Applied Biological Research 25(3): 348-360; (2023)
DOI: 10.5958/0974-4517.2023.00043.5

ANALYSIS OF MISSENSE SNPs IN HUMAN LRP5 GENE BY *in silico* APPROACH

Mrinmoyee Sengupta and Mitali Das*

Laboratory of Molecular Cell Biology and Genetics, Department of Zoology, University of Gour Banga, Malda - 732 101 West Bengal (India)
*****e-mail: *mitalibhu@gmail.com*

(Received 13 April, 2023; accepted 21 July, 2023)

ABSTRACT

Low-density lipoprotein receptor-related protein 5 (LRP5) is a protein, encoded by the LRP5 gene in humans. LRP5, the key component of the LRP5/LRP6/ Frizzled co-receptor group is involved in canonical Wnt pathway, a regulator of bone metabolism. Mutations in LRP5 can lead to considerable changes in bone mass following bone diseases. Our study aimed to predict the effect of missense SNPs in human LRP5 gene by in silico methods using publicly available online software tools and database. We predicted the functional effects of mutations by using VEP, PROVEAN, SNPs & GO, and PANTHER, effects on the stability of the protein by I-Mutant 2.0 and on the 3-Dstructure of the protein by Project HOPE. We used GeneMANIA to predict the interaction of LRP5 with 20 other genes. Over all this is a comprehensive study provides all the available information about the clinically significant missense SNPs of LRP5 at a glance.

Keywords: Bone diseases, 3D structural effect, functional effect, *in silico*, LRP5, missense SNPs analysis

INTRODUCTION

Low-density lipoprotein receptor-related protein 5 (LRP5) have received attention as a candidate gene for bone related diseases from mid 90s when the chromosomal region 11q12-1 3 was found linked with osteoporosis-pseudoglioma and high bone mass of human subjects (Gong *et al.*, 1996; Lara-Castilo and Johnson, 2015). Since then, many scientific groups around the world identified several single nucleotide polymorphisms (SNPs) of LRP5 associated with bone related diseases along with other ones. LRP5 protein in humans is encoded by LRP5 gene. LRP5 acts as a co-receptor with LRP6 and the Frizzled protein family members for transducing signals by Wnt proteins through the canonical Wnt pathway (Williams and Insogna, 2009). This pathway regulates the metabolism of osteoblasts (Laine *et al.,* 2011). Lrp5 gene is located on human chromosome 11q13 (Hey e*t al.,* 1998, Guo *et al.*, 2006) and possesses 23 coding exons (Laine *et al.,* 2011). LRP5 is a transmembrane receptor protein made up of 1615 amino acids (Norwitz e*t al.,* 2019). The LRP5 protein contains an extracellular domain containing a signal peptide, an intracellular domain, a single transmembrane domain, four EGF-like domains intersected by four β-propeller domains, and three low density lipoprotein receptor (LDLR) repeats. Each β-propeller domain contains six blades and five YWTD (tyrosine-tryptophan-threonine-aspartic acid) motifs (Mao *et al*., 2011).

As per the NCBI's dictionary of genetics terms, SNPs are defined as a DNA sequence variation that occurs when a single nucleotide in the genome sequence is altered and the particular alterations present in at least 1% population (Shastry, 2009; Keats and Sherman 2013). SNPs are functionally classified into various types such as 3′splice site, 3'UTR (untranslated region), 5′splice site, 5′UTR, coding synonymous, frameshift, intron, missense, nonsense, and gained. Among these, missense SNPs result in amino acid substitutions. The impact of such SNP mediated amino acid changes on protein structure and function is necessary to understand the complex pathophysiology of the associated human diseases (Teng *et al.,* 2010). SNPs can be used as biomarkers for predicting the

genetic predisposition of diseases. Recently the association of SNPs in LRP5 gene with diseases has been investigated (Xu *et al.,* 2014; Liu *et al.,* 2017; Pekkinen *et al.,* 2017; Wang *et al.,* 2020). However, the perusal of literature has revealed that the *in silico* analysis of the damaging effects of missense SNPs in LRP5 gene have not been reported as yet. The present study was aimed to find missense SNPs in LRP5 gene, the related amino acid substitutions and their possible effects on protein attributes depending on the altered physiochemical properties (such as size, charge, hydrophobicity, structure, domain, and conservation) of the amino acids by using bioinformatic methods.

MATERIALS AND METHODS

Mining for protein and SNP database of LRP5 gene

The National Centre for Biotechnology Information (NCBI) dbSNP database was utilized to extract the reported SNPs in LRP5 gene (NCBI Gene ID: 4041) in April 2022 (*https://www.ncbi.nlm.nih.gov/snp*). We have used "LRP5" as search term and assigned filters to limit the search results into missense category. The SNP accession numbers of missense SNPs were retrieved for further use. Missense SNPs were selected because they may result into amino acid substitutions in wild type protein sequence and thus affect the protein structure and function. The missense SNPs were further filtered depending on the clinical significance and selected as benign, likely benign, likely pathogenic and pathogenic. The amino acid sequence of the protein (UniProt Accession No: O75197) encoded by LRP5 gene was obtained from the UniProt database (*https://www.uniprot.org/).*

Predicting the effect of amino acid change on protein structure and function

The effects of missense SNPs on LRP5 protein were predicted by several freely available online software tools (Abdelraheem *et al*., 2016; Nimir *et al.*, 2017). The SNP Accession numbers of all the SNPs were uploaded one by one into Variant Effect Predictor (VEP) (*http://www.ensembl.org/Tools/VEP*) (Martin et al., 2023) and enabled "SIFT" (Sorting Intolerant from Tolerant) *(http://siftdna.org/www/SIFT_dbSNP.html*) (Vaser *et al.,* 2016) and "PolyPhen" (Polymorphism Phenotyping v2) (*http://genetics.bwh.harvard.edu/pph2/*) (Ueki et al., 2017) and the program was run. SIFT and PolyPhen estimate the effects of an amino acid substitution on protein structure and function depending on sequence homology, physical properties of amino acids and comparative physical properties. These two algorithms provided the respective predictions of functional significance of each SNP. VEP provides a SIFT and PolyPhen prediction and score. Depending on SIFT prediction, both deleterious and tolerated variants were selected for next level of analysis. The particular amino acid substitution for each SNP was uploaded to three other software *viz*., PROVEAN (Protein Variation Effect Analyzer) *(http://provean.jcvi.org/index.php*) (Sandell and Sharp 2022), SNPs & GO *(http://snps.biofold.org/snps-and-go/snps-and-go.html*) (Manfredi *et al.,* 2022) and PANTHER (Protein Analysis Through Evolutionary Relationships) *(http://www.pantherdb.org/tools/csnpScoreForm.jsp*) (Mi *et al*., 2021) to get their predictions and scores. SNP & GO predicts the relatedness of SNP with a disease. PROVEAN and PANTHER predict the possible effect of an amino acid substitution on protein activity. The effect of these SNPs on protein stabilization was determined by using I-Mutant 2.0 (*http://folding.biofold.org/i-mutant/i-mutant2.0.html*) (Lim *et al.,* 2021). Beside, the effect of amino acid substitutions on 3D structures of proteins was predicted using an automatic program project HOPE *(http://www.cmbi.ru.nl/hope/method/)* (Venselaar *et al.,* 2010).

Predicting the genetic interactome of LRP5

To delineate the genetic interactome of LRP5 as a candidate gene, the GeneMANIA web server (*https://genemania.org/*) (Franz et al., 2018) was used. LRP5 was uploaded as our search term and consequently a complex network of various genes interacting with LRP5 in terms of physical interactions, gene co-expression, predicted, co-localization, pathway, genetic interactions and shared protein domains were retrieved.

