Pakistan J. Zool., pp 1-13, 2021. **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)

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 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

* Corresponding author: munazzaraza@iccs.com
0030-9923/2021/0001-0001 $ 9.00/0

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).

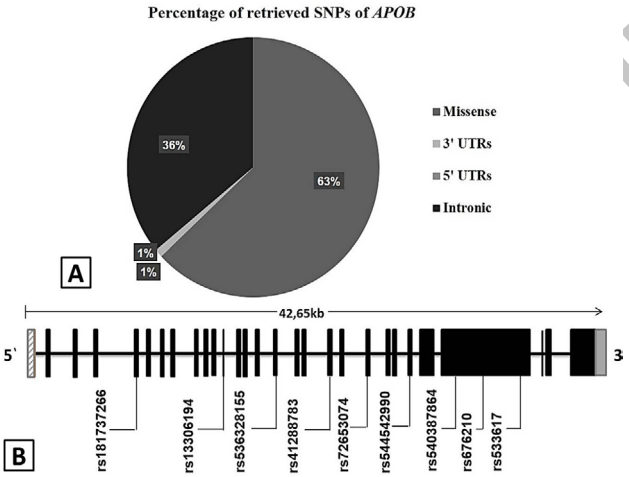

Fig. 1. **A**, Pie chart of retrieved validated SNPs of *APOB* from NCBI and Ensembl data bases. It includes 295 (62%) missense, 170 (36%) intronic, 4 (1%) 3′ UTR and 5 (1%) 5′-UTR. **B**, Convergent deleterious and functionally important SNPs are located in distinct exonic region of *APOB* gene. The 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.

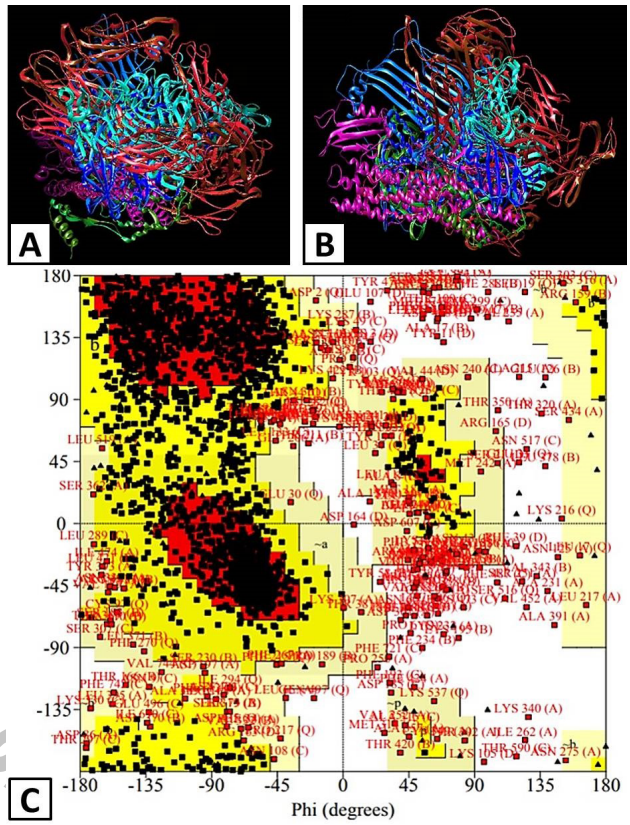

Fig. 2. Predicted Model of APOB protein by using UCSF Chimera. a) Front view of the predicted model. B) Lateral view of the predicted model. C) Ramachandran plot of predicted APOB model by using PROCHECH. Region of A, B and L were considered core regions. a, b, l and p showed allowed region, ~a, ~b,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

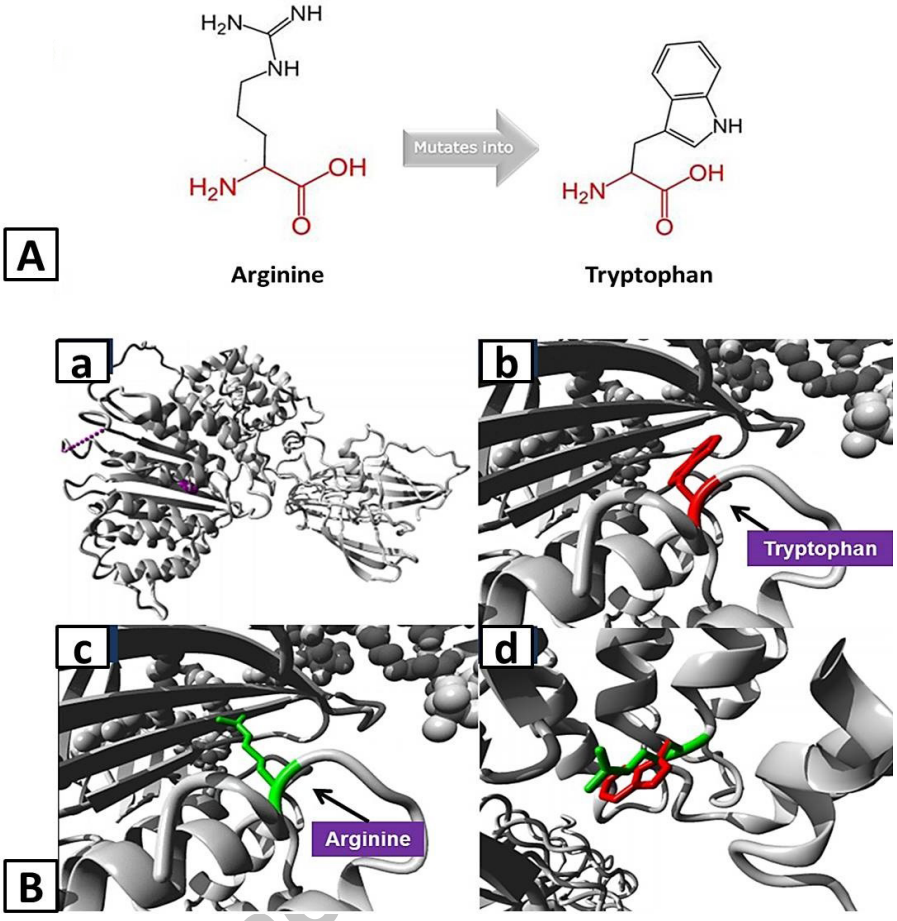

Fig. 3. **A**, Schematic structures of the original (left) Arginine and the mutant (right) amino acid tryptophan for variant Arg532Trp of *APOB* created by HOPE tool. The backbone, which is the same for each amino acid, is colored red and the side chain, unique for each amino acid, is colored black. **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

Pakistan J. Zool., pp 1-13, 2021. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20210713140738

Supplementary Material

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*

Supplementary Table I.- Working inputs and prediction score criteria used for 18 computational tools used in prediction of deleterious *APOB* SNPs.

S. No.	Tools	Criteria	Input
Sequence Based Tools and their prediction criteria			
1.	SIFT	Score > 0.5 Tolerable Score < 0.5 Deleterious	rs dbSNP
2.	PROVEAN	Score < -2.5 Deleterious Score > -2.5 Neutral	ENSP00000224605 63 A
3.	Mutation Assessor	Neutral /Low Medium/High	UniProt ID position Wild and variant
4.	PON-P2	Pathogenic Neutral Unknown	FASTA-format amino acid wild and mutant in header
5.	PhD-SNP	RI > 0.5 DiseaseRI < 0.5 Neutral	FASTA seq. or Swiss Port ID, AA position and new variable
Supervised Learned Based Tools and their prediction criteria			
6.	SNAP	-100 Neutral +100 Deleterious	FASTA Sequence
7.	SuSPect	0 - 100	UniPort Accession and mutations
8.	MutPred	Score > 0.5 Deleterious Score < 0.5 Neutral	Protein FASTA sequence along amino acid substitution in sequence header
Structure Based Tools and their prediction criteria			
9.	PolyPhen	0 – Benign 1 - Probably Damaging Between 0 – 1 Possibly Damaging	Protein FASTA seq. AA substitution and position
10.	I-Mutant3	Score > - 0.5 Stable Score < -0.5 Unstable	FASTA seq. or Swiss Port ID, AA position and new variable

