International Journal of Genetics and Genomics
2022; 10(4): 94-102
http://www.sciencepublishinggroup.com/j/ijgg
doi: 10.11648/j.ijgg.20221004.12
ISSN: 2376-7340 (Print); ISSN: 2376-7359 (Online)

The Effect of Non-synonymous Single Nucleotide Polymorphisms Variants in *ALPL* Gene, a Computational Approach

Nabaa Kamal Alshafei[1], Intisar Hassan Saeed[2], Mona Abdelrahman Mohamed Khaier[3, *]

[1]Department of Biochemistry, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

[2]Department of Physiology, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

[3]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

Email address:

munakhaier@gmail.com (Mona Abdelrahman Mohamed Khaier), muna.khaier@bahri.edu.sd (Mona Abdelrahman Mohamed Khaier)

[*]Corresponding author

To cite this article:

Nabaa Kamal Alshafei, Intisar Hassan Saeed, Mona Abdelrahman Mohamed Khaier. The Effect of Non-synonymous Single Nucleotide Polymorphisms Variants in *ALPL* Gene, a Computational Approach. *International Journal of Genetics and Genomics.* Vol. 10, No. 4, 2022, pp. 94-102. doi: 10.11648/j.ijgg.20221004.12

Received: July 6, 2022; **Accepted**: September 26, 2022; **Published**: December 29, 2022

Abstract: *Background*: Phosphorus is one of the major macronutrient essential for normal growth and development of living organisms. Alkaline phosphatase (ALP) is an enzyme present in human and animals responsible for solubilization and mineralization of organic phosphate and makes it readily available for the body. The disease Hypophosphatasia (HPP) is an autosomal recessive inherited one, branded by malfunctioning mineralization of bone, dental problems, and low serum ALP levels. This study aimed to scrutinize the effect of non-synonymous SNPs (nsSNPs) of *ALPL* gene on protein function and structure using different computational software. *Material and Methods:* Different Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) and protein related sequences were attained from NCBI and ExPASY databases. Deleterious and damaging effect of nsSNPs were analyzed using SIFT, Polyphen-2, Provean and SNPs&GO software. Protein stability was inspected using I-Mutant and MUpro software. The interaction of ALPL with other genes was studied using GeneMANIA software. The structural and functional influence of point mutations was predicted using Project Hope software. *Results: ALPL* gene was found to have an association with 20 other genes such as *TRAF3* and *TPP1* using GeneMANIA. It comprises a total of 485 SNPs out of that 188 were found to be synonymous, 298 were nsSNPs. Analysis of the nsSNPs by SIFT predicts 33 as deleterious and 265 as tolerated ones. Using Provean software, 26 were deleterious while 7 nsSNPs were neutral. Taking the deleterious nsSNPSs to Polyphen-2, 24 nsSNPs were damaging, while 2 were benign. Using SNPs&GO 13 nsSNPs were predicted as disease-related while 11 were predicted to be neutral. By using PHD SNPs 11 nsSNPs were predicted as disease-related while 13 were predicted to be neutral. Project Hope analyzes the mutations according to their size, charge, hydrophobicity, and conservancy. *Conclusion:* This study reveals 11 nsSNPs as being possibly pathogenic variants. Seven of them were already reported from previous studies by DNA sequencing, while the remaining four were predicted in this study for the first time.

Keywords: *ALPL* Gene, Computational Analysis, GeneMANIA, Hypophosphatasia, Non Synonymous SNP, SIFT, Polyphen-2

1. Introduction

Mutations in the *ALPL* gene cause hypophosphatasia. This gene affords commands for making an enzyme entitled tissue-nonspecific alkaline phosphatase (TNSALP), which plays a critical role in mineralization of the skeleton and teeth. Mutations in the *ALPL* gene lead to the manufacturing of an abnormal type of TNSALP that cannot participate effectively in the mineralization process [1]. A shortage of TNSALP allows several other substances, which are normally processed by the enzyme, to build up abnormally in