RESULTS AND DISCUSSION

Prediction of the impact of missense SNPs on LRP5 protein function and stability

The NCBI dbSNP database provided a total of 58334 SNPs for human LRP5 gene. Among these 1523 SNPs were filtered as missense. After using category-wise dual filtration with clinical significance, 19 benign, 16 likely benign, 21 likely pathogenic and 29 pathogenic missense SNPs were extracted from NCBI. The SNPs denoted by NCBI as "merged" from the above list were excluded, while the SNPs only under one category were included if they were repeated in two different categories at a time and proceeded with rest 69 missense SNPs (11 benign, 11 likely benign, 21 likely pathogenic and 26 pathogenic) for further analysis (Fig. 1). Finally, the data for 56 missense SNPs out of 69 were retrieved.

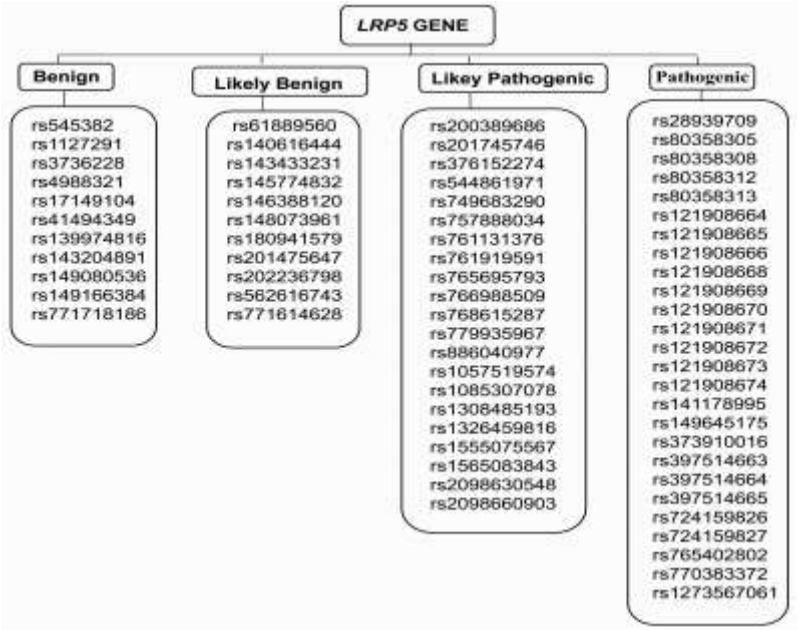


Fig. 1: LRP5 missense SNPs categorized according to clinical significance by NCBI

As per SIFT analysis 15 missense SNPs were found tolerated (Fig. 2) and 41 SNPs as deleterious (Fig. 3). The data retrieved from running 56 SNPs through all the online software are summarized in Table 1. The SNP IDs included under each and every functional category derived from SIFT, PolyPhen, PROVEAN, SNPs & GO, PANTHER are depicted in Fig. 3 and 4. The results of I-Mutant 2.0 software tool showed that 51 SNPs have decreasing effect on the stability of LRP5, while the other 5 SNPs have an increasing effect (Table 1).

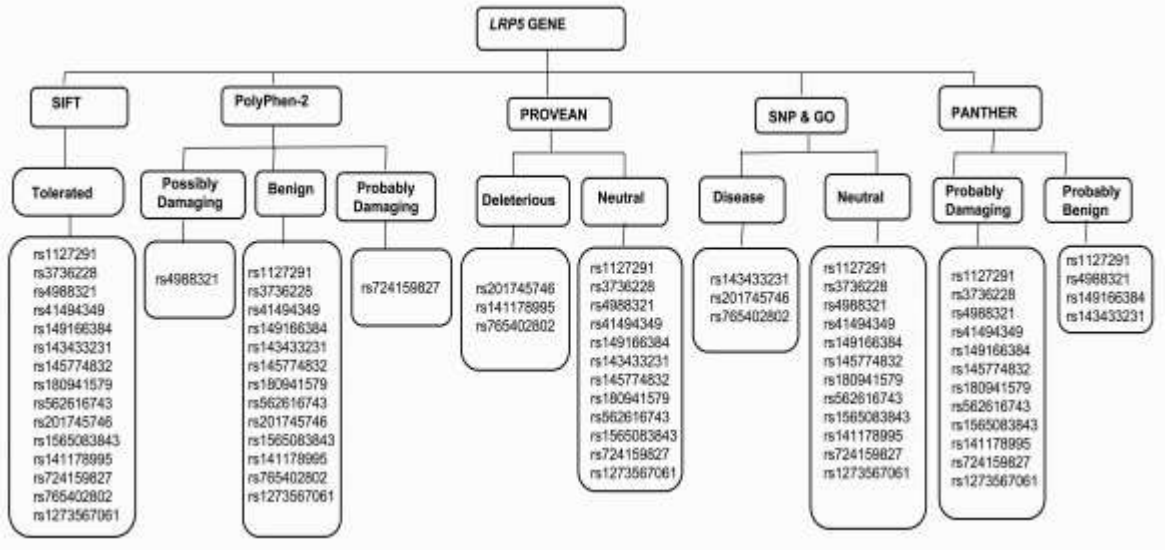


Fig. 2: LRP5 missense SNPs (tolerated) categorized by different online software as per the functional effect of the mutations

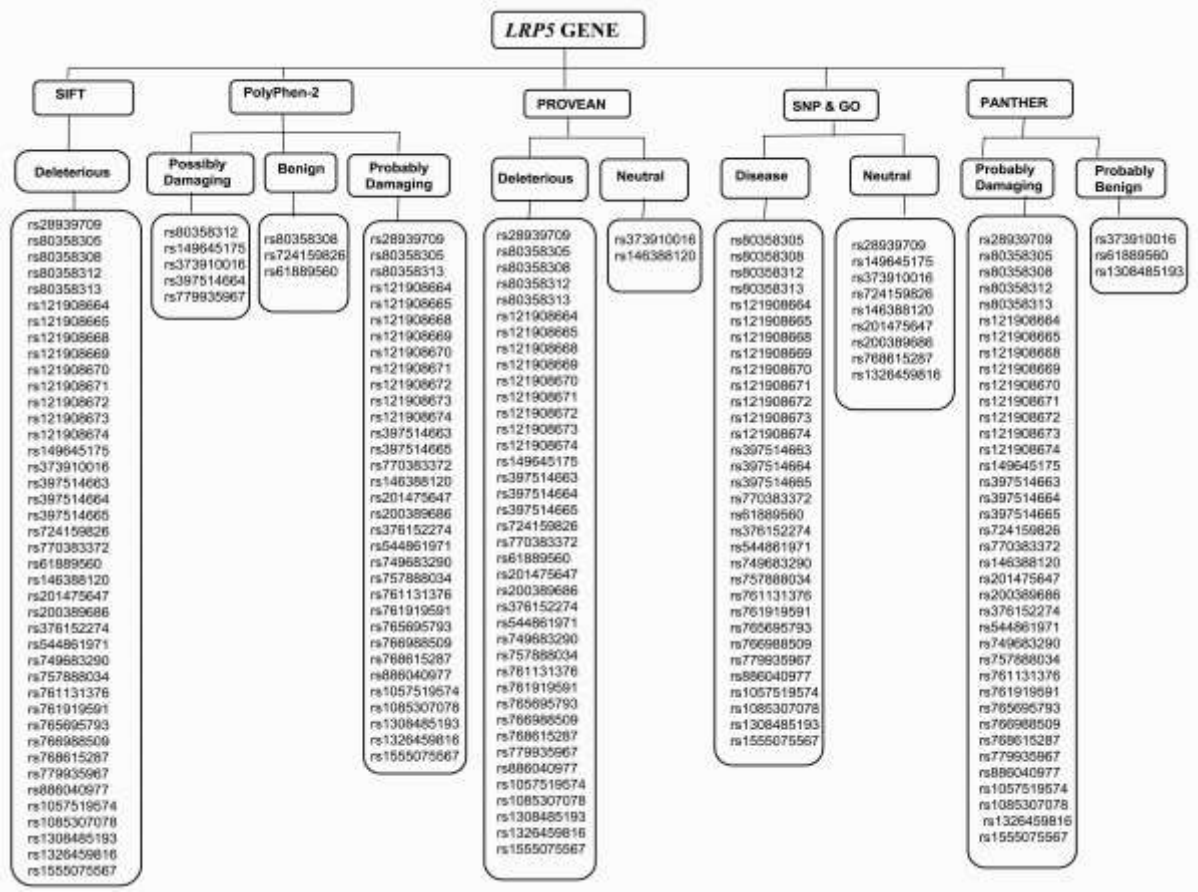


Fig. 3: LRP5 missense SNPs (deleterious) categorized by different online software according to the functional effect of the mutations