* Corresponding author: munazzaraza@iccs.com
0030-9923/2021/0001-0001 $ 9.00/0

S. No.	Tools	Criteria	Input
Consensus Based Tools and their prediction criteria			
11.	Condel	0 – Neutral 1.0 − Deleterious	Identifier Substitutions
12.	MetaSNP	Output > 0.5 = Deleterious	Protein sequence and a list of comma separated mutations
13.	PredictSNP	Green – Neutral Red − Deleterious	Protein FASTA sequence along substitution
Evolutionary conservation based method			
14.	PANTHER	PT > 450 my Damaging PT < 200 my Neutral	FASTA seq. and substitutions
Tools used for prediction of non- coding region SNPs			
15.	RegulomeDB	1 (Likely to affect binding and liked expression) **2** (Likely to affect binding) **3** (Less likely to affect) **4 ,5, 6** (Minimal binding)	rs dbSNP ID
16.	PolymiRTs	**D** (Disturb) **N** (Disturb non conserved site) **C** (Create new allele) **O (**Ancestor allele not determined)	rs dbSNP ID
17.	SNPinfo	Y sSNP Splicing Regulation Stop Codon Polyphen Prediction SNPs3D Prediction TFBS Prediction miRNA Binding Site Prediction	rs dbSNP ID
18	HOPE	Structural features Different properties of amino acid Generate 3D models	Protein sequence and the mutation position

Supplementary Table II.- Retrieved Data of missense SNPs present in *APOB* by using from NCBI dbSNP data base.