the body. The accumulation of one of these compounds, namely inorganic pyrophosphate (PPi), underlies the defective mineralization of bones and teeth in people with hypophosphatasia [2]. Most Hypophosphatasia (HPP) is inherited as recessive variants with only a subset of moderate forms being inherited dominantly. Most patients are homozygotes or compound heterozygotes of ALPL variants [2]. The *ALPL* gene is localized on chromosome 1p36.12 and codes tissue nonspecific alkaline phosphatase (TNSALP), and categorized by defective mineralization of bone, dental problems, and low serum ALP levels. TNSALP forms a homodimer which is required to exert an enzymatic activity. [3, 4]. Alkaline phosphatases are specific enzymes present in nearly all living organisms catalyzing the dephosphorylation of pyrophosphate (PPi) and pyridoxal-50-phosphate (PLP) in vivo, presumably as well as nucleotides like ATP and proteins like osteopontin [5-7]. Low TNSALP activity results in extracellular accumulation of the TNSALP substrates inorganic pyrophosphate (PPi) and pyridoxal 5′-phosphate (PLP). Impaired bone mineralization due to PPi accumulation can lead to rickets, osteomalacia, and fractures [8, 9]. Hypophosphatasia is generally classified into the following subtypes based on the age of onset, clinical features and disease severity; perinatal severe, perinatal benign, infantile, childhood, adult and odonto HPP. To date, more than 400 *ALPL* gene mutations have been identified and listed in the *ALPL* gene mutation database (http://alplmutationdatabase.hypophosphatasie.*Com*). [3, 4]. TNSALP is expressed in fundamentally all tissues (especially in hepatic, skeletal, and renal tissues) and shows high activity in mineralizing bone where it is localized in the plasma membrane of osteoblastic cells [10]. *ALPL* gene mutations have been reported to be genetic risk factors of low bone mineral density and typical femoral fractures [11, 12].

The human *ALPL* gene encodes, according to the ENSEMBL database, seven different transcripts. Only three human transcript variants (ENST00000374840.8, ENST00000539907.5, ENST00000540617.5) have been validated and display tissue-specific alternative splicing, e.g., in the liver and bone [13, 14]. The reference transcript (ENST00000374840.8; NM_000478.6) spans a genomic region of more than 50 kb, contains overall 12 exons, including 11 protein-coding exons, has a transcript length of 2536 bp, and corresponds to the bony-type ALPL transcript described by [14]. Humans have four different genes encoding distinct alkaline phosphatase isoforms [10]. Three of these are considered as being expressed in a tissue-specific manner in the placenta, intestine, and germ cells (ALPP (NCBI GeneID: 250), ALPI (NCBI GeneID: 248), and ALPG (NCBI GeneID: 251) and one is characterized as tissue-nonspecific (ALPL (NCBI GeneID: 249). More than 400 *ALPL* pathogenic variants have been identified [15] and considerable clinical variability has been observed, with manifestations that can present at any time from in utero and infancy through adult- hood [16]. The effects of nsSNPs on ALPL protein structure and functions still remain indefinable; therefore, in this present study, we analyze the deleterious effect of nsSNPs on the *ALPL* gene by using various computational databases and bioinformatics tools, as it is fast and cost effective screening for pathologic nsSNPs.

2. Material and Methods

2.1. Collection of Data

Data was rescued from the SNP database of the National Center for Biotechnology Information (dbSNP) (http://www.ncbi.nlm.nih.gov/snp), March 2022. The primary sequence of the protein (Uniprot accession number: P05186) encoded by the *ALPL* gene was obtained from the UniProt database (March 2022).

2.2. Gene Association for ALPL Gene

Gene MANIA Software is a flexible user-friendly website for making hypotheses interface that finds other genes related to a set of input genes, using a very large set of functional association data. Association data include genomics and proteomics data, pathways, co-expression, co-localization and protein domain similarity in this mode; it weights each functional genomic dataset according to its predictive value for the query. http://www.genemania.org [17].

2.3. Functional and Structural Analysis of the nsSNPs

Only nsSNPs were selected from the NCBI SNPs database as they can modify the sequence of the amino acid encoded by the protein and has the potential to disturb the structural arrangement and function of the proteins. The functional effect of the SNPs on the protein was investigated using SIFT, Provean, Polyphen-2, SNPs& GO, and PHD-SNP. The stability of the protein as the result of the mutation was studied using I- Mutant and MUpro, and finally the effect of the SNPs on the structure was predicted using Project Hope software.