The SNPs are situated in different domains and repeat sequences of protein. The SNP IDs and their respective domains and repeats are enlisted in Table 4. The LRP5 domains EGF-like1, 3 and LDL-receptor class B13, 14, 15 do not contain any SNPs at all. The EGF-like 2 domain harbors the deleterious SNPrs80358313 resulted in glycine to arginine substitution (G610R) (Table 3). The substitution of the most flexible wild type glycine with a completely different amino acid may disrupt the protein function as the flexibility may be necessary for the purpose. The EGF-like4 domain contains deleterious SNP rs146388120 and rs768615287 resulted in D1233N, and C1253Y amino acid substitutions, respectively. In D1233N substitution, aspartic acid forms a salt bridge with lysine at position 1231. The difference in charge of the mutated residue will disturb the original ionic interaction. C1253Y substitution could possibly disrupt the metal-ion interactions, because tyrosine is larger as compared to cysteine. This may also disturb the protein function as ligand binding is important for protein function. The wild-type residue formed a cysteine bridge, important for stability of LRP5. Mutation disrupting this interaction drastically affects the 3D-structure of the protein.

The domain LDL-receptor class A 1 contains rs1326459816, resulted in the mutation of a cysteine into a tyrosine at position 1286. The mutated residue is not in contact with a metal; however, one of the neighbouring residues does make a metal-contact that might be affected by the nearby mutation. The wild-type residue is involved in a cysteine bridge whose disruption disturbs the stability of the protein. LDL-receptor class A2 domain contains rs3736228, rs149166384 having substitution A1330V and A1301T respectively. In both the cases, the wild type alanine is substituted by bigger amino acids leading to bumps which can disturb this domain and abolish its function. This particular domain contains only these two mutations categorized by SIFT as tolerated. LDL-receptor class A3

domain contains rs28939709 substitution of glutamic acid into a lysine at position 1367 (E1367K). The mutation introduces a residue with a charge opposite to the wild-type. This can cause repulsion with other residues in the protein or ligands which can disturb this domain and abolish its function.

Some mutations also occurred in other regions apart from the above-mentioned domains of protein. In case of rs149645175 (R1534G), introduction of a very flexible glycine affected the required rigidity of protein. The SNP rs544861971 (A67T) and rs1057519574 (D69Y) are located in

Table 1: Outcome of SIFT, POLYPHEN, PROVEAN, SNPs & GO, PANTHER and I-MUTANT software

S. No.	SNP ID	Nucleotide change	Amino acid change	SIFT		POLYPHEN-2		PROVEAN		SNP&GO	PANTHER	I-Mutant	
				Result	Score	Result	Score	Result	Score		Result	Result	Score
1	rs28939709	G>A	E1367K	Delt	0.04	PD	1	Delt	-3.122	Neutral	PD	D	5
2	rs80358305	C>T	L145F	Delt	0	PD	0.986	Delt	-3.002	Disease	PD	D	7
3	rs80358308	C>T	R444C	Delt	0	B	0.426	Delt	-6.409	Disease	PD	D	2
4	rs80358312	G>A	R570Q	Delt	0	PD	0.686	Delt	-2.974	Disease	PD	D	9
5	rs80358313	G>A,C,T	G610R	Delt	0	PD	0.993	Delt	-7.057	Disease	PD	D	9
6	rs121908664	G>A	R494Q	Delt	0	PD	0.995	Delt	-3.44	Disease	PD	D	9
7	rs121908665	C>T	R570W	Delt	0	PD	0.816	Delt	-6.336	Disease	PD	D	5
8	rs121908668	G>T	G171V	Delt	0	PD	1	Delt	-7.67	Disease	PD	D	6
9	rs121908669	G>C	G171R	Delt	0	PD	1	Delt	-6.818	Disease	PD	D	3
10	rs121908670	G>A	A242T	Delt	0	PD	0.999	Delt	-2.933	Disease	PD	D	4
11	rs121908671	G>A	A214T	Delt	0.01	PD	1	Delt	-2.812	Disease	PD	D	7
12	rs121908672	C>T	A214V	Delt	0.01	PD	1	Delt	-3.112	Disease	PD	D	4
13	rs121908673	C>T	T253I	Delt	0	PD	0.999	Delt	-5.058	Disease	PD	I	2
14	rs121908674	C>G,T	R752G	Delt	0	PD	0.996	Delt	-6.312	Disease	PD	D	9
15	rs149645175	C>G,T	R1534G	Delt	0	PD	0.598	Delt	-3.680	Neutral	PD	D	6
16	rs373910016	G>A,T	V454M	Delt	0	PD	0.675	Neutral	-1.882	Neutral	PB	D	8
17	rs397514663	C>T	T552M	Delt	0	PD	0.973	Delt	-3.815	Disease	PD	D	2
18	rs397514664	C>T	P382L	Delt	0	PD	0.907	Delt	-7.569	Disease	PD	D	5
19	rs397514665	C>G,T	T244M	Delt	0.01	PD	0.998	Delt	-4.674	Disease	PD	D	0
20	rs724159826	G>A,C	R1529S	Delt	0	B	0.326	Delt	-3.455	Neutral	PD	D	8
21	rs770383372	G>C,T	W167S	Delt	0	PD	1	Delt	-11.855	Disease	PD	D	9
22	rs61889560	G>A,C,T	R1036Q	Delt	0.03	B	0.051	Delt	-3.383	Disease	PB	D	6
23	rs146388120	G>A	D1233N	Delt	0.01	PD	0.999	Neutral	-2.196	Neutral	PD	D	3
24	rs201475647	C>T	P1504L	Delt	0	PD	1	Delt	-8.329	Neutral	PD	D	6
25	rs200389686	T>A,C	L1149Q	Delt	0	PD	0.949	Delt	-4.976	Neutral	PD	D	9
26	rs376152274	G>A	E441K	Delt	0.01	PD	0.975	Delt	-3.242	Disease	PD	D	7
27	rs544861971	G>A	A67T	Delt	0.02	PD	0.999	Delt	-2.774	Disease	PD	D	8
28	rs749683290	G>A	E528K	Delt	0.01	PD	0.996	Delt	-3.473	Disease	PD	D	8
29	rs757888034	G>A,T	D434N	Delt	0.01	PD	0.99	Delt	- 4.467	Disease	PD	D	4
30	rs761131376	G>A	P1026R	Delt	0	PD	0.85	Delt	-5.638	Disease	PD	D	4
31	rs761919591	C>T	A422V	Delt	0	PD	0.997	Delt	-3.575	Disease	PD	D	4
32	rs765695793	C>G,T	R450C	Delt	0	PD	0.996	Delt	-7.012	Disease	PD	D	5
33	rs766988509	G>A	G503R	Delt	0.01	PD	0.986	Delt	-4.185	Disease	PD	D	3
34	rs768615287	G>A,T	C1253Y	Delt	0	PD	1	Delt	-9.308	Neutral	PD	I	1
35	rs779935967	G>A	R1002Q	Delt	0	PD	0.781	Delt	- 2.75	Disease	PD	D	8
36	rs886040977	A>G	D381G	Delt	0.01	PD	0.971	Delt	-6.253	Disease	PD	D	2
37	rs1057519574	G>T	D69Y	Delt	0	PD	1	Delt	-7.066	Disease	PD	I	1
38	rs1085307078	A>G	N198S	Delt	0	PD	0.998	Delt	-4.265	Disease	PD	D	6
39	rs1308485193	T>C,G	L1081R	Delt	0	PD	0.493	Delt	-4.322	Disease	PB	D	9
40	rs1326459816	G>A	C1286Y	Delt	0	PD	1	Delt	-9.223	Neutral	PD	I	2
41	rs1555075567	A>T	N198Y	Delt	0	PD	1	Delt	-6.824	Disease	PD	I	0
42	rs1127291	C>G,T	A1525V	Tol	0.29	B	0.079	Neutral	-0.581	Neutral	PB	D	1
43	rs3736228	C>T	A1330V	Tol	0.22	B	0.005	Neutral	-1.435	Neutral	PD	D	3
44	rs4988321	G>A,C	V667M	Tol	0.11	PD	0.691	Neutral	-1.315	Neutral	PB	D	8
45	rs41494349	A>G	Q89R	Tol	0.62	B	0	Neutral	0.36	Neutral	PD	D	0
46	rs149166384	G>A	A1301T	Tol	0.64	B	0.009	Neutral	0.033	Neutral	PB	D	5
47	rs143433231	C>T	A97V	Tol	0.25	B	0.001	Neutral	-1.108	Disease	PB	D	6
48	rs145774832	A>G	M1086V	Tol	1	B	0	Neutral	0.099	Neutral	PD	D	8
49	rs180941579	G>A	D666N	Tol	0.32	B	0.003	Neutral	-1.619	Neutral	PD	D	9
50	rs562616743	C>T	L1163F	Tol	0.61	B	0	Neutral	-0.166	Neutral	PD	D	6
51	rs201745746	C>G,T	P1026R	Tol	0.11	B	0.35	Delt	-5.638	Disease	PD	D	4
52	rs1565083843	A>G,T	K697R	Tol	0.08	B	0.031	Neutral	-2.171	Neutral	PD	D	6
53	rs141178995	C>G,T	R1188W	Tol	0.1	B	0.024	Delt	-6.05	Neutral	PD	D	4
54	rs724159827	G>A,T	D1551N	Tol	0.33	PD	0.91	Neutral	-0.959	Neutral	PD	D	5
55	rs765402802	C>G,T	R1078G	Tol	0.05	B	0.014	Delt	-3.958	Disease	PD	D	8
56	rs1273567061	A>G	T409A	Tol	0.79	B	0.215	Neutral	-1.012	Neutral	PD	D	8