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
1	ENSP00000233242.1	rs676210	21008652	0.3662	missense	C/A,	P [Pro] ⇒ Q [Gln]	NP_000375.2
2	ENSP00000233242.1	rs679899	21028042	0.485	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
3	ENSP00000233242.1	rs1042031	21002881	0.1278	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
4	ENSP00000233242.1	rs1042034	21002409	0.3704	missense	G/A	S [Ser] ⇒ N [Asn]	NP_000375.2
5	ENSP00000233242.1	rs1367117	21041028	0.1693	missense	C/T	T [Thr] ⇒ I [Ile]	NP_000375.2
6	ENSP00000233242.1	rs1801695	21001981	0.0142	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
7	ENSP00000233242.1	rs1801696	21009172	0.0006	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
8	ENSP00000233242.1	rs1801701	21005955	0.0385	missense	G/A	R [Arg] ⇒ Q [Gln]	NP_000375.2
9	ENSP00000233242.1	rs1801702	2100261	0.0627	missense	G/C	R [Arg] ⇒ T [Thr]	NP_000375.2
10	ENSP00000233242.1	rs13306194	21029662	0.0274	missense	C/T	R [Arg] ⇒ W [Trp]	NP_000375.2
11	ENSP00000233242.1	rs61743502	21002628	0.002	missense	T/C	V [Val] ⇒ A [Ala]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
12	ENSP00000233242.1	rs72653092	21009583	0.0004	missense	T/A, T/G	S [Ser] ⇒ T [Thr], S [Ser] ⇒ A [Ala]	NP_000375.2
13	ENSP00000233242.1	rs72654423	21002482	0.0042	missense	A/G	I [Ile] ⇒ V [Val]	NP_000375.2
14	ENSP00000233242.1	rs533617	21011100	0.0162	missense	A/G	H [His] ⇒ R [Arg]	NP_000375.2
15	ENSP00000233242.1	rs568413	21012603	---	missense	G/A, G/C, G/T	C [Cys] ⇒ Y [Tyr], C [Cys]	NP_000375.2
16	ENSP00000233242.1	rs584542	21009931	0.0138	missense	G/T, G/A	V [Val] ⇒ I [Ile], V [Val] ⇒ F [Phe]	NP_000375.2
17	ENSP00000233242.1	rs1042023	21006574	0.0028	missense	C/G	Q [Gln] ⇒ E [Glu]	NP_000375.2
18	ENSP00000233242.1	rs1800480	21044060	0.0172	UTR-5	----/G	-------	------
19	ENSP00000233242.1	rs1801698	21004631	0.0004	missense	A/G	T [Thr] ⇒ A [Ala]	NP_000375.2
20	ENSP00000233242.1	rs1801699	21011127	0.0112	missense	A/G	N [Asn] ⇒ S [Ser]	NP_000375.2
21	ENSP00000233242.1	rs1801703	21003040	0.0034	missense	G/A, G/C, G/T	V [Val] ⇒ M [Met],V [Val]	NP_000375.2
22	ENSP00000233242.1	rs2163204	21008515	0.0118	missense	A/C	N [Asn] ⇒ H [His]	NP_000375.2
23	ENSP00000233242.1	rs6752026	21038062	0.0349	missense	C/T	P [Pro] ⇒ S [Ser]	NP_000375.2
24	ENSP00000233242.1	rs9282603	21041014	0.001	missense	T/C	Y [Tyr] ⇒ H [His]	NP_000375.2
25	ENSP00000233242.1	rs12691202	21026844	0.0142	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
26	ENSP00000233242.1	rs12713450	21001971	0.0383	missense	C/T	T [Thr] ⇒ M [Met]	NP_000375.2
27	ENSP00000233242.1	rs12713540	21005467	0.0004	missense	T/A	S [Ser] ⇒ T [Thr]	NP_000375.2
28	ENSP00000233242.1	rs12713559	21006196	0.0002	missense	C/T	R [Arg] ⇒ C [Cys]	NP_000375.2
29	ENSP00000233242.1	rs12713675	21009501	0.0222	missense	C/A	A [Ala] ⇒ D [Asp]	NP_000375.2
30	ENSP00000233242.1	rs12713681	21009973	0.0088	missense	G/C	D [Asp] ⇒ H [His]	NP_000375.2
31	ENSP00000233242.1	rs12713843	21015495	0.0028	missense	G/A	R [Arg] ⇒ H [His]	NP_000375.2
32	ENSP00000233242.1	rs12713844	21015541	0.003	missense	G/C	D [Asp] ⇒ H [His]	NP_000375.2
33	ENSP00000233242.1	rs12714192	21026810	0.0218	missense	C/A	T [Thr] ⇒ N [Asn]	NP_000375.2
34	ENSP00000233242.1	rs12714214	21028495	0.003	missense	C/G, C/T	P [Pro] ⇒ R [Arg] ,P [Pro] ⇒ L [Leu]	NP_000375.2
35	ENSP00000233242.1	rs12714225	21032483	0.0082	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
36	ENSP00000233242.1	rs12720763	21001550	0.0102	UTR-3	---/A	-------	--------
37	ENSP00000233242.1	rs12720854	21007033	0.0072	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
38	ENSP00000233242.1	rs12720855	21006988	0.0224	missense	T/C	S [Ser] ⇒ P [Pro]	NP_000375.2
39	ENSP00000233242.1	rs13306187	21013213	0.004	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
40	ENSP00000233242.1	rs13306190	21032408	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
41	ENSP00000233242.1	rs13306198	21037212	0.0142	missense	C/T	T [Thr] ⇒ M [Met]	NP_000375.2
42	ENSP00000233242.1	rs13306206	21019859	0.0018	missense	C/T	P [Pro] ⇒ S [Ser]	NP_000375.2
43	ENSP00000233242.1	rs41288783	21019741	0.001	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
44	ENSP00000233242.1	rs61736761	21015135	0.0284	missense	C/A	L [Leu] ⇒ M [Met]	NP_000375.2
45	ENSP00000233242.1	rs61741164	21014550	0.0006	missense	A/G	Y [Tyr] ⇒ C [Cys]	NP_000375.2
46	ENSP00000233242.1	rs61741625	21035651	0.0026	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
47	ENSP00000233242.1	rs61741974	21005611	0.0012	missense	T/C	F [Phe] ⇒ L [Leu]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
48	ENSP00000233242.1	rs61742331	21006807	0.0018	missense	C/G	A [Ala] ⇒ G [Gly]	NP_000375.2
49	ENSP00000233242.1	rs61743299	21002725	0.0224	missense	T/A	S [Ser] ⇒ T [Thr]	NP_000375.2
50	ENSP00000233242.1	rs61743313	21002656	0.0014	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
51	ENSP00000233242.1	rs61744153	21005391	0.0006	missense	C/A	T [Thr] ⇒ K [Lys]	NP_000375.2
52	ENSP00000233242.1	rs61744288	21006088	0.0018	missense	T/C	W[Trp] ⇒ R [Arg]	NP_000375.2
53	ENSP00000233242.1	rs61746672	21037187	0.0034	missense	A/T	E [Glu] ⇒ D [Asp]	NP_000375.2
54	ENSP00000233242.1	rs72653059	21037186	0.0042	missense	A/G	I [Ile] ⇒ V [Val]	NP_000375.2
55	ENSP00000233242.1	rs72653077	21015451	0.0015	missense	C/T	P [Pro] ⇒ S [Ser]	NP_000375.2
56	ENSP00000233242.1	rs72653078	21014578	0.001	missense	C/A	L [Leu] ⇒ I [Ile]	NP_000375.2
57	ENSP00000233242.1	rs72653091	21009626	0.003	missense	A/C	E [Glu] ⇒ D [Asp]	NP_000375.2
58	ENSP00000233242.1	rs72653095	21008406	0.0062	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
59	ENSP00000233242.1	rs72653099	21007642	0.0014	missense	C/A	L[Leu] ⇒ M [Met]	NP_000375.2
60	ENSP00000233242.1	rs72654409	21005107	0.0034	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
62	ENSP00000233242.1	rs72654424	21002320	0.0008	missense	C/G	Q [Gln] ⇒ E [Glu]	NP_000375.2
63	ENSP00000233242.1	rs72654428	21001719	0.0084	UTR-3	----	------	-----
64	ENSP00000233242.1	rs72654430	21001501	0.003	UTR-3	-----	-----	-----
65	ENSP00000233242.1	rs140027955	21009645	0.0004	missense	C/T	S [Ser] ⇒ F [Phe]	NP_000375.2
66	ENSP00000233242.1	rs141225768	21012205	0.0002	missense	A/G	I [Ile] ⇒ V [Val]	NP_000375.2
67	ENSP00000233242.1	rs142114415	21032396	0.001	missense	G/A, G/T	R[Arg] ⇒ H [His] , R [Arg] ⇒ L [Leu]	NP_000375.2
68	ENSP00000233242.1	rs142151703	21001551	0.0022	UTR-3	----	-----	-----
69	ENSP00000233242.1	rs142422341	21007057	0.0002	missense	G/A	G [Gly] ⇒ S [Ser]	NP_000375.2
70	ENSP00000233242.1	rs143269114	21007377	0.0006	missense	C/T	T[Thr] ⇒ M [Met]	NP_000375.2
71	ENSP00000233242.1	rs143282164	21013198	0.001	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
72	ENSP00000233242.1	rs143425834	21026862	0.0002	missense	G/T	G [Gly] ⇒ C [Cys]	NP_000375.2
73	ENSP00000233242.1	rs143710616	21005514	0.0002	missense	C/T	T [Thr] ⇒ I [Ile]	NP_000375.2
74	ENSP00000233242.1	rs144467873	21006289	0.0004	missense	C/T	R[Arg] ⇒ W [Trp]	NP_000375.2
75	ENSP00000233242.1	rs145142090	21023544	0.0004	missense	T/C	V [Val] ⇒ A [Ala]	NP_000375.2
76	ENSP00000233242.1	rs145862664	21028508	0.0006	missense	G/C	D[Asp] ⇒ H [His]	NP_000375.2
77	ENSP00000233242.1	rs146152405	21037152	0.0004	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
78	ENSP00000233242.1	rs148170480	21009253	0.0016	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
79	ENSP00000233242.1	rs149227065	21033362	0.0002	missense	A/G, A/G	E [Glu] ⇒ G [Gly],E [Glu] ⇒ V [Val]	NP_000375.2
80	ENSP00000233242.1	rs151009667	21011802	0.0004	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
81	ENSP00000233242.1	rs151333262	21002239	0.0002	missense	G/A	G [Gly] ⇒ S [Ser]	NP_000375.2
82	ENSP00000233242.1	rs180874451	21011409	0.0006	missense	A/C	K [Lys] ⇒ T [Thr]	NP_000375.2
83	ENSP00000233242.1	rs183117027	21004340	0.0008	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
84	ENSP00000233242.1	rs183398286	21022990	0.0002	missense	A/G	N [Asn] ⇒ S [Ser]	NP_000375.2
85	ENSP00000233242.1	rs186299244	21006985	0.0004	missense	T/C	Y [Tyr] ⇒ H [His]	NP_000375.2
86	ENSP00000233242.1	rs186544754	21041033	0.0022	missense	G/T	Q[Gln] ⇒ H [His]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
87	ENSP00000233242.1	rs199510126	21011178	0.0002	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
88	ENSP00000233242.1	rs199668351	21003112	0.0006	missense	C/A	L[Leu] ⇒ M [Met]	NP_000375.2
89	ENSP00000233242.1	rs199859104	21015453	0.0002	missense	C/T	S [Ser] ⇒ L [Leu]	NP_000375.2
90	ENSP00000233242.1	rs200106845	21010212	0..0002	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