2.3.1. SIFT (Sorting Intolerant from Tolerant)

This software it is an online tool that was developed by [18]. It predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids by using sequence homology. It performs analysis based on different algorithms and it interprets the homologous sequences using the Swiss-Prot (version 51.3) and TrEMBL (version 34.3) [18]. The SIFT prediction is given as a tolerance index (TI) score ranging from 0.0 to 1.0, which is the normalized probability that the amino acid change is tolerated. The threshold intolerance score for SNPs is 0.05 or less [19].

2.3.2. Provean (Protein Variation Effect Analysis)

Is a software tool that predicts whether an amino acid substitution has an effect on the biological function of the protein or not. Provean is useful for filtering sequence variants to identify nonsynonymous variants that are predicted to be functionally important. The performance of PROVEAN is comparable to popular tools such as SIFT or

PolyPhen-2 [20]. A fast computation approach to obtain pairwise sequence alignment scores enabled the generation of precomputed PROVEAN predictions for 20 single AA substitutions at every amino acid position of all protein sequences in humans and mice [21].

2.3.3. Polyphen-2 (Polymorphism Phenotyping v2)

It is a multiple sequence alignment server that aligns sequences using structural information. Input for the PolyPhen-2 server is either a protein sequence or a SWALL database ID or accession number together with sequence position with two amino acid variants [22]. It estimates the position-specific independent count score (PSIC) for every variant and then determines the difference between them, the higher the PSI, the higher the functional impact of the amino acid on the protein function may be. Prediction outcomes could be classified as probably damaging, possibly damaging or benign according to the score ranging from (0–1) [23].

2.4. Association of Predicted SNPs to Disease

2.4.1. SNPs &GO (Single Nucleotide Polymorphism & Gene Ontology)

SNPs&GO, an accurate method that, starting from a protein sequence, can predict whether a mutation is disease-related or not by exploiting the protein functional annotation. SNPs&GO collects in unique framework information derived from protein sequence, evolutionary information, and function as encoded in the Gene Ontology terms, and outperforms other available predictive methods [24].

2.4.2. PHD-SNP, (Predictor of Human Deleterious SNP)

PHD-SNP, also known as predictor of human deleterious single-nucleotide polymorphisms (SNPs) https://snps.biofold.org/phd-snp/phd-snp.html is an SVM classifier that is optimized to assess whether a nonsynonymous single-point mutation can be grouped as disease-associated or neutral polymorphism from protein sequence information [25]. The techniques are calculated by using BLAST software [26] against the UniRef 90 database [27] as entry data about the mutation, such as its background setting and status at the genetically altered site. PHD-SNP yields an output grade (range 0–1) for each mutation, which reflects the likelihood means; this nsSNP will be connected with disease. The technique claims that 0.5 is the threshold below which disease-associated nsSNPs are expected to be [28].

2.5. Prediction of Protein Stability

Two software were used to predict the effect of a missense mutation on the protein's stability.

2.5.1. I-Mutant 3.0 http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi

This software offers the opportunity to predict automatically protein stability changes upon single-site mutations starting from protein sequence alone or protein structure when available. Moreover, it can predict deleterious Single Nucleotide Polymorphism starting from the protein sequence alone. [25].

2.5.2. MUpro: http://mupro.proteomics.ics.uci.edu/

It is a machine-learning approach based on support-vector machines to predict the protein stability changes for single site mutations in two contexts taking into account structure-dependent and sequence-dependent information, respectively [29].

2.6. Prediction of Protein Modeling

This was achieved by using project Hope software https://www3.cmbi.umcn.nl/hope/. HOPE is a next-generation software application for automatic mutant analysis. HOPE was designed to explain the molecular origin of a disease-related phenotype caused by mutations in human proteins. HOPE collects information from data sources such as the protein's 3D structure and the UniProt database of well-annotated protein sequences. For each protein, this data is stored in a Postgre SQL-based information system. A decision scheme is used to process these data and predict the effects of the mutation on the 3D structure and the protein's function. [30].

3. Results

In this study ALPL gene was found to have an association with 20 other different genes. Among the most important ones is the TRAF3 and TPP1 gene which is also a trans membrane protein (Figure 1. and Table 1). The physical interaction and co expression of this gene with other related gene (Figure 1).