Delt = Deleterious; Tol = Tolerated; PD = Probably Damaging; B = Benign; PB = Probably benign; D = Decrease; I = Increase

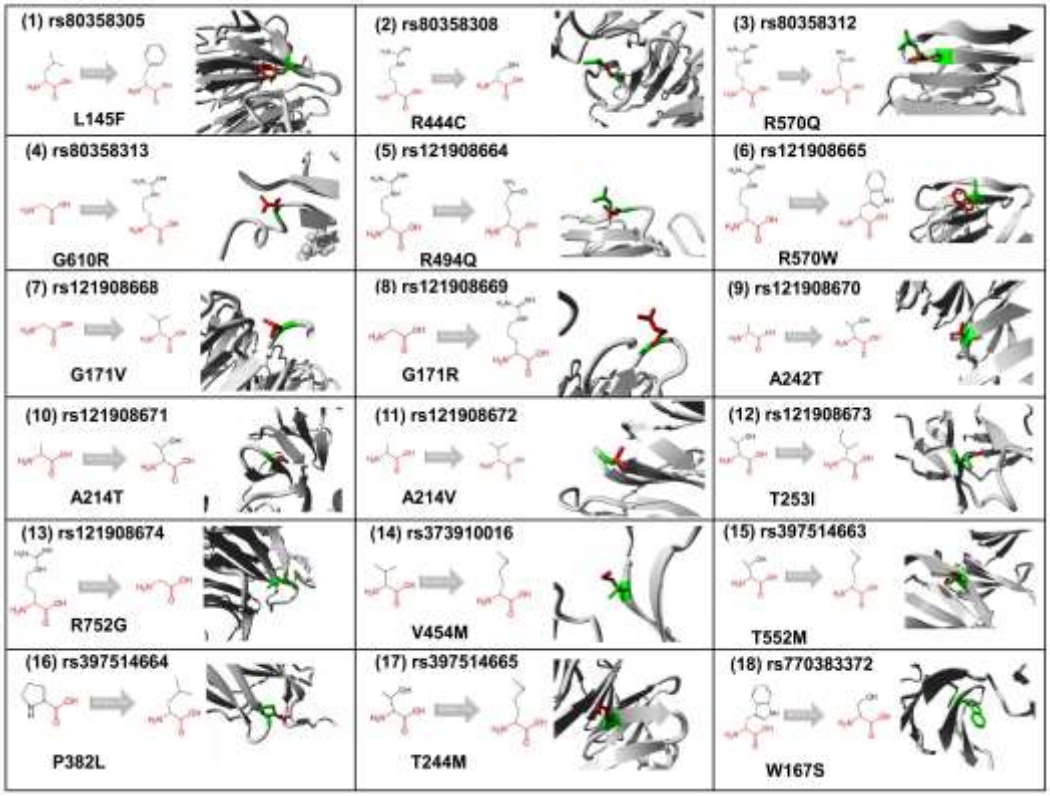
Table 2: Difference between wild- type and mutant-type amino acid properties obtained from Project HOPE software

S. No.	SNP ID	Amino acid change	Wild-type amino acids			Mutant type amino acid		
			Size	Charge	Hydrophobicity	Size	Charge	Hydrophobicity
1	rs28939709	E1367K	<	- charge		>	+ charge	
2	rs80358305	L145F	<			>		
3	rs80358308	R444C	>	+ charge	<	<	neutral	>
4	rs80358312	R570Q	>	+ charge		<	neutral	
5	rs80358313	G610R	<	Neutral	>	>	+ charge	<
6	rs121908664	R494Q	>	+ charge		<	Neutral	
7	rs121908665	R570W	<	+ charge	<	>	Neutral	>
8	rs121908668	G171V	<		<	>		>
9	rs121908669	G171R	<	Neutral	>	>	+ charge	<
10	rs121908670	A242T	<		>	>		<
11	rs121908671	A214T	<		>	>		<
12	rs121908672	A214V	<			>		
13	rs121908673	T253I	<		<	>		>
14	rs121908674	R752G	>	+ charge	<	<	neutral	>
15	rs149645175	R1534G	>	+ charge	<	<	neutral	>
16	rs373910016	V454M	<			>		
17	rs397514663	T552M	<		<	>		>
18	rs397514664	P382L	<			>		
19	rs397514665	T244M	<		<	>		>
20	rs724159826	R1529S	>	+ charge	<	<	neutral	>
21	rs770383372	W167S	>		>	<		<
22	rs61889560	R1036Q	>	+ charge		<	neutral	
23	rs146388120	D1233N		- charge			neutral	
24	rs201475647	P1504L	<			>		
25	rs200389686	L1149Q	<		>	>		<
26	rs376152274	E441K	<	- charge		>	+ charge	
27	rs544861971	A67T	<		>	>		<
28	rs749683290	E528K	<	- charge		>	+ charge	
29	rs757888034	D434N		- charge			neutral	
30	rs761131376	P1026R	<	Neutral	>	>	+ charge	<
31	rs761919591	A422V	<			>		
32	rs765695793	R450C	>	+ charge	<	<	neutral	>
33	rs766988509	G503R	<	Neutral	>	>	+ charge	<
34	rs768615287	C1253Y	<		>	>		<
35	rs779935967	R1002Q	>	+ charge		<	neutral	
36	rs886040977	D381G	>	- charge	<	<	neutral	>
37	rs1057519574	D69Y	<	- charge	<	>	neutral	>
38	rs1085307078	N198S	>		<	<		>
39	rs1308485193	L1081R	<	Neutral	>	>	+ charge	<
40	rs1326459816	C1286Y	<		>	>		<
41	rs1555075567	N198Y	<		<	>		>
42	rs1127291	A1525V	<			>		
43	rs3736228	A1330V	<			>		
44	rs4988321	V667M	<			>		
45	rs41494349	Q89R	<	Neutral		>	+ charge	
46	rs149166384	A1301T	<		>	>		<
47	rs143433231	A97V	<			>		
48	rs145774832	M1086V	>			<		
49	rs180941579	D666N		- charge			neutral	
50	rs562616743	L1163F	<			>		
51	rs201745746	P1026R	<	Neutral	>	>	+ charge	<
52	rs1565083843	K697R	<			>		
53	rs141178995	R1188W	<	+ charge	<	>	neutral	>
54	rs724159827	D1551N		- charge			neutral	
55	rs765402802	R1078G	>	+ charge	<	<	neutral	>
56	rs1273567061	T409A	>		<	<		>