91	ENSP00000233242.1	rs375894411	21015174	0.0004	missense	G/A	D[Asp] ⇒ N [Asn]	NP_000375.2
92	ENSP00000233242.1	rs531273434	21009015	0.0004	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
93	ENSP00000233242.1	rs544542990	21014461	0.0002	missense	C/T, A/G	L [Leu] ⇒ F [Phe], R[Arg] ⇒ G [Gly]	NP_000375.2
94	ENSP00000233242.1	rs551178628	21005425	0.0004	missense	G/A, G/T	V[Val] ⇒ M [Met],V [Val] ⇒ L [Leu]	NP_000375.2
95	ENSP00000233242.1	rs558304720	21012282	0.0002	missense	G/A, G/T	G[Gly] ⇒ D [Asp] G[Gly] ⇒ V [Val]	NP_000375.2
96	ENSP00000233242.1	rs574725520	21007229	0.0002	missense	C/A	N [Asn] ⇒ K [Lys]	NP_000375.2
97	ENSP00000233242.1	rs575505383	21019079	0.0016	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
98	ENSP00000233242.1	rs12720801	21016515	0.0002	missense	G/A, G/C	G [Gly] ⇒ S [Ser], G [Gly] ⇒ R [Arg]	NP_000375.2
99	ENSP00000233242.1	rs61741467	21016646	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
100	ENSP00000233242.1	rs61742323	21007728	0.0034	missense	C/G, C/T	T [Thr] ⇒ R [Arg], T [Thr] ⇒ M [Met]	NP_000375.2
101	ENSP00000233242.1	rs61744151	21005956	0.0002	missense,	C/A	R [Arg] ⇒ R [Arg]	NP_000375.2
102	ENSP00000233242.1	rs61746686	21005148	0.0006	missense	A/G	D [Asp] ⇒ G [Gly]	NP_000375.2
103	ENSP00000233242.1	rs72653060	21034825	0.001	missense	T/G	F [Phe] ⇒ V [Val]	NP_000375.2
104	ENSP00000233242.1	rs72653074	21016551	0.0002	missense	G/A	G [Gly] ⇒ R [Arg]	NP_000375.2
105	ENSP00000233242.1	rs72653076	21015496	0.0004	missense	C/T	R [Arg] ⇒ C [Cys]	NP_000375.2
106	ENSP00000233242.1	rs72653088	21010006	0.0002	missense	C/T	H [His] ⇒ Y [Tyr]	NP_000375.2
107	ENSP00000233242.1	rs72653100	21007626	0.0014	missense	G/A, G/C	S [Ser] ⇒ N [Asn],S [Ser] ⇒ T [Thr]	NP_000375.2
108	ENSP00000233242.1	rs72653101	21007547	0.0006	missense	C/G	N [Asn] ⇒ K [Lys]	NP_000375.2
109	ENSP00000233242.1	rs72653102	21007462	0.0004	missense	C/T	R [Arg] ⇒ C [Cys]	NP_000375.2
110	ENSP00000233242.1	rs72654404	21007001	0.0002	missense,	C/A, C/T	D [Asp] ⇒ E [Glu],T [Thr] ⇒ T [Thr]	NP_000375.2
111	ENSP00000233242.1	rs72654416	21004645	0.0026	missense	C/T	T [Thr] ⇒ M [Met]	NP_000375.2
112	ENSP00000233242.1	rs72654422	21002619	0.003	missense	T/C	M [Met] ⇒ T [Thr]	NP_000375.2
113	ENSP00000233242.1	rs72654425	21002183	0.0002	missense	C/G	N [Asn] ⇒ K [Lys]	NP_000375.2
114	ENSP00000233242.1	rs78875649	21015162	0.0006	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
115	ENSP00000233242.1	rs80179164	21041083	0.0002	missense	G/C	V [Val] ⇒ L [Leu]	NP_000375.2
116	ENSP00000233242.1	rs138391809	21008559	0.0006	missense	C/T	T [Thr] ⇒ I [Ile]	NP_000375.2
117	ENSP00000233242.1	rs138529007	21007651	0.0002	missense	A/G	N [Asn] ⇒ D [Asp]	NP_000375.2
118	ENSP00000233242.1	rs139599466	21005607	0.0002	missense	C/T	T [Thr] ⇒ I [Ile]	NP_000375.2
119	ENSP00000233242.1	rs139857448	21010000	0.0002	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
120	ENSP00000233242.1	rs139929439	21006392	0.001	Missense,	T/A	I [Ile] ⇒ I [Ile]	NP_000375.2
121	ENSP00000233242.1	rs140246447	21012391	0.0002	missense	T/C	S [Ser] ⇒ P [Pro]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
122	ENSP00000233242.1	rs140647761	21006513	0.0002	missense	C/G, C/T	P [Pro] ⇒ R [Arg],P [Pro] ⇒ L [Leu]	NP_000375.2
123	ENSP00000233242.1	rs140773510	21013262	0.0002	missense	T/A	S [Ser] ⇒ T [Thr]	NP_000375.2
124	ENSP00000233242.1	rs140858817	21011824	0.0004	missense	A/G	M[Met] ⇒ V [Val]	NP_000375.2
125	ENSP00000233242.1	rs140877474	21012493	0.0014	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
126	ENSP00000233242.1	rs141591543	21008019	0.0004	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
127	ENSP00000233242.1	rs142638069	21011758	0.0006	missense	G/A, G/T	A [Ala] ⇒ T [Thr] ,A [Ala] ⇒ S [Ser]	NP_000375.2
128	ENSP00000233242.1	rs142756262	21007986	0.0002	missense	A/G, A/T	N [Asn] ⇒ S [Ser],N [Asn] ⇒ I [Ile]	NP_000375.2
129	ENSP00000233242.1	rs143612660	21002671	0.0002	missense	A/G, G/C	I [Ile] ⇒ V [Val], S [Ser] ⇒ S [Ser]	NP_000375.2
130	ENSP00000233242.1	rs145324793	21012588	0.0002	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
131	ENSP00000233242.1	rs145424390	21010849	0.0008	missense	G/A	V[Val] ⇒ M [Met]	NP_000375.2
132	ENSP00000233242.1	rs146687572	21001752	0.0004	missense	G/C	G[Gly] ⇒ A [Ala]	NP_000375.2
133	ENSP00000233242.1	rs146687604	21006931	0.001	missense	C/A	L [Leu] ⇒ I [Ile]	NP_000375.2
134	ENSP00000233242.1	rs147173587	21027964	0.0002	missense	C/A	S [Ser] ⇒ Y [Tyr]	NP_000375.2
135	ENSP00000233242.1	rs147447233	21009964	0.0008	missense	G/A	G [Gly] ⇒ R [Arg]	NP_000375.2
136	ENSP00000233242.1	rs147564959	21005635	0.0002	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
137	ENSP00000233242.1	rs148126873	21033413	0.0002	missense	T/A	I [Ile] ⇒ N [Asn]	NP_000375.2
138	ENSP00000233242.1	rs148190577	21024958	0.0004	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
139	ENSP00000233242.1	rs148451559	21011717	0.0002	missense	G/A	M [Met] ⇒ I [Ile]	NP_000375.2
140	ENSP00000233242.1	rs148502464	21025111	0.0004	missense	G/A	G [Gly] ⇒ E [Glu]	NP_000375.2
141	ENSP00000233242.1	rs148503464	21041049	0.0002	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
142	ENSP00000233242.1	rs148987924	21008904	0.0002	missense	C/A, C/G	T [Thr] ⇒ N [Asn], T [Thr] ⇒ S [Ser]	NP_000375.2
143	ENSP00000233242.1	rs149349129	21009828	0.0002	missense	G/A	R[Arg] ⇒ K [Lys]	NP_000375.2
144	ENSP00000233242.1	rs149357946	21019061	0.0004	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
145	ENSP00000233242.1	rs149887185	21006466	0.0002	missense	A/G	M[Met] ⇒ V [Val]	NP_000375.2
146	ENSP00000233242.1	rs150110189	21004370	0.0004	missense	G/A	G [Gly] ⇒ S [Ser]	NP_000375.2
147	ENSP00000233242.1	rs181143364	21029669	0.0002	Missense	G/A	Q[Gln] ⇒ Q [Gln]	NP_000375.2
148	ENSP00000233242.1	rs181570298	21016629	0.0004	missense	A/T	T [Thr] ⇒ S [Ser]	NP_000375.2
149	ENSP00000233242.1	rs181737266	21038061	0.0004	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
150	ENSP00000233242.1	rs181877973	21011034	0.0002	missense	A/G	D[Asp] ⇒ G [Gly]	NP_000375.2
151	ENSP00000233242.1	rs183394803	21008487	0.0002	missense	C/T	T [Thr] ⇒ I [Ile]	NP_000375.2
152	ENSP00000233242.1	rs183475426	21008028	0.0006	missense	G/C	S [Ser] ⇒ T [Thr]	NP_000375.2
153	ENSP00000233242.1	rs183812948	21002053	0.0008	missense	G/A	D[Asp] ⇒ N [Asn]	NP_000375.2
154	ENSP00000233242.1	rs183950016	21024971	0.0002	missense	C/A	L[Leu] ⇒ M [Met]	NP_000375.2
155	ENSP00000233242.1	rs184512808	21010317	0.0002	missense	A/G	Y [Tyr] ⇒ C [Cys]	NP_000375.2
156	ENSP00000233242.1	rs184685800	21009814	0.0002	missense	C/G	Q [Gln] ⇒ E [Glu]	NP_000375.2
157	ENSP00000233242.1	rs185224477	21007420	0.0002	missense	T/C	F [Phe] ⇒ L [Leu]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
158	ENSP00000233242.1	rs185540148	21016457	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
159	ENSP00000233242.1	rs186346145	21011685	0.0002	missense	A/G	N [Asn] ⇒ S [Ser]	NP_000375.2
160	ENSP00000233242.1	rs186480094	21011134	0.0004	missense	A/T	N[Asn] ⇒ Y [Tyr]	NP_000375.2
161	ENSP00000233242.1	rs187506285	21019873	0.0002	missense	C/T	T[Thr] ⇒ M [Met]	NP_000375.2
162	ENSP00000233242.1	rs188019153	21001699	0.0004	UTR' 3		-----	NP_000375.2
163	ENSP00000233242.1	rs188592517	21004369	0.0002	missense	G/A, G/T	G[Gly] ⇒ D [Asp], G[Gly] ⇒ V [Val]	NP_000375.2
164	ENSP00000233242.1	rs189341276	21011179	0.0006	missense	C/A, C/T	R [Arg] ⇒ S [Ser] ,R [Arg] ⇒ C [Cys]	NP_000375.2
165	ENSP00000233242.1	rs189880634	21023593	-----	missense	T/A	L[Leu] ⇒ M [Met]	NP_000375.2
166	ENSP00000233242.1	rs190114930	21006438	0.0002	missense	T/C	V [Val] ⇒ A [Ala]	NP_000375.2
167	ENSP00000233242.1	rs190134103	21037963	0.0002	missense	T/A, T/C	F [Phe] ⇒ I [Ile] ,F [Phe] ⇒ L [Leu]	NP_000375.2
168	ENSP00000233242.1	rs191145848	21012073	0.0004	missense	C/T	R [Arg] ⇒ C [Cys]	NP_000375.2
169	ENSP00000233242.1	rs192004342	21032448	-----	missense	G/A, G/C	E [Glu] ⇒ K [Lys], E [Glu] ⇒ Q [Gln]	NP_000375.2
170	ENSP00000233242.1	rs192491009	21008393	0.0002	missense	G/C	K[Lys] ⇒ N [Asn]	NP_000375.2
171	ENSP00000233242.1	rs192869978	21028037	0.0002	missense	A/G	K [Lys] ⇒ E [Glu]	NP_000375.2
172	ENSP00000233242.1	rs199585500	21011288	0.0002	missense	G/T	Q[Gln] ⇒ H [His]	NP_000375.2
173	ENSP00000233242.1	rs199590149	21007461	0.0002	missense	G/A, G/C	R[Arg] ⇒ H [His] R [Arg] ⇒ P [Pro]	NP_000375.2
174	ENSP00000233242.1	rs199689957	21006601	0.0002	missense	G/A	V[Val] ⇒ M [Met]	NP_000375.2
175	ENSP00000233242.1	rs199815987	21012003	0.0002	missense	A/G	H [His] ⇒ R [Arg]	NP_000375.2
176	ENSP00000233242.1	rs200034452	21009774	0.0004	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