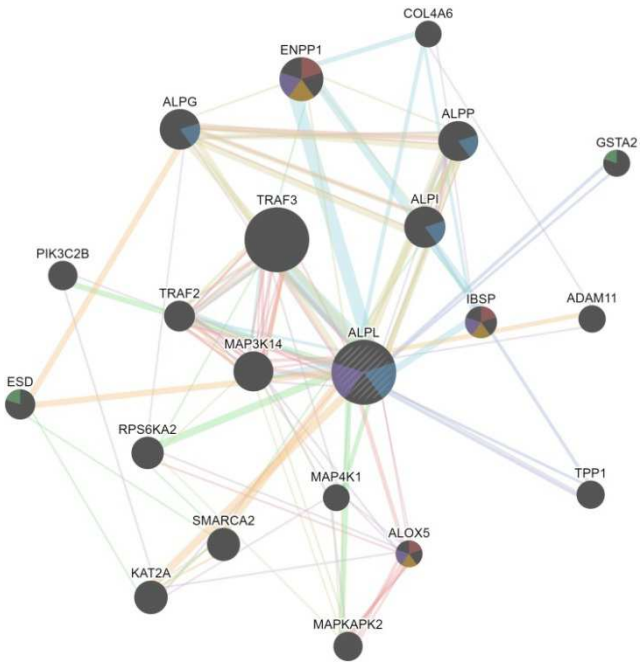


Figure 1. *GeneMANIA result for ALPL Gene.*

Table 1. *Gene Description Rank Using Gene MANIA.*

Gene	Description
ALPL	alkaline phosphatase, biomineralization associated
TRAF3	TNF receptor associated factor 3
ENPP1	Ectonucleotide pyrophosphatase/phosphodiesterase 1
ALPI	alkaline phosphatase, intestinal
ALPG	ALPG alkaline phosphatase, germ cell
ALPP	ALPP alkaline phosphatase, placental
MAP3K14	MAP3K14 mitogen-activated protein kinase kinase kinase 14
KAT2A	Lysine acetyl transferase 2A
RPS6KA2	ribosomal protein S6 kinase A2
IBSP	integrin binding sialoprotein
TRAF2	TNF receptor associated factor 2
PIK3C2B	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 beta
MAPKAPK2	MAPK activated protein kinase 2
TPP1	tripeptidyl peptidase 1
ADAM11	ADAM metallopeptidase domain 11
GSTA2	glutathione S-transferase alpha 2
ALOX5	arachidonate 5-lipoxygenase
MAP4K1	mitogen-activated protein kinase kinase kinase kinase 1
COL4A6	collagen type IV alpha 6 chain

The SNPs of the *ALPL* gene systematically examined in this study were recovered from the NCBI SNP database. The protein was recovered from UniProtKB. The human *ALPL* gene comprises a total of 485 SNPs out of that 188 were found to be synonymous, 298 were nonsynonymous SNPs (nsSNPs). Analysis of the nsSNPs by SIFT predicts 33 as deleterious and 402 as tolerated ones. Using Provean only 26 were deleterious while 7 SNPs were neutral. Taking the deleterious nsSNPSs to Polyphen-2, 24 nsSNPs were damaging (21 were probably damaging and 3 were possibly damaging), while two were benign. Using SNPs&GO 13 nsSNPs were predicted as disease-related while 11 were predicted to be neutral. By using PHD SNP 11 nsSNPs were predicted as disease-related while 13 were predicted to be neutral, Tables 2 and 3.

Table 2. *The Results of Different Software.*

Software	Results
Retrieved SNPs	188 synonymous and 298 non- synonymous
SIFT	33 Deleterious and 265Tolerated
Provean	26 deleterious 7 neutral
Polyphen-2	21 probably damaging 3 possibly damaging 2 benign
SNPs & GO	13 SNPs had a disease association and 11 neutral
PHD SNPs	11 SNPs had a disease association and 13 neutral

Table 3. *List of nsSNPs predicted to be deleterious by SIFT, Provean and PolyPhen-2 coding region of ALPL gene.*