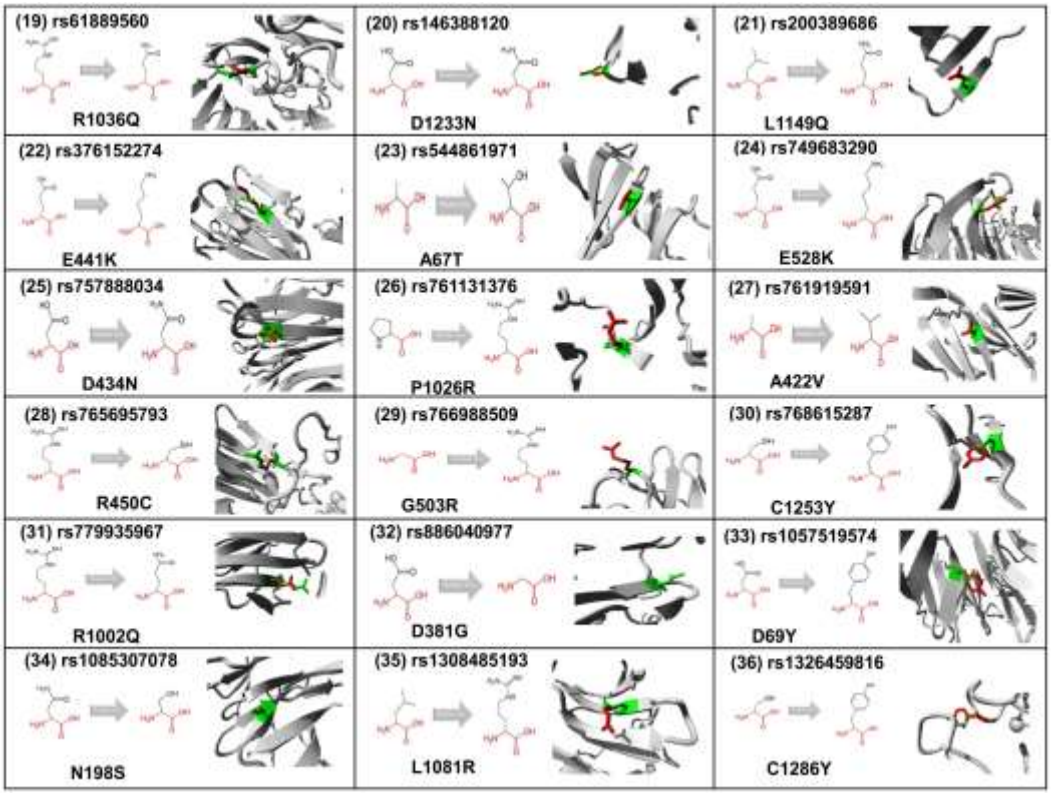
Table 3: The position of deleterious and tolerated SNPs in different domains and repeats within LRP5 protein

Domain	Deleterious (SNP ID)	Tolerated (SNP ID)
EGF-like 1	Nil	Nil
EGF-like 2	rs80358313	Nil
EGF-like 3	Nil	Nil
EGF-like 4	rs146388120, rs768615287	Nil
LDL-receptor class A 1	rs1326459816	Nil
LDL-receptor class A 2	Nil	rs3736228, rs149166384
LDL-receptor class A 3	rs28939709	Nil
Others	rs149645175, rs397514664, rs201475647, rs544861971, rs886040977, rs1057519574	rs4988321, rs180941579
Repeat	**Deleterious (SNP ID)**	**Tolerated (SNP ID)**
LDL-receptor class B 1	Nil	rs41494349, rs143433231
LDL-receptor class B 2	rs80358305	Nil
LDL-receptor class B 3	rs121908668, rs121908669, rs770383372 rs1085307078, rs1555075567	Nil
LDL-receptor class B 4	rs121908670, rs121908671, rs121908672, rs397514665	Nil
LDL-receptor class B 5	rs121908673	Nil
LDL-receptor class B 6	rs761919591	rs1273567061
LDL-receptor class B 7	rs80358308, rs373910016, rs376152274 rs757888034, rs765695793	Nil
LDL-receptor class B 8	rs121908664, rs766988509	Nil
LDL-receptor class B 9	rs397514663, rs749683290	Nil
LDL-receptor class B 10	rs80358312, rs121908665	Nil
LDL-receptor class B 11	Nil	rs1565083843
LDL-receptor class B 12	rs121908674	Nil
LDL-receptor class B 13	Nil	Nil
LDL-receptor class B 14	Nil	Nil
LDL-receptor class B 15	Nil	Nil
LDL-receptor class B 16	rs761131376, rs779935967	rs201745746
LDL-receptor class B 17	rs61889560	rs765402802
LDL-receptor class B 18	rs1308485193	rs145774832
LDL-receptor class B 19	rs200389686	rs562616743
LDL-receptor class B 20	Nil	rs141178995

in Beta-propeller 1. The differences in amino acid properties can disturb this region and disturb its function. TheSNPrs397514664 (P382L) and rs886040977 (D381G) are located in a special region Beta-propeller 2. In case of rs397514664, wild type proline is meant for providing conformational rigidity to the protein. Mutation of the proline can disrupt the conformation of the protein. In case of rs886040977, the mutated glycine can provide additional flexibility and disturb the conformation of the protein. Beta-propeller 3, a special region in the LRP5 protein contains SNP rs4988321 and rs180941579 representing mutation of a valine into a methionine at position 667 and aspartic acid into a asparagine at position 666 respectively. Though these two SNPs are included in tolerated category according to SIFT, the differences in amino acid properties can disturb the function of this region. The PPPSP motif A includes rs201475647 (P1504L). The replacement of the wild type proline affects the rigidity of the particular motif. The repeats of protein also contain mutations that may disturb the LRP5 function. The repeats LDL-receptor class B13, 14, 15 showed no SNPs for LRP5 gene at all. LDL-receptor class B1 repeat shows rs41494349, rs143433231 with substitution Q89R, A97V respectively. The SNP rs41494349 shows the mutation of glutamine into arginine at position 89. The wild-type residue forms a hydrogen bond with tyrosine at position 78. Protein with this mutation is unable to make the same hydrogen bond as the wild type residue due to the size difference between wild-type and mutant residue. The SNP rs143433231 shows mutation of alanine