177	ENSP00000233242.1	rs200143030	21009537	0.0012	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
178	ENSP00000233242.1	rs200184366	21006280	0.0002	missense	G/A	V[Val] ⇒ M [Met]	NP_000375.2
179	ENSP00000233242.1	rs200222843	21003286	0.0002	missense	C/T	R[Arg] ⇒ W [Trp]	NP_000375.2
180	ENSP00000233242.1	rs200231583	21009139	0.0002	missense	A/C	M[Met] ⇒ L [Leu]	NP_000375.2
181	ENSP00000233242.1	s200318200	21041038	0.0002	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
182	ENSP00000233242.1	rs200321839	1012618	0.0002	missense	C/T	T [Thr] ⇒ M [Met]	NP_000375.2
183	ENSP00000233242.1	rs200367807	21008233	0.0004	missense	A/T	N [Asn] ⇒ Y [Tyr]	NP_000375.2
184	ENSP00000233242.1	rs200370790	21010351	0.0002	missense	C/G	Q [Gln] ⇒ E [Glu]	NP_000375.2
185	ENSP00000233242.1	rs200464882	21043527	0.0002	missense	A/C	S [Ser] ⇒ T [Thr]	NP_000375.2
186	ENSP00000233242.1	rs200469887	21002080	0.0002	missense	A/T	I [Ile] ⇒ F [Phe]	NP_000375.2
187	ENSP00000233242.1	rs200524554	21025057	0.0002	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
188	ENSP00000233242.1	rs200584734	21007345	0.0002	missense	G/C	A [Ala] ⇒ P [Pro]	NP_000375.2
189	ENSP00000233242.1	rs200759464	21009627	0.0002	missense	A/T	E [Glu] ⇒ V [Val]	NP_000375.2
190	ENSP00000233242.1	rs200783423	21009496	0.0002	missense	G/A, G/C	E [Glu] ⇒ K [Lys] ,E [Glu] ⇒ Q [Gln]	NP_000375.2
191	ENSP00000233242.1	rs200824333	21008143	0.0002	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
192	ENSP00000233242.1	rs200940198	21009879	0.0002	missense	T/A, T/C	V [Val] ⇒ E [Glu], V [Val] ⇒ A [Ala]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
193	ENSP00000233242.1	rs201152495	21009144	0.0002	missense	A/T, A/T	K [Lys] ⇒ I [Ile],A [Ala] ⇒ A [Ala]	NP_000375.2
194	ENSP00000233242.1	rs201270852	21013220	0.0008	missense	C/T	R[Arg] ⇒ W [Trp]	NP_000375.2
195	ENSP00000233242.1	rs201368319	21038109	0.0004	missense	A/G	Y [Tyr] ⇒ C [Cys]	NP_000375.2
196	ENSP00000233242.1	rs201448956	21003274	0.0002	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
197	ENSP00000233242.1	rs201461940	21008695	0.0002	missense	A/C	T [Thr] ⇒ P [Pro]	NP_000375.2
198	ENSP00000233242.1	rs201519465	21010096	0.0002	missense	A/G	I [Ile] ⇒ V [Val]	NP_000375.2
199	ENSP00000233242.1	rs201595604	21025036	0.0002	missense	C/A	R[Arg] ⇒ H [His]	NP_000375.2
200	ENSP00000233242.1	rs201606169	21037137	0.0004	missense	C/A	R[Arg] ⇒ H [His]	NP_000375.2
201	ENSP00000233242.1	rs201764222	21019889	0.0002	missense	G/T	V [Val] ⇒ F [Phe]	NP_000375.2
202	ENSP00000233242.1	rs201874707	21008757	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
203	ENSP00000233242.1	rs201985284	21008617	0.0002	missense/	G/A, G/T	E [Glu] ⇒ K [Lys] ,E[Glu]	NP_000375.2
204	ENSP00000233242.1	rs201990496	21005124	0.0002	missense	C/G, C/T	S [Ser] ⇒ C [Cys] ,S [Ser] ⇒ F [Phe]	NP_000375.2
205	ENSP00000233242.1	rs202001155	21027991	0.0002	missense	G/A	R[Arg] ⇒ Q [Gln]	NP_000375.2
206	ENSP00000233242.1	rs202213088	21005706	0.0002	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
207	ENSP00000233242.1	rs202229735	21009840	----	missense	A/G, A/T	E [Glu] ⇒ G [Gly],E [Glu] ⇒ V [Val]	NP_000375.2
208	ENSP00000233242.1	rs267599180	21006843	0.0008	missense	C/T	S [Ser] ⇒ F [Phe]	NP_000375.2
209	ENSP00000233242.1	rs367788462	21028530	0.0004	missense/	G/A	E [Glu] ⇒ E [Glu]	NP_000375.2
210	ENSP00000233242.1	rs368779816	21013297	0.0002	missense	C/T	T[Thr] ⇒ M [Met]	NP_000375.2
211	ENSP00000233242.1	rs369430233	21007848	0.0002	missense	T/A, T/C	M[Met] ⇒ K [Lys], M[Met] ⇒ T [Thr]	NP_000375.2
212	ENSP00000233242.1	rs369472561	21008013	0.0002	missense	G/T	G[Gly] ⇒ V [Val]	NP_000375.2
213	ENSP00000233242.1	rs371987644	21016525	0.0002	missense	A/C	E [Glu] ⇒ D [Asp]	NP_000375.2
214	ENSP00000233242.1	rs372035579	21009927	0.006	missense	T/C	L [Leu] ⇒ P [Pro]	NP_000375.2
215	ENSP00000233242.1	rs372154910	21013525	0.0002	missense	G/A	R[Arg] ⇒ Q [Gln]	NP_000375.2
216	ENSP00000233242.1	rs373272476	21009291	0.0002	missense	T/C	M[Met] ⇒ T [Thr]	NP_000375.2
217	ENSP00000233242.1	rs373995960	21006007	0.0002	missense	G/A, G/T	A [Ala] ⇒ T [Thr] ,A [Ala] ⇒ S [Ser]	NP_000375.2
218	ENSP00000233242.1	rs374402922	21011669	0.0002	missense	A/T	Q[Gln] ⇒ H [His]	NP_000375.2
219	ENSP00000233242.1	rs374411400	21005850	0.0002	missense	A/G	K [Lys] ⇒ R [Arg]	NP_000375.2
220	ENSP00000233242.1	rs375294575	21016458	0.0002	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
221	ENSP00000233242.1	rs375855688	21009109	0.0004	missense	G/C	V [Val] ⇒ L [Leu]	NP_000375.2
222	ENSP00000233242.1	rs376250460	21008385	0.0002	missense	T/C, G/T	V [Val] ⇒ A [Ala], T [Thr] ⇒ T [Thr]	NP_000375.2
223	ENSP00000233242.1	rs528503563	21011474	0.0002	missense	A/G	T [Thr] ⇒ A [Ala]	NP_000375.2
224	ENSP00000233242.1	rs530171166	21032507	0.0004	missense	G/A	R[Arg] ⇒ H [His]	NP_000375.2
225	ENSP00000233242.1	rs530601244	21012550	0.0002	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
226	ENSP00000233242.1	rs530659716	21006064	0.0002	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
227	ENSP00000233242.1	rs531341535	21015435	0.0002	missense	T/A	L [Leu] ⇒ H [His]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
228	ENSP00000233242.1	rs532531549	21009513	0.0006	missense	G/C	G[Gly] ⇒ A [Ala]	NP_000375.2
229	ENSP00000233242.1	rs532575330	21002364	0.0002	missense	G/A	C[Cys] ⇒ Y [Tyr]	NP_000375.2
230	ENSP00000233242.1	rs533551460	21002457	0.0002	missense	T/G	F [Phe] ⇒ C [Cys]	NP_000375.2
231	ENSP00000233242.1	rs534116066	21008505	0.0008	missense	G/A, G/T	G[Gly] ⇒ D [Asp], G[Gly] ⇒ V [Val]	NP_000375.2
232	ENSP00000233242.1	rs534924739	21012249	0.0002	missense	G/T	G[Gly] ⇒ V [Val]	NP_000375.2
233	ENSP00000233242.1	rs535011070	21006202	0.0002	missense	C/A, C/T	L [Leu] ⇒ I [Ile] ,L [Leu] v F [Phe]	NP_000375.2
234	ENSP00000233242.1	rs535977033	21006385	0.0002	missense	T/A, C/T	S [Ser] ⇒ T [Thr], V [Val] ⇒ V [Val]	NP_000375.2
235	ENSP00000233242.1	rs536328155	21025043	0.0002	missense	T/C	Y [Tyr] ⇒ H [His]	NP_000375.2
236	ENSP00000233242.1	rs536537735	21037107	0.0002	missense	A/G	K [Lys] ⇒ R [Arg]	NP_000375.2
237	ENSP00000233242.1	rs537054558	21014559	0.0004	missense	G/A	S [Ser] ⇒ N [Asn]	NP_000375.2
238	ENSP00000233242.1	rs539614975	21015499	0.0002	missense	C/T	P [Pro] ⇒ S [Ser]	NP_000375.2
239	ENSP00000233242.1	rs540075453	21001479	0.0002	UTR-3	A/G	----------	-----
240	ENSP00000233242.1	rs540192015	21002629	0.0002	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
241	ENSP00000233242.1	rs540387864	21007339	0.0002	missense	T/A, T/C	Y [Tyr] ⇒ N [Asn]	NP_000375.2
242	ENSP00000233242.1	rs540530658	21038007	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
243	ENSP00000233242.1	rs540978306	21016481	0.0004	missense	A/G	Q [Gln] ⇒ R [Arg]	NP_000375.2
244	ENSP00000233242.1	rs542107305	21026829	0.0002	missense	G/A	V[Val] ⇒ M [Met]	NP_000375.2
245	ENSP00000233242.1	rs543698457	21035637	0.0004	missense/	T/A	H [His] ⇒ Q [Gln]	NP_000375.2
246	ENSP00000233242.1	rs544303971	21002250	0.0002	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
247	ENSP00000233242.1	rs544521341	21007174	0.0004	missense	A/G, C/T	K [Lys] ⇒ E [Glu]T [Thr] ⇒ T [Thr]	NP_000375.2
248	ENSP00000233242.1	rs546392030	21009396	0.0002	missense	T/C	L [Leu]⇒ S [Ser]	NP_000375.2
249	ENSP00000233242.1	rs546684381	21032559	0.0002	missense	G/A, G/T	V [Val] ⇒ I [Ile] V [Val] ⇒ F [Phe]	NP_000375.2
250	ENSP00000233242.1	rs546747242	21007551	0.0002	missense	A/G	N [Asn] ⇒ S [Ser]	NP_000375.2
251	ENSP00000233242.1	rs547372326	21009611	0.0002	missense	C/G	F [Phe] ⇒ L [Leu]	NP_000375.2
252	ENSP00000233242.1	rs548108916	21027992	0.0002	Missense/	C/A	R [Arg] ⇒ R [Arg]	NP_000375.2
253	ENSP00000233242.1	rs548321407	21027842	0.0002	missense	G/C	A [Ala] ⇒ P [Pro]	NP_000375.2
254	ENSP00000233242.1	rs549741449	21012192	0.0004	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2
255	ENSP00000233242.1	rs549983245	21012114	0.0004	missense	A/T	N [Asn] ⇒ I [Ile]	NP_000375.2
256	ENSP00000233242.1	rs551773082	21012643	0.0002	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
257	ENSP00000233242.1	rs552302745	21002383	0.0002	missense	T/G	L [Leu] ⇒ V [Val]	NP_000375.2
258	ENSP00000233242.1	rs552447556	21008910	0.0002	missense	T/G	L [Leu] ⇒ R [Arg]	NP_000375.2
259	ENSP00000233242.1	rs552692138	21007029	0.0002	missense	T/A	M[Met] ⇒ K [Lys]	NP_000375.2
260	ENSP00000233242.1	rs553689397	21044034	0.0006	UTR-5	---/C	----------	-----
261	ENSP00000233242.1	rs554045156	21009330	0.0002	missense	G/A	R[Arg] ⇒ Q [Gln]	NP_000375.2
262	ENSP00000233242.1	rs554648284	21009718	0.0002	missense	G/T	V [Val] ⇒ F [Phe]	NP_000375.2
263	ENSP00000233242.1	rs556582055	21010805	0.0002	Missense/	C/G	D[Asp] ⇒ E [Glu]	NP_000375.2