SNP ID	Amino acids change	SIFT prediction	SIFT score	Provean Prediction	Provean score	Polyphen-2 prediction	Polyphen-2 Score
rs143358506	L37P	deleterious	0.001	Deleterious	-5.945	probably damaging	1
rs139811782	V95M	deleterious	0.001	Neutral	-0.989	probably damaging	0.997
rs139811782	A40M	deleterious	0.001	Deleterious	-4.154	probably damaging	1
rs11586344	N47S	deleterious	0.037	Neutral	-1.977	benign	0.063
rs11586344	N124S	deleterious	0.037	Deleterious	-4.762	benign	0.063
rs11586344	A69S	deleterious	0.037	Deleterious	-2.682	probably damaging	0.998
rs146517700	T127N	deleterious	0.013	Deleterious	-4.47	probably damaging	0.997
rs146517700	V50N	deleterious	0.015	Neutral	-1.368	possibly damaging	0.931
rs146517700	I72N	deleterious	0.015	Deleterious	-6.576	probably damaging	1
rs200621180	G75C	deleterious	0.015	Deleterious	-8.013	probably damaging	1
rs200621180	L97C	deleterious	0.001	Deleterious	-4.734	probably damaging	1
rs200621180	R152C	deleterious	0.001	Deleterious	-3.628	benign	0.94
rs374558572	S149F	deleterious	0.001	Deleterious	-5.793	probably damaging	1
rs374558572	T127F	deleterious	0.046	Deleterious	-5.82	probably damaging	0.999
rs374558572	I204F	deleterious	0.046	Deleterious	-3.634	benign	0.146
rs199665722	E146Q	deleterious	0.046	Deleterious	-2.751	probably damaging	0.976
rs199665722	R168Q	deleterious	0.027	Deleterious	-3.266	probably damaging	1
rs199665722	R223Q	deleterious	0.027	Deleterious	-3.554	probably damaging	0.997

SNP ID	Amino acids change	SIFT prediction	SIFT score	Provean Prediction	Provean score	Polyphen-2 prediction	Polyphen-2 Score
rs200133602	R321W	deleterious	0.027	Deleterious	-3.315	probably damaging	0.997
rs200133602	G244W	deleterious	0.035	Deleterious	-7.372	probably damaging	1
rs200133602	S266W	deleterious	0.038	Deleterious	-3.261	probably damaging	0.997
rs371243939	L37C	deleterious	0.038	Deleterious	-4.02	possibly damaging	0.901
rs371243939	V314C	deleterious	0.038	Neutral	-1.748	possibly damaging	0.934
rs371243939	I336C	deleterious	0.007	Deleterious	-4.523	probably damaging	1
rs371243939	R391C	deleterious	0.007	Deleterious	-7.63	probably damaging	1
rs138690664	A96C	deleterious	0.007	Deleterious	-3.71	probably damaging	1
rs376354718	G120I	deleterious	0.001	Deleterious	0	probably damaging	1
rs376354718	G397I	deleterious	0.001	Deleterious	-9.031	probably damaging	1
rs376354718	P419I	deleterious	0.001	Deleterious	-9.368	probably damaging	1
rs376354718	V474I	deleterious	0.041	Neutral	-0.839	possibly damaging	0.921
rs34605986	M467A	deleterious	0.041	Deleterious	-3.818	possibly damaging	0.481
rs34605986	S445A	deleterious	0.041	Neutral	0.671	benign	0.001
rs34605986	V522A	deleterious	0.001	Neutral	0.052	benign	0.98

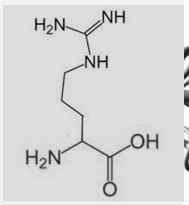
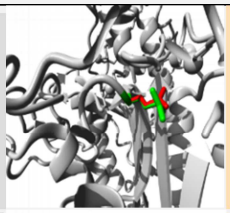
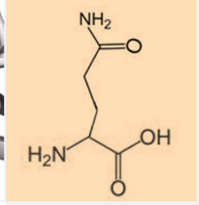
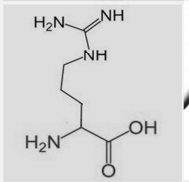
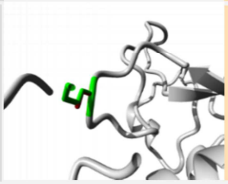
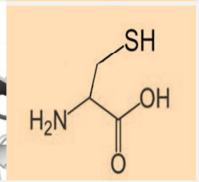
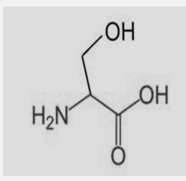
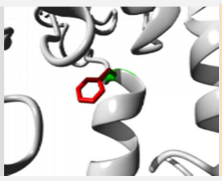
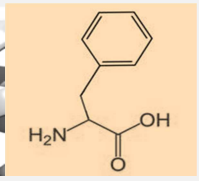
Using additional software SNPs&GO showed that 13 SNPs had a disease effect and 11 were neutral. For protein stability, I-Mutant software was used; all disease-related mutations resulting from SNPs&Go were predicted to decrease the protein stability with varied probabilities, Table 4.