A



B

Fig. 4: 3D model of structural effect of the SNPs predicted by Project HOPE (A) Deleterious SNPs, B) deleterious SNPs, C) deleterious SNP 37, tolerated SNPs 38 to 48 [C on next page]

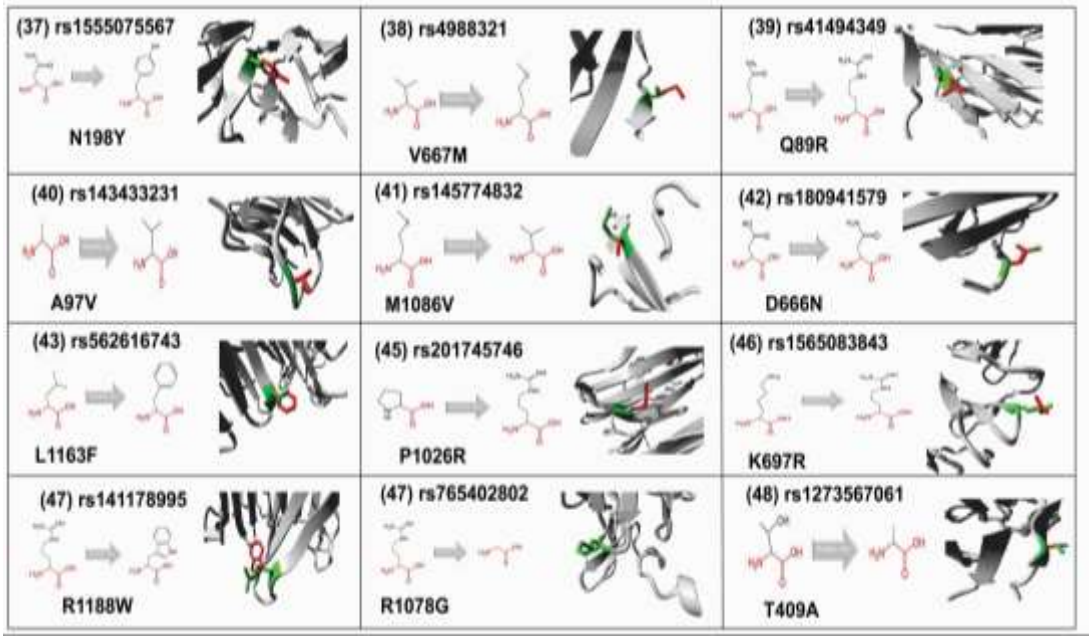


Fig. 4C: see earlier page for details

into valine at position 97. The mutant residue is bigger than the wild-type residue. The wild type residue is located on the surface of the protein, hence, its mutation can disturb interactions with other molecules or other parts of the protein. Both of these SNPs are denoted as tolerated by SIFT analysis.

LDL-receptor class B2 repeat contains SNP rs80358305 having leucine into phenylalanine substitution (L145F) at position 145. The mutation might disturb this repeat. The LDL-receptor class B3 repeat shows rs121908668, rs121908669, rs770383372, rs1085307078 and rs1555075567 with G171V, G171R, W167S, N198S and N198Y amino acid substitutions respectively (Table 2). These mutations might have a possible effect on the function of the LDL-receptor class B 3 repeats.

The LDL-receptor class B4 repeat contains SNP rs121908670, rs121908671, rs121908672, and rs397514665 with A242T, A214T, A214V and T244M amino acid substitutions, respectively (Table 4). These mutations also might have a possible effect on the function of the LDL-receptor class B4 repeats. The LDL-receptor class B 5 repeat contains SNP rs121908673 resulted in T253I mutation where the hydrophobicity of the wild type and mutant residue differs. The mutation causes loss of hydrogen bonds in the core of protein, so disturb the correct folding. All the SNPs included in LDL-receptor class B2 to B5 were categorized as deleterious by SIFT analysis. The LDL-receptor class B6 repeat possesses one deleterious (rs761919591) and one tolerated (rs1273567061) SNP with the mutation A422V and T409A, respectively. The differences in amino acid properties (Table 2) can disturb this region and thus its function. The repeat LDL-receptor class B7 contains the deleterious SNPs rs80358308, rs373910016, rs376152274, and rs757888034, rs765695793 with substitutions R444C, V454M, E441K, D434N, and R450C, respectively. These mutations might have a possible effect on the function of LDL-receptor class B7 repeats.

LDL-receptor class B8 contains SNPs rs121908664 & rs766988509 with substitution R494Q, and G503R, respectively. In rs121908664, wild-type arginine residue forms a salt bridge with glutamic acids at position 456 and/or 485. Charge difference (Table 2) due to mutation disturbs the ionic interaction made by the original wild type residue. In SNPrs766988509, the mutation of glycine into arginine at position 503 causes difference in charge between the wild-type and mutant amino acid (Table 2). The mutant residue introduces a charge at the position of a core buried wild-type residue which can lead to problems in protein folding. The mutant residue is bigger and probably will not fit to the core. The torsion angles for this residue are unusual. Only glycine is flexible enough to make these torsion angles. Mutation into another residue forces the local backbone into an incorrect

conformation and disturbs the local structure.

LDL-receptor class B9 repeat contains the SNP rs397514663, rs749683290 with substitution T552M, E528K respectively. In the SNP rs397514663 the wild-type residue forms a hydrogen bond with alanine at position 593. The size difference between wild-type and mutant residue makes that the new residue not in the correct position to make the same hydrogen bond as the original wild-type residue did. It also causes difference in hydrophobicity. In SNP rs749683290 shows the mutation of a glutamic acid into a lysine at position 528. The wild-type residue forms a salt bridge with lysine at position 526 and/or arginine at position 537. The difference in charge will disturb the ionic interaction made by the original wild-type residue. LDL-receptor class B 10 repeat possess the SNP rs80358312 and rs121908665 with substitution R570Q, R570W respectively. The wild-type arginine forms a salt bridge with glutamic acid at position 342 and/or aspartic acid at position 579. The SNP rs80358312 shows the mutation of an arginine into a glutamine at position 570. The SNP rs121908665 shows the mutation of an arginine into a tryptophan at the same position. The difference in charge will disturb the ionic interaction made by the original, wild-type residue.

The LDL-receptor class B 11 repeat contains SNP rs1565083843 (tolerated as per SIFT analysis). There is substitution of K697R i.e., the mutation of lysine into arginine at position 697. The wild-type residue forms a hydrogen bond with aspartic acid at position 718. The mutant residue is bigger than the wild-type residue (Table 2). The size difference between wild-type and mutant residue alters the capacity to make hydrogen bond like the original wild-type amino acid. The wild-type residue forms a salt bridge with aspartic acid at position 718 and/or aspartic acid at position 736. The wild type residue is located on the surface of protein and mutation of this residue can disturb interactions with other molecules or other parts of the protein.

The LDL-receptor class B 12 contains rs121908674 which is deleterious as per SIFT analysis with substitution R752G i.e. the mutation of arginine into glycine at position 752. The wild-type residue forms a salt bridge with glutamic acid at position743 and/or alanine at position 929. The difference in charge will disturb the ionic interaction made by the original, wild-type residue. LDL-receptor class B 16 repeat contains deleterious SNPs rs761131376 (P1026R) and rs779935967 (R1002Q) and tolerated SNP rs201745746 (P1026R). In rs761131376 and rs201745746, the wild type proline is substituted by arginine at position 1026. The mutant residue is bigger than wild-type residue. The wild-type residue was buried in the core of protein. The bigger mutant residue probably may not fit to the core. The hydrophobicity of wild-type and mutant residue also differs (Table 2). The mutation causes loss of hydrophobic interactions in the core of protein. Both SNPs and rs201745746 cause proline to arginine substitution at position 1026 but show difference in their respective nucleotide substitutions. In rs761131376, G to A and in rs201745746, C to G, T transitions have occurred. In rs779935967 the wild-type residue forms a hydrogen bond with tyrosine at position 992. The size between wild-type and mutant residue differs which makes the new residue not in the correct position to make the same hydrogen bond as the original wild-type residue did. The wild-type residue forms a salt bridge with aspartic acid at position 985 and/or aspartic acid at position 988. The difference in charge disturbs the ionic interaction made by the original, wild-type residue.