No.	ENSP ID	SNP ID	Chromo. Position	MAF Value	Type of SNP	Nuc. Chan	AA Change	Protein ID
264	ENSP00000233242.1	rs557580138	21011760	0.0002	missense	C/A	A [Ala] ⇒ D [Asp]	NP_000375.2
265	ENSP00000233242.1	rs558626935	21015538	0.0002	missense	A/G	T [Thr] ⇒ A [Ala]	NP_000375.2
266	ENSP00000233242.1	rs559224730	21005385	0.0002	missense	C/T	P [Pro] ⇒ L [Leu]	NP_000375.2
267	ENSP00000233242.1	rs561304247	21002277	0.0002	missense	T/C	M[Met] ⇒ T [Thr]	NP_000375.2
268	ENSP00000233242.1	rs561774487	21009939	0.0008	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
269	ENSP00000233242.1	rs562608061	21032376	0.0002	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
270	ENSP00000233242.1	rs562624518	21008918	0.0002	missense	T/G	F [Phe] ⇒ L [Leu]	NP_000375.2
271	ENSP00000233242.1	rs564030306	21023012	0.0002	missense	G/A	V[Val] ⇒ M [Met]	NP_000375.2
272	ENSP00000233242.1	rs565114521	21012510	0.0002	missense	T/C	L [Leu] ⇒ P [Pro]	NP_000375.2
273	ENSP00000233242.1	rs565184018	21006055	0.0002	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
274	ENSP00000233242.1	rs565963363	21013304	0.0002	missense	C/G, C/T	L [Leu] ⇒ V [Val]	NP_000375.2
275	ENSP00000233242.1	rs566501266	21013243	0.0002	missense	C/A	T [Thr] ⇒ N [Asn]	NP_000375.2
276	ENSP00000233242.1	rs566674173	21023538	0.0006	missense	T/A	L [Leu] ⇒ Q [Gln]	NP_000375.2
277	ENSP00000233242.1	rs568054660	21002743	0.0002	missense	G/A	V [Val] ⇒ I [Ile]	NP_000375.2
278	ENSP00000233242.1	rs568176871	21009751	0.0002	missense	A/T	T [Thr] ⇒ S [Ser]	NP_000375.2
279	ENSP00000233242.1	rs569794936	21009667	0.0002	missense	G/C	V [Val] ⇒ L [Leu]	NP_000375.2
280	ENSP00000233242.1	rs570097888	21012169	0.0002	missense	G/A	E [Glu] ⇒ K [Lys]	NP_000375.2
281	ENSP00000233242.1	rs570383610	21011700	0.0002	missense	G/T	S [Ser] ⇒ I [Ile]	NP_000375.2
282	ENSP00000233242.1	rs570782024	21002841	0.0002	missense	T/C	I [Ile] ⇒ T [Thr]	NP_000375.2
283	ENSP00000233242.1	rs571485213	21005512	0.0008	missense	C/T	L [Leu] ⇒ F [Phe]	NP_000375.2
284	ENSP00000233242.1	rs571626569	21009249	0.0018	missense	G/T	G[Gly] ⇒ V [Val]	NP_000375.2
285	ENSP00000233242.1	rs571728141	21012283	0.0002	missense	G/A	G [Gly] ⇒ S [Ser]	NP_000375.2
286	ENSP00000233242.1	rs572653083	21011094	0.0002	missense	G/A	G [Gly] ⇒ E [Glu]	NP_000375.2
287	ENSP00000233242.1	rs572869977	21042384	0.0002	missense	A/G	S [Ser] ⇒ G [Gly]	NP_000375.2
288	ENSP00000233242.1	rs573286918	21011926	0.0002	missense	G/A	G [Gly] ⇒ S [Ser]	NP_000375.2
289	ENSP00000233242.1	rs573308525	21037144	0.0002	missense	C/T	P [Pro] ⇒ S [Ser]	NP_000375.2
290	ENSP00000233242.1	rs573670976	21006319	0.0002	missense	G/A	A [Ala] ⇒ T [Thr]	NP_000375.2
291	ENSP00000233242.1	rs574659380	21002013	0.0002	missense	C/A	S [Ser] ⇒ Y [Tyr]	NP_000375.2
292	ENSP00000233242.1	rs574791609	21003113	0.0002	Missense/	G/A	K[Lys] ⇒ K [Lys]	NP_000375.2
293	ENSP00000233242.1	rs577157373	21008554	0.0002	missense	G/A, G/C	D[Asp] ⇒ N [Asn]	NP_000375.2
294	ENSP00000233242.1	rs577157440	21002175	0.0002	missense	T/C	V [Val] ⇒ A [Ala]	NP_000375.2
295	ENSP00000233242.1	rs577437570	21009351	0.0002	missense	C/T	A [Ala] ⇒ V [Val]	NP_000375.2