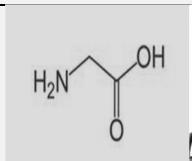
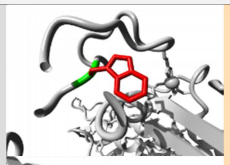
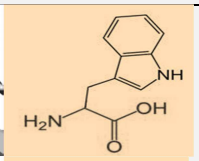
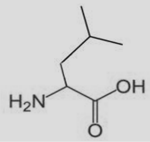
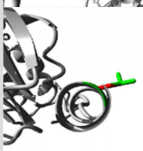
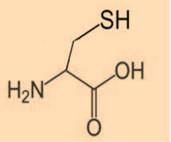
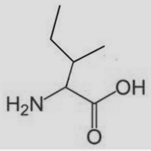
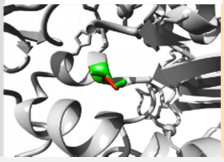
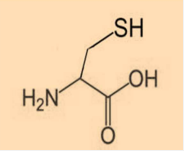
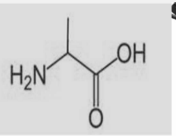
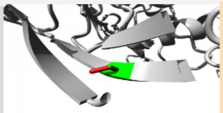
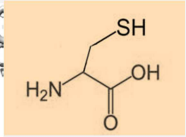
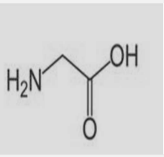
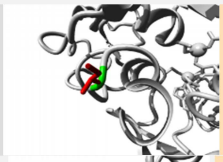
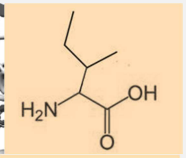
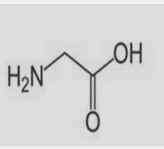
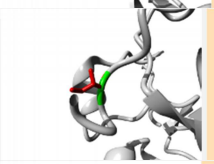
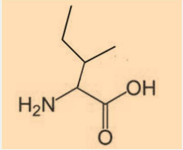
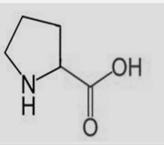
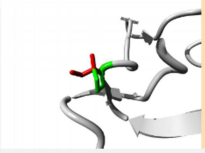
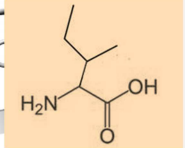
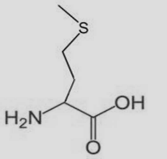
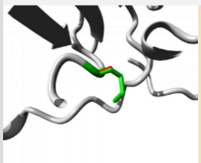
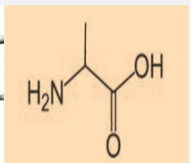
Table 4. *Results of SNPs&GO, PHD SNP and I-Mutant software.*

Amino acids change	SNP&GO Prediction	SNP&GO RI	PHD SNP Prediction	PHD RI	I-Mutant Prediction	I-Mutant RI	MUpro
E146Q	Disease	5	Neutral	4	Decrease	8	Decrease
R168Q	Disease	4	Neutral	0	Decrease	6	Decrease
S149F	Disease	9	Disease	7	Decrease	5	Decrease
R223Q	Disease	6	Disease	6	Decrease	7	Decrease
G244W	Disease	7	Disease	8	Decrease	3	Decrease
L37C	Disease	2	Disease	5	Decrease	5	Decrease
I336C	Disease	7	Disease	6	Decrease	7	Decrease
R391C	Disease	8	Disease	7	Decrease	2	Decrease
A96C	Disease	1	Disease	7	Decrease	3	Decrease
G120I	Disease	6	Disease	1	Decrease	2	Decrease
G397I	Disease	8	Disease	7	Decrease	4	Decrease
P419I	Disease	7	Disease	6	Decrease	7	Decrease
M467A	Disease	2	Disease	1	Decrease	6	Decrease

The structural impact of the SNPs on protein structure and function was investigated using Project hope. 13 which were damaging, disease related and affect the protein stability were analyzed using Project Hope, Table 5.