LDL-receptor class B 17 contains SNP rs61889560 (deleterious) and rs765402802 (tolerated) with amino acid substitution R1036Q and R1078G, respectively. In SNP rs61889560, wild-type residue forms a salt bridge with aspartic acid at position 985 and/or aspartic acid at position 988. The difference in charge with the mutated residue disturbs the ionic interaction made by the original, wild-type residue. In SNP rs765402802 there is a substitution of arginine to glycine at position 1078. The wild-type residue forms a salt bridge with glutamic acid at position1077 and/or aspartic acid at position 1120. The difference in charge may disturb the ionic interaction made by the original, wild-type residue. LDL-receptor class B 18 repeat contains SNP rs1308485193 (deleterious) and SNP rs1457748329 (tolerated) with amino acid substitution L1081R and M1086V, respectively. The SNP rs1308485193 shows mutation of leucine into arginine at position 1081. Due to charge difference between wild-type and mutant amino acid the mutant residue introduces a charge in a buried residue which leads to protein folding problems. The wild-type residue was buried in the core of protein

whereas mutant residue was bigger and probably may not fit. The hydrophobicity of wild-type and mutant residue differs. The mutation causes loss of hydrophobic interactions in protein core. In SNP rs1457748329 there is substitution of methionine into valine at position1086. The mutant residue is smaller than the wild-type residue and will cause a possible loss of external interactions.

LDL-receptor class B 19 repeat contains SNP rs200389686 (deleterious) and SNP rs562616743 (tolerated) with substitution L1149Q and L1163F, respectively. The SNP rs200389686 with substitution L1149Q, shows mutation of leucine to glutamine at position 1149. The mutant residue is bigger than wild-type residue and is located on protein surface, so mutation in leucine residue disturbs interactions with other molecules or other parts of protein. The hydrophobicity of wild-type and mutant residue differs. The mutation may cause the loss of hydrophobic interactions with other molecules on protein surface. In case of SNP rs562616743 there is a substitution of leucine into a phenylalanine at position 1163. The wild-type and mutant amino acids differ in size. The mutant residue is bigger than wild-type residue. The wild-type residue is buried in the core of protein. The bigger mutant residue will probably not fit to the core of the protein.

LDL-receptor class B 20 repeat contains SNP rs141178995 (tolerated as per SIFT analysis). There is a substitution of arginine into tryptophan at position 1188 (R1188W). The wild-type residue forms a hydrogen bond with glutamic acid at position 1178. Due to size difference between wild-type and mutant residue the new residue is not in correct position to make the same hydrogen bond which the original wild-type residue did. The difference in hydrophobicity affects hydrogen bond formation. The wild-type residue forms a salt bridge with glutamic acid at position1150 and/or glutamic acid at position 1178. The difference in charge disturbs the ionic interaction of the original wild-type residue.

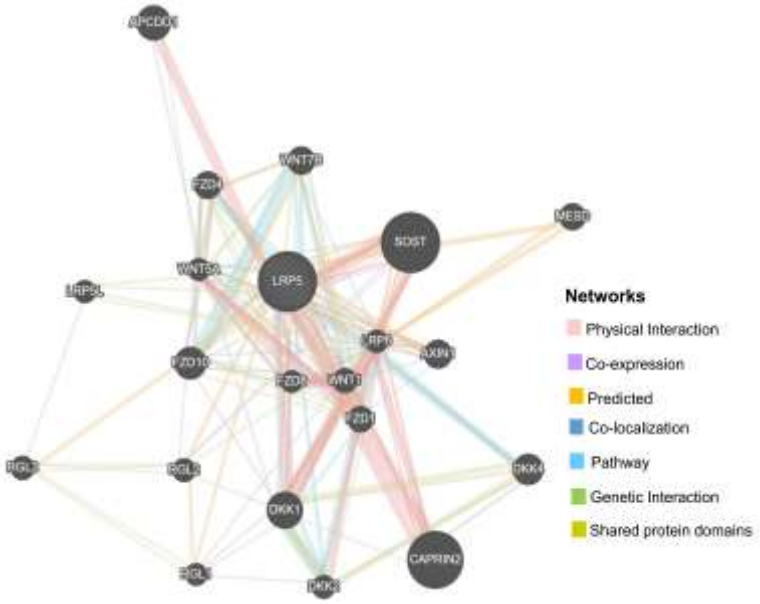


Fig. 5: Genetic network of LRP5 predicted by GeneMANIA

Gene-gene interactions

LRP5 is situated at the centre of a complex genetic network to regulate bone metabolism. The crosstalk of LRP5 gene with another 20 genes in the form of a network is shown in Fig. 5. According to Gene MANIA software tool SOST, CAPRIN2, DKK1 are ranked in first, second and third position respectively in terms of the interaction with LRP5. SOST (sclerostin) is a negative regulator of bone growth (Qin *et al*., 2013). CAPRIN2 (caprin family member 2) promotes canonical WNT signalling by regulating LRP5 phosphorylation (Ding *et al.*, 2008), DKK1 (dickkopf WNT signalling pathway inhibitor 1) inhibits WNT signalling pathway by binding to LRP5/6 (Ahn *et al.*, 2011).

Conclusions: This work is a comprehensive report of almost all the available online information about the effects of missense SNPs of LRP5. Still there is a scope to generate more experimental data about each and every amino acid substitution to fill up the gaps in information like in some cases Project HOPE could not generate 3D structure due to lack of information. Some tolerated SNPs (like rs3736228) were found significantly associated with bone related disease such as osteoporosis in some populations of world (Liu *et al*., 2017; Nakano *et al*., 2021). Hence categorization of the SNPs as deleterious and tolerated by software must be reconsidered by alternative methods. Otherwise, this article will be very helpful in planning experimental research on LRP5 SNPs in future.

REFERENCES

Abdelraheem, N.E., Osman, M.M., Elgemaabi, O.M., Alla, A.A.F., Ismail, M.M., Osman, S.A. and Hassan, M.A. 2016. Computational Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) in Human MutS Homolog6 (MSH6) Gene. *American Journal of Bioinformatics Research*, **6**(2): 56-97.

Ahn, V.E., Chu, M.L., Choi, H.J., Tran, D., Abo, A. and Weis, W.I. 2011. Structural basis of Wnt signaling inhibition by Dickkopf binding to LRP5/6. *Developmental Cell*, **21**(5): 862-873.

Ding, Y., Xi, Y., Chen, T., Wang, J.Y., Tao, D.L., Wu, Z.L., Li, Y.P., Li, C., Zeng, R. and Li, L. 2008. Caprin-2 enhances canonical Wnt signaling through regulating LRP5/6 phosphorylation. *The Journal of Cell Biology*, **182**(5): 865-872.

Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G.D. and Morris, Q. 2018. Gene MANIA update 2018. *Nucleic Acids Research*, **46**(W1): W60-W64. [*https://doi.org/10.1093/nar/gky311*].