Supplementary Table III.- *APOB* SNPs present in non-coding region with their Regulome DB scores.

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
1	chr2:21224372..21224373	rs72654430	4	0.60906	3 UTR
2	chr2:21224421..21224422	rs12720763	4	0.60906	3 UTR
3	chr2:21224422..21224423	rs142151703	4	0.60906	3 UTR
4	chr2:21224590..21224591	rs72654428	4	0.60906	3 UTR

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
5	chr2:21044060	rs1800480	4	0.60906	5 UTR
6	chr2:21226400..21226401	rs141644795	4	0.60906	intronic
7	chr2:21226486..21226487	rs187296678	5	0.58955	intronic
8	chr2:21226873..21226874	rs542404934	4	0.60906	intronic
9	chr2:21227850..21227851	rs72654410	5	0.58955	intronic
10	chr2:21227935..21227936	rs139704306	4	0.60906	intronic
11	chr2:21235564..21235565	rs12713771	3a	0.24403	intronic
12	chr2:21235604..21235605	rs566982035	3a	1	intronic
13	chr2:21235618..21235619	rs72653080	3a	0.65649	intronic
14	chr2:21235809..21235810	rs183128193	3a	0.75931	intronic
15	chr2:21235827..21235828	rs118002564	4	0.60906	intronic
16	chr2:21235847..21235848	rs12720844	5	0.58955	intronic
17	chr2:21235943..21235944	rs190865121	5	0.58955	intronic
18	chr2:21235981..21235982	rs12713776	4	0.60906	intronic
19	chr2:21236481..21236482	rs488329	2b	0.66759	intronic
20	chr2:21236732..21236733	rs143870044	3a	0.75713	intronic
21	chr2:21236739..21236740	rs536749646	3a	0.50689	intronic
22	chr2:21236831..21236832	rs147270037	5	0.955	intronic
23	chr2:21237145..21237146	rs139828082	5	0.58955	intronic
24	chr2:21237236..21237237	rs187401851	5	0.9375	intronic
25	chr2:21237296..21237297	rs569993957	4	0.60906	intronic
26	chr2:21237674..21237675	rs534008508	4	0.60906	intronic
27	chr2:21237851..21237852	rs78580569	4	0.60906	intronic
28	chr2:21238515..21238516	rs182561283	5	0.58955	intronic
29	chr2:21238763..21238764	rs146764589	4	0.60906	intronic
30	chr2:21238797..21238798	rs140225127	4	0.60906	intronic
31	chr2:21238825..21238826	rs541923839	4	0.60906	intronic
32	chr2:21238983..21238984	rs150956240	5	0.58955	intronic
33	chr2:21238992..21238993	rs12720800	5	0.58955	intronic
34	chr2:21239082..21239083	rs12720802	5	0.58955	intronic
35	chr2:21239583..21239584	rs12713868	3a	0.35836	intronic
36	chr2:21239584..21239585	rs12713869	3a	0.29248	intronic
37	chr2:21239594..21239595	rs12713870	5	0.225	intronic
38	chr2:21239610..21239611	rs12713871	5	0.58955	intronic
39	chr2:21239622..21239623	rs13306202	5	0.58955	intronic
40	chr2:21239652..21239653	rs191090985	5	0.88556	intronic
41	chr2:21239682..21239683	rs545349871	5	0.58955	intronic
42	chr2:21239711..21239712	rs187182863	5	0.58955	intronic
43	chr2:21239718..21239719	rs12720819	5	0.58955	intronic
44	chr2:21239727..21239728	rs571625076	5	0.58955	intronic

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
45	chr2:21239771..21239772	rs12720817	5	0.58955	intronic
46	chr2:21239817..21239818	rs549394363	5	0.99722	intronic
47	chr2:21239913..21239914	rs138999114, rs573286852	5	0.83105	intronic
48	chr2:21239920..21239921	rs112745407	3a	0.63742	intronic
49	chr2:21240044..21240045	rs115755929	3a	0.47482	intronic
50	chr2:21240110..21240111	rs144876159	4	0.60906	intronic
51	chr2:21240125..21240126	rs189046815	3a	1	intronic
52	chr2:21240147..21240148	rs180931378	3a	0.8738	intronic
53	chr2:21240184..21240185	rs535824398	4	0.60906	intronic
54	chr2:21240407..21240408	rs572186909	2c	0.58033	intronic
55	chr2:21240502..21240503	rs76693756	4	0.60906	intronic
56	chr2:21240537..21240538	rs12713911	4	0.60906	intronic
57	chr2:21240937..21240938	rs147100637	4	0.60906	intronic
58	chr2:21240958..21240959	rs374897675	4	0.60906	intronic
59	chr2:21241210..21241211	rs147720589	5	0.58955	intronic
60	chr2:21241289..21241290	rs12720825	4	0.60906	intronic
62	chr2:21241301..21241302	rs12720821	3a	0.75713	intronic
63	chr2:21241641..21241642	rs181346861	4	0.60906	intronic
64	chr2:21242235..21242236	rs560821395	4	0.60906	intronic
65	chr2:21242328..21242329	rs185721706	4	0.60906	intronic
66	chr2:21242795..21242796	rs72653069	5	0.72379	intronic
67	chr2:21242852..21242853	rs561979139	5	0.58955	intronic
68	chr2:21243043..21243044	rs12720785	4	0.60906	intronic
69	chr2:21243400..21243401	rs139313355	2c	1	intronic
70	chr2:21243967..21243968	rs114471564	5	1	intronic
71	chr2:21244126..21244127	rs563121607	4	0.60906	intronic
72	chr2:21244127..21244128	rs531867731	4	0.60906	intronic
73	chr2:21244259..21244260	rs191606984	4	0.60906	intronic
74	chr2:21244655..21244656	rs542314665	5	0.85033	intronic
75	chr2:21244983..21244984	rs111825526	4	0.60906	intronic
76	chr2:21245081..21245082	rs564571506	4	0.60906	intronic
77	chr2:21245203..21245204	rs149677861	4	0.60906	intronic
78	chr2:21245365..21245366	rs145511260	5	0.99471	intronic
79	chr2:21245560..21245561	rs372178138	5	0.58955	intronic
80	chr2:21245630..21245631	rs558233757	4	0.60906	intronic
81	chr2:21246381..21246382	rs72653066	3a	0.4701	intronic
82	chr2:21246960..21246961	rs142557833	5	0.58955	intronic
83	chr2:21247013..21247014	rs529641466	5	0.87837	intronic
84	chr2:21247022..21247023	rs547734093	5	0.99722	intronic