Table 5. *The effect mutation on protein using Project Hope prediction.*

SNP ID	Wild	3D structure	Mutant	Effect
rs199665722 Arginine into a Glutamine R223Q				This mutation is more likely damaging to the protein. The difference in charge will disturb the ionic interaction made by the original, wild-type residue.
rs371243939 Arginine into a Cysteine R391C				The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. There is also difference in the charge between the wild and mutant type. Mutation of the residue might disturb protein function
rs374558572 Serine into a Phenylalanine S149F				The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.

SNP ID	Wild	3D structure	Mutant	Effect
rs200133602 Glycine into a Tryptophan G244W				The mutant residue is bigger than the wild-type residue. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.
rs371243939 Leucine into a Cysteine L37C				The wild-type and mutant amino acids differ in size Based on conservation scores this mutation is probably damaging to the protein.
rs371243939 Isoleucine into a Cysteine I336C				The damaging effect is due to wild-type and mutant amino acids differ in size. The mutant residue is smaller; this might lead to loss of interactions.
rs138690664 Alanine into a Cysteine A96C				The mutated residue is located in a domain that is important for the activity of the protein and in contact with residues in another domain The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit
rs376354718 Glycine into a Isoleucine G120I				The damaging effect is due to mutant residue is bigger and probably will not fit. Glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation.
rs376354718 Glycine into a Isoleucine G397I				The damaging effect is due to mutant residue is bigger and probably will not fit. Glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation.
rs376354718 Proline into a Isoleucine P419I				The damaging effect is due to increased size and conservancy. Prolines are known to have a very rigid structure, mutation changes a proline with such a function into another residue, thereby disturbing the local structure.
rs34605986 Methionine into a Alanine M467A				The damaging effect is since; the mutation is found in a conserved region of the protein and important for its activity. The mutant residue is smaller than the wild residue, causing an empty space in the core of the protein.

4. Discussion

The phenotype of HPP varies greatly in patients, even among those with identical ALPL genotype [31, 32]. It is probable that not a few patients with HPP and mild symptoms are currently mis- or under-diagnosed [33]. The input set of query sequences were submitted to evaluate the evolutionary relationship in terms of extent of similarities and differences among them. For this purpose, the input sequences were aligned by multiple sequence alignment (MSA). The aligned regions were used for nonaligned regions and considered as deletions [34]. Therefore, it has been revealed that highly unstable mutations tend to change the community structure in a protein more radically than mutations that are less unstable [35, 36].

In this study, a total of 11 SNPs were shown to be deleterious, damaging and disease related using six different software. Seven nsSNPs namely rs371243939 (L37C), rs138690664 (A96C), rs199665722 (R223Q), rs200133602 (G244W), rs371243939 (I336C), rs371243939 (R391C) and rs34605986 (M467A) have already been previously reported as pathogenic mutation in *ALP*L gene in patients with HPP through direct DNA sequencing, while the four mutations rs376354718 (G120I), rs374558572 (S149F) rs376354718 (G397I), and rs376354718 (P419I) were not reported in ClinVar database.

For rs371243939 (R391C) which was found in this study to be pathogenic using 6 different software, this mutation was predicted by other researchers who mentioned that this mutation was pathogenic and has recently been reported as one of the important mutations in *ALPL* gene, specific heterozygous variants were experimentally associated with 10.3% and 4.0% of wild-type enzyme activity, respectively,

predicting a severe phenotype [37].

These nsSNPs, rs199665722 (R223Q) and rs371243939 (R391C) were predicted by other researchers to be pathogenic, in this study they were predicted to be damaging and disease related. Some missense variants in *ALPL* have been proposed to exert dominant negative effect on enzyme function, especially those located within functional domains [2]. However, only R391C variant had been tested for this in vitro with negative results [38], the results reported here support the view that the R391C variants impair ALP protein function.

Variants predicted to be pathogenic by bioinformatics techniques were found in heterozygous form, and most were located in functionally important domains of the protein: The Ca++ binding domain, which is crucial for the main function of the ALP protein (R223Q), the crown domain involved in binding to collagen (R391C, R450C) and the homodimeric domain, which binds ALP to form the active homodimer (R391C, R450C). The R391C variant also affects a residue that is thought to be subject to phosphorylation [39].