Gong, Y., Vikkula, M., Boon, L., Liu, J., Beighton, P., Ramesar, R., Peltonen, L., Somer, H., Hirose, T., Dallapiccola, B., De Paepe, A., Swoboda, W., Zabel, B., Superti-Furga, A., Steinmann, B., Brunner, H.G., Jans, A., Boles, R.G., Adkins, W., van den Boogaard, M.J. and Warman, M.L. 1996. Osteoporosis-pseudoglioma syndrome, a disorder affecting skeletal strength and vision, is assigned to chromosome region 11q12-13. *American Journal of Human Genetics*, **59**(1): 146-151.

Guo, Y.F., Xiong, D.H., Shen, H., Zhao, L.J., Xiao, P., Guo, Y., Wang, W., Yang, T.L., Recker, R.R. and Deng, H.W. 2006. Polymorphisms of the low-density lipoprotein receptor-related protein 5 (LRP5) gene are associated with obesity phenotypes in a large family-based association study. *Journal of Medical Genetics*, **43**(10): 798-803.

Hey, P.J., Twells, R.C., Phillips, M.S., Yusuke Nakagawa, Brown, S.D., Kawaguchi, Y., Cox, R., Guochun Xie, Dugan, V., Hammond, H., Metzker, M.L., Todd, J.A. and Hess, J.F. 1998. Cloning of a novel member of the low-density lipoprotein receptor family. *Gene*, **216**(1): 103-111.

Keats, B. and Sherman, S.L. 2013. Population genetips. pp. 1-12. ***In:*** *Emery and Rimoin's Principles and Practice of Medical Genetics*. Elsevier, [*doi:10.1016/b978-0-12-383834-6.00015-x*].

Laine, C.M., Chung, B.D., Susic, M., Prescott, T., Semler, O., Fiskerstrand, T., d'Eufemia, P., Castori, M., Pekkinen, M., Sochett, E. and Cole, W.G. 2011. Novel mutations affecting LRP5 splicing in patients with osteoporosis-pseudoglioma syndrome (OPPG). *European Journal of Human Genetics*, **19**(8): 875-881.

Lara-Castillo, N. and Johnson, M.L. 2015. LRP receptor family member associated bone disease. *Reviews in Endocrine* and *Metabolic Disorders*, **16**(2): 141-148.

Lim, S.W., Tan, K.J., Azuraidi, O. M., Sathiya, M., Lim, E. C., Lai, K.S., Yap, W.S. and Afizan, N.A.R.N.M. 2021. Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer. *Scientific Reports*, **11**(1), 24206. [*https://doi.org/10.1038/s41598-021-03624-x*].

Liu, K., Tan, L.J., Wang, P., Chen, X.D., Zhu, L.H., Zeng, Q., Hu, Y. and Deng, H.W. 2017. Functional relevance for associations between osteoporosis and genetic variants. *PLoS One*, **12**(4): e0174808. [*https://doi.org/10.1371/journal.pone.0174808*].

Manfredi, M., Savojardo, C., Martelli, P.L. and Casadio, R. 2022. E-SNPs & GO: Embedding of protein sequence and function improves the annotation of human pathogenic variants. *Bioinformatics (Oxford, England)*, **38**(23): 5168-5174.

Mao, W., Wordinger, R.J. and Clark, A.F. 2011. Functional analysis of disease-associated polymorphism LRP5.Q89R. *Molecular Vision*, **17**: 894-902.

Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhurji, S.K., Boddu, S., Branco Lins, P.R., Brooks, L., Ramaraju, S.B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., Davidson, C. and Flicek, P. 2023. Ensembl 2023. *Nucleic Acids Research*, **51**(D1): D933-D941. [*https://doi.org/10.1093/nar/gkac958*].

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.P., Mushayamaha, T. and Thomas, P.D. 2021. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, **49**(D1): D394-D403. [*https://doi.org/10.1093/nar/gkaa1106*].

Nakano, M., Yui, H., Kikugawa, S., Tokida, R., Sakai, N., Kondo, N., Endo, N., Haro, H., Shimodaira, H., Suzuki, T., Kato, H., Takahashi, J. and Nakamura, Y. 2021. Associations of *LRP5* and *MTHFR* gene variants with osteoarthritis Prevalence in Elderly Women: A Japanese Cohort Survey Randomly Sampled from a Basic Resident Registry. *Therapeutics and Clinical Risk Management*, **17**: 1065-1073.

NCI's Dictionary of Genetics Terms [online] 2022. [*https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/single-nucleotide-polymorphism*]. (Accessed 19 July 2022)

Nimir, M., Abdelrahim, M., Abdelrahim, M., Abdalla, M., Ahmed, W.E., Abdullah, M. and Hamid, M. M.A. 2017. *In silico* analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene. *F1000Research*, **6**: 243. [*https://doi.org/10.12688/f1000research.10937.2*].

Norwitz, N.G., Mota, A.S., Misra, M. and Ackerman, K.E. 2019. LRP5, Bone Density, and Mechanical Stress: A Case Report and Literature Review. *Frontiers in Endocrinology*, **10**: 184. [*https://doi.org/10.3389/fendo.2019.00184*].

Pekkinen, M., Grigelioniene, G., Akin, L., Shah, K., Karaer, K., Kurtoğlu, S., Ekbote, A., Aycan, Z., Sağsak, E., Danda, S., Åström, E. and Mäkitie, O. 2017. Novel mutations in the LRP5 gene in patients with Osteoporosis-pseudoglioma syndrome. *American Journal of Medical Genetics. Part A*, **173**(12): 3132-3135.

Qin, L.J., Ding, D.X., Cui, L.L. and Huang, Q.Y. 2013. Expression and regulation of the SOST gene. *Hereditas*, **35**(8): 939-947.

Sandell, L. and Sharp, N.P. 2022. Fitness effects of mutations: An assessment of PROVEAN predictions using mutation accumulation data. *Genome Biology and Evolution*, **14**(1): evac004. [*https://doi.org/10.1093/gbe/evac004*].

Shastry B.S. 2009. SNPs: Impact on gene function and phenotype. **In**: *Methods in Molecular Biology* (ed. N.J. Clifton), **578**: 3-22.[*https://doi.org/10.1007/978-1-60327-411-1_1*].

Teng, S., Srivastava, A.K., Schwartz, C.E., Alexov, E. and Wang, L. 2010. Structural assessment of the effects of amino acid substitutions on protein stability and protein-protein interaction. *International Journal of Computational Biology and Drug Design*, **3**(4): 334-349.

Ueki, M., Takeshita, H., Utsunomiya, N., Chino, T., Oyama, N., Hasegawa, M., Kimura-Kataoka, K., Fujihara, J., Iida, R. and Yasuda, T. 2017. Survey of single-nucleotide polymorphisms in the gene encoding human deoxyribonuclease I-like 2 producing loss of function potentially implicated in the pathogenesis of parakeratosis. *PloS One*, **12**(4): e0175083. [*https://doi.org/10.1371/journal.pone.0175083*].

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. and Ng, P.C. 2016. SIFT missense predictions for genomes. *Nature Protocols*, **11**(1): 1-9.

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**: 548. [*https://doi.org/10.1186/1471-2105-11-548*].

Wang, Q.F., Bi, H.S., Qin, Z.L., Wang, P., Nie, F.F. and Zhang, G.W. 2020. Associations of LRP5 gene with bone mineral density, bone turnover markers, and fractures in the elderly with osteoporosis. *Frontiers in Endocrinology*, **11**: 571549. [*https://doi.org/10.3389/fendo.2020.571549*].

Williams, B.O., and Insogna, K.L. 2009. Where Wnts went: The exploding field of Lrp5 and Lrp6 signaling in bone. *Journal of Bone and Mineral Research*, **24**(2): 171-178.

Xu, G.Y., Qiu, Y. and Mao, H.J. 2014. Common polymorphism in the LRP5 gene may increase the risk of bone fracture and osteoporosis. *BioMed Research International*, **2014**: 290531. [*https://doi.org/10.1155/2014/290531*].