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
85	chr2:21247174..21247175	rs151334814	3a	0.93733	intronic
86	chr2:21247286..21247287	rs12720839	3a	0.41664	intronic
87	chr2:21247728..21247729	rs78341608	3a	0.54807	intronic
88	chr2:21247764..21247765	rs72653065	4	0.60906	intronic
89	chr2:21248100..21248101	rs12720759	4	0.60906	intronic
90	chr2:21248236..21248237	rs566255530	4	0.60906	intronic
91	chr2:21248349..21248350	rs116255207	4	0.60906	intronic
92	chr2:21248484..21248485	rs186200586	3a	0.66969	intronic
93	chr2:21249179..21249180	rs531424311	4	0.60906	intronic
94	chr2:21249267..21249268	rs552569557	4	0.60906	intronic
95	chr2:21249403..21249404	rs190552864	4	0.60906	intronic
96	chr2:21249415..21249416	rs548617467	3a	0.30476	intronic
97	chr2:21249445..21249446	rs186828301	3a	0.144	intronic
98	chr2:21249870..21249871	rs78941963	3a	0.63742	intronic
99	chr2:21250254..21250255	rs78504102	4	0.60906	intronic
100	chr2:21250596..21250597	rs145100968	2b	0.79558	intronic
101	chr2:21251527..21251528	rs12714215	5	1	intronic
102	chr2:21252095..21252096	rs182312395	3a	0.59935	intronic
103	chr2:21252096..21252097	rs570904180	2b	0.59816	intronic
104	chr2:21252102..21252103	rs548067874	2b	0.33595	intronic
105	chr2:21252103..21252104	rs12720840	2b	0.549	intronic
106	chr2:21252183..21252184	rs184358787	3a	0.61437	intronic
107	chr2:21252303..21252304	rs146536216	4	0.60906	intronic
108	chr2:21252338..21252339	rs113010686	3a	0.6352	intronic
109	chr2:21252364..21252365	rs572597107, rs67556837	3a	0.8507	intronic
110	chr2:21252715..21252716	rs12720765	3a	0.33881	intronic
111	chr2:21253196..21253197	rs554117371	5	0.58955	intronic
112	chr2:21253200..21253201	rs6727706	5	0.58955	intronic
113	chr2:21253955..21253956	rs115200831	5	0.85033	intronic
114	chr2:21254108..21254109	rs140830032	5	0.58955	intronic
115	chr2:21254116..21254117	rs572158013	5	1	intronic
116	chr2:21254168..21254169	rs12720766	5	0.58955	intronic
117	chr2:21254384..21254385	rs12720769	4	0.60906	intronic
118	chr2:21254563..21254564	rs12720768	4	0.60906	intronic
119	chr2:21254735..21254736	rs201106138, rs377355276	2c	0.9925	intronic
120	chr2:21254738..21254739	rs143452815	2c	0.9925	intronic
121	chr2:21254837..21254838	rs114883958	5	0.58955	intronic
122	chr2:21255165..21255166	rs12714224	5	0.58955	intronic
123	chr2:21255763..21255764	rs184507838	4	0.60906	intronic

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
124	chr2:21255947..21255948	rs12714226	5	0.58955	intronic
125	chr2:21256002..21256003	rs576038819	5	0.58955	intronic
126	chr2:21256065..21256066	rs540340913	3a	0.75409	intronic
127	chr2:21256404..21256405	rs72653061	4	0.60906	intronic
128	chr2:21256412..21256413	rs12720811	4	0.60906	intronic
129	chr2:21256488..21256489	rs78610189	5	0.98	intronic
130	chr2:21257257..21257258	rs569321939	3a	0.68065	intronic
131	chr2:21257376..21257377	rs12720793	4	0.60906	intronic
132	chr2:21257413..21257414	rs78652877	4	0.60906	intronic
133	chr2:21257627..21257628	rs182257545	4	0.60906	intronic
134	chr2:21257676..21257677	rs148944625	4	0.60906	intronic
135	chr2:21257926..21257927	rs12720842	4	0.60906	intronic
136	chr2:21257975..21257976	rs114659957	4	0.60906	intronic
137	chr2:21257982..21257983	rs541372741	4	0.60906	intronic
138	chr2:21259328..21259329	rs12720853	4	0.60906	intronic
139	chr2:21259739..21259740	rs566540157	4	0.60906	intronic
140	chr2:21260189..21260190	rs191618417	2b	0.78832	intronic
141	chr2:21260298..21260299	rs13306197	4	0.60906	intronic
142	chr2:21260360..21260361	rs12720846	4	0.60906	intronic
143	chr2:21260371..21260372	rs139266444	4	0.60906	intronic
144	chr2:21261197..21261198	rs530005654	3a	0.57256	intronic
145	chr2:21261211..21261212	rs559799871	3a	0.57256	intronic
146	chr2:21261242..21261243	rs12714239	4	0.60906	intronic
147	chr2:21261269..21261270	rs569472006	3a	0.8227	intronic
148	chr2:21261845..21261846	rs142229577	2b	0.67017	intronic
149	chr2:21261997..21261998	rs12720796	3a	0.6851	intronic
150	chr2:21262034..21262035	rs146194270	3a	0.58328	intronic
151	chr2:21262190..21262191	rs12720797	2b	0.7516	intronic
152	chr2:21262314..21262315	rs12720798	3a	0.75713	intronic
153	chr2:21262374..21262375	rs12720794	4	0.60906	intronic
154	chr2:21262379..21262380	rs12720795	4	0.60906	intronic
155	chr2:21262667..21262668	rs12714248	5	0.58955	intronic
156	chr2:21262746..21262747	rs142459501	4	0.60906	intronic
157	chr2:21262755..21262756	rs12720789	3a	0.8507	intronic
158	chr2:21262771..21262772	rs555323073	4	0.60906	intronic
159	chr2:21262870..21262871	rs12720788	3a	0.57268	intronic
160	chr2:21262916..21262917	rs12720786	3a	0.39922	intronic
161	chr2:21263233..21263234	rs140139277	3a	0.57429	intronic
162	chr2:21263356..21263357	rs562375034	4	0.60906	intronic
163	chr2:21263636..21263637	rs1367116	5	0.97858	intronic

S No.	Chromosome location	dbSNP IDs	Rank	Score	SNP type
164	chr2:21263766..21263767	rs9282605	5	0.58955	intronic
165	chr2:21264077..21264078	rs548513084	4	0.60906	intronic
166	chr2:21264106..21264107	rs563016779	4	0.60906	intronic
167	chr2:21264181..21264182	rs191372114	4	0.60906	intronic
168	chr2:21264423..21264424	rs144729882	4	0.60906	intronic
169	chr2:21264998..21264999	rs531023775	2b	0.81166	intronic
170	chr2:21265897..21265898	rs12714268	2a	0.79886	intronic
171	chr2:21265934..21265935	rs72653055	4	0.60906	intronic
172	chr2:21266362..21266363	rs570818403	4	0.60906	intronic
173	chr2:21266658..21266659	rs12720762	2b	0.66261	Intronic

Supplementary Table IV.- Extracted non-coding SNPs of *APOB* through Regulome DB and SNP info.

dbSNP IDs	Regulome DB rank	Description	SNP info
rs12714268	2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak	
rs488329	2b	TF binding + any motif + DNase Footprint + DNase peak	
rs145100968	2b	-----	
rs570904180	2b	-----	
rs548067874	2b	-----	
rs12720840	2b	-----	
rs191618417	2b	-----	
rs142229577	2b	-----	
rs12720797	2b	-----	
rs531023775	2b	-----	
rs12720762	2b	-----	Y (TFBS)
rs572186909	2c	TF binding + matched TF motif + DNase peak	
rs139313355	2c	------	
rs201106138, rs377355276	2c	-----	
rs143452815	2c	-----	

TFBS, Transcription factor binding cite; TF, Transcription factor.