The novel four variants rs376354718 (G120I), rs374558572 (S149F) rs376354718 (G397I), and rs376354718 (P419I) were not reported in ClinVar. The mutant residues are bigger than the wild-types and mutations of these residues can disturb interactions with other molecules or other parts of the protein. Also the damaging effect is due to loss of hydrogen bonds in the core of the protein and as a result disturb correct folding. On the other hand, rs376354718 (G120I) and rs376354718 (G397I), only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

The variants identified in a study by [40], R301W was also classified as pathogenic based on the criteria suggested by the American College of Medical Genetics and the Association for Molecular Pathology but in this study they show to have pathogenic effect of nsSNPs rs199665722 (R223Q) and rs371243939 (R391C) while the latter SNP the substituted amino acid appeared at the different position and different amino acid. In a Japanese family with low serum ALP level the mutation in the *ALPL* gene Gly82Arg has been reported [41] but in this study it was not detected, while (G244W) which is recognized in the same study as pathogenic also predicted in our study as pathogenic variant using all software.

The genotype-phenotype association in the proband's pedigree suggested that both mutations contributed individually to low serum ALP levels, combination of heterozygous ALPL mutations increased the severity of HPP [42]. Both affected residues were highly conserved and mapped to exon 12 of the *ALPL* gene. The mutations localized to the dimer interface region, and bioinformatics analysis suggested molecular disruptions within and between TNSALP monomers resulting from altered amino acid properties.

5. Conclusion

Analysis of *ALPL* gene coding region was carried out using different computational tools aiming to investigate the effect of nsSNPs on structure and function of the protein. This study reveal 11 nsSNPs as being possibly pathogenic variants. Seven of the 11 nsSNPs were already reported from previous studies by DNA sequencing, while the remaining four were predicted in this study for the first time. These are: (rs376354718 (G120I), rs374558572 (S149F) rs376354718 (G397I), and rs376354718 (P419I).

References

[1] Whyte, M. P., 2018. Hypophosphatasia and how alkaline phosphatase promotes mineralization. In Genetics of bone biology and skeletal disease (pp. 481-505). Academic Press.

[2] Mornet, E., 2018. Hypophosphatasia. Metabolism, 82, pp. 142-155.

[3] Mornet, E., 2015. Molecular genetics of hypophosphatasia and phenotype-genotype correlations. Neuronal Tissue-Nonspecific Alkaline Phosphatase (TNAP), pp. 25-43.

[4] Michigami, T., Tachikawa, K., Yamazaki, M., Kawai, M., Kubota, T. and Ozono, K., 2020. Hypophosphatasia in Japan: ALPL mutation analysis in 98 unrelated patients. Calcified Tissue International, 106 (3), pp. 221-231.

[5] Millán, J. L. and Whyte, M. P., 2016. Alkaline phosphatase and hypophosphatasia. Calcified tissue international, 98 (4), pp. 398-416.

[6] Simao, A. M. S., Bolean, M., Hoylaerts, M. F., Millán, J. L. and Ciancaglini, P., 2013. Effects of pH on the production of phosphate and pyrophosphate by matrix vesicles' biomimetics. Calcified tissue international, 93 (3), pp. 222-232.

[7] Narisawa, S., Yadav, M. C. and Millán, J. L., 2013. In vivo overexpression of tissue-nonspecific alkaline phosphatase increases skeletal mineralization and affects the phosphorylation status of osteopontin. Journal of bone and mineral research, 28 (7), pp. 1587-1598.

[8] Whyte, M. P., 2010. Physiological role of alkaline phosphatase explored in hypophosphatasia. Annals of the New York Academy of Sciences, 1192 (1), pp. 190-200.

[9] Rockman-Greenberg C. Hypophosphatasia. Pediatric endocrinology reviews: PER. 2013 Jun 1; 10: 380-8.

[10] Harris, H., 1990. The human alkaline phosphatases: what we know and what we don't know. Clinica chimica acta, 186 (2), pp. 133-150.

[11] Nguyen, H. H., van de Laarschot, D. M., Verkerk, A. J., Milat, F., Zillikens, M. C. and Ebeling, P. R., 2018. Genetic risk factors for atypical femoral fractures (AFFs): a systematic review. JBMR plus, 2 (1), pp. 1-11.

[12] Nielson, C. M., Zmuda, J. M., Carlos, A. S., Wagoner, W. J., Larson, E. A., Orwoll, E. S. and Klein, R. F., 2012. Rare coding variants in ALPL are associated with low serum alkaline phosphatase and low bone mineral density. Journal of bone and mineral research, 27 (1), pp. 93-103.

[13] Matsuura, S., Kