One*, **10**: e0121812. https://doi.org/10.1371/journal.pone.0121812

Ilmonen, M., Heliö, T., Bütler, R., Palotie, A., Pietinen, P. and Huttunen, J.K., 1995. Two new immunogenetic polymorphisms of the apoB gene and their effect on serum lipid levels and responses to changes in dietary fat intake. *Arterioscler. Thromb., Vasc. Biol.,* **15**: 1287-1293. https://doi.org/10.1161/01.ATV.15.9.1287

Joshi, B.B., Koringa, P.G., Mistry, K.N., Patel, A.K., Gang, S. and Joshi, C.G., 2015. *In silico* analysis of functional nsSNPs in human TRPC6 gene associated with steroid resistant nephrotic syndrome. *Gene,* **572**: 8-16. https://doi.org/10.1016/j.gene.2015.06.069

Kaur, B., Singh, J. and Kaur, M., 2017. Structural and functional impact of SNPs in P-selectin gene: A comprehensive *in silico* analysis. *Open Life Sci.,* **12**: 19-33. https://doi.org/10.1515/biol-2017-0003

Kaur, T., Thakur, K., Singh, J., Kamboj, S.S. and Kaur, M., 2017. Identification of functional SNPs in human LGALS3 gene by *in silico* analyses. *Egypt. J. med. Hum. Genet.,* **18**: 321-328. https://doi.org/10.1016/j.ejmhg.2017.02.001

Liwen, R., Jie, Y., Wan, L., Xiangjin, Z., Jinyi, L., Jinhua, W. and Guanhua, D., 2019. Apolipoproteins and cancer. *Cancer Med.,* **8**: 7032–7043. https://doi.org/10.1002/cam4.2587

Mah, J.T., Low, E.S. and Lee, E., 2011. *In silico* SNP analysis and bioinformatics tools: A review of the state of the art to aid drug discovery. *Drug Discov. Today,* **16**: 800-809. https://doi.org/10.1016/j.

drudis.2011.07.005

Mäkelä, K.M., Traylor, M., Oksala, N., Kleber, M.E., Seppäläm I. and Lyytikäinenm L.P., 2014. Association of the novel single-nucleotide polymorphism which increases oxidized low-density lipoprotein levels with cerebrovascular disease events. *Atherosclerosis,* **234**: 214-217. https://doi.org/10.1016/j.atherosclerosis.2014.03.002

Navarese, E.P., Robinson, J.G., Kowalewski, M., Kołodziejczak, M., Andreotti, F. and Bliden, K., 2018. Association between baseline LDL-C level and total and cardiovascular mortality after LDL-C lowering: A systematic review and meta-analysis. *J. Am. med. Assoc.,* **319**: 1566-1579. https://doi.org/10.1001/jama.2018.2525

Nicoloso, M.S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P. and Shimizu, M., 2010. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.,* **70**: 2789-2798. https://doi.org/10.1158/0008-5472.CAN-09-3541

Niroula, A., Urolagin, S. and Vihinen, M., 2015. PON-P2: Prediction method for fast and reliable identification of harmful variants. *PLoS One*, **10**: e0117380. https://doi.org/10.1371/journal.pone.0117380

Nordestgaard, B.G., Langlois, M.R., Langsted, A., Chapman, M.J., Aakre, K.M., Baum, H. and Borén, J., 2020. Quantifying atherogenic lipoproteins for lipid-lowering strategies: Consensus-based recommendations from EAS and EFLM. *Atherosclerosis*, **294**: 46-61. https://doi.org/10.1016/j.atherosclerosis.2019.12.005

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N. and Nam, H.J., 2020. MutPred2: Inferring the molecular and phenotypic impact of amino acid variants. *Nat. Comun*., **1**: 5918. https://doi.org/10.1038/s41467-020-19669-x

Pires, A.S., Porto, W.F., Franco, O.L. and Alencar, S.A., 2017. *In silico* analyses of deleterious missense SNPs of human apolipoprotein E3. *Scient. Rep.,* **7**: 1-9. https://doi.org/10.1038/s41598-017-01737-w

Quann, K., Jing, Y. and Rigoutsos, I., 2015. Post-transcriptional regulation of BRCA1 through its coding sequence by the miR-15/107 group of miRNAs. *Front. Genet.,* **6**: 242. https://doi.org/10.3389/fgene.2015.00242

Reumers, J., Maurer-Stroh, S., Schymkowitz, J. and Rousseau, F., 2006. SNPeffect v2. 0: A new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. *Bioinformatics,* **22**: 2183-2185. https://doi.org/10.1093/bioinformatics/btl348

Reva, B., Antipin, Y. and Sander, C., 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucl. Acids Res.,* **39**: e118–e118. https://doi.org/10.1093/nar/gkr407

Rost, B., 2001. Protein secondary structure prediction continues to rise. *J. Struct. Biol.,* **134**: 204-218. https://doi.org/10.1006/jsbi.2001.4336

Shoulders, C.C., Brett, D.J., Bayllss, D.J., Narcisi, M.E., Jarmuz, A. and Grantham, T.T., 1993. Abetalipoproteinemia is caused by defects of the gene encoding the 97 kDa subunit of a microsomal triglyceride transfer protein. *Hum. mol. Genet.,* **2**: 2109-2116. https://doi.org/10.1093/hmg/2.12.2109

Shoulders, C.C., Narcisi, T.M.E., Read, J., Chester, S.A., Brett, D.J., Scott, J., Timothy, A., Anderson, T.A., Levitt, D.G. and Banaszak, L.J., 1994. The abetalipoproteinemia gene is a member of the vitellogenin family and encodes an α–helical domain. *Nat. Struct. Biol.,* **1**: 285-286. https://doi.org/10.1038/nsb0594-285

Tang, C.S., Zhang, H., Cheung, C.Y., Xu, M., Ho, J.C. and Zhou, W., 2015. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. *Nat. Commun.,* **6**: 1-9. https://doi.org/10.1038/ncomms10206

Tang, H. and Thomas, P.D., 2016. PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics,* **32**: 2230-2232. https://doi.org/10.1093/bioinformatics/btw222

Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G., 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics,* **11**: 1-10. https://doi.org/10.1186/1471-2105-11-548

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.,* **34**: 275–305. https://doi.org/10.1186/1297-9686-34-3-275

Wang, P., Dai, M., Xuan, W., McEachin, R.C., Jackson, A.U. and Scott, L.J., 2006. SNP Function Portal: A web database for exploring the function implication of SNP alleles. *Bioinformatics,* **22**: e523-e529. https://doi.org/10.1093/bioinformatics/btl241

Wang, Y.T., Li, Y., Ma, Y.T., Yang, Y.N., Ma, X. and Li, X.M., 2018. Association between apolipoprotein B genetic polymorphism and the risk of calcific aortic stenosis in Chinese subjects, in Xinjiang, China. *Lipids Hlth. Dis*., **17**: 1-7. https://doi.org/10.1186/

s12944-018-0696-6

Xiao, R. Sun, S. Zhang, J. Ouyang, Y. Zhang, N. Yang, M. Jin, T. and Xia, Y. 2017. Association analysis of APO gene polymorphisms with ischemic stroke risk: a case-control study in a Chinese Han population. *Oncotarget,* **8**: 60496–60503.

Xu, Z. and Taylor, J.A., 2009. SNPinfo: Integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucl. Acids Res.,* **37**(Suppl_2): W600-W605. https://doi.org/10.1093/nar/gkp290

Yue, P., Melamud, E. and Moult, J., 2006. SNPs3D: Candidate gene and SNP selection for association studies. *BMC Bioinformatics,* **7**: 1-15. https://doi.org/10.1186/1471-2105-7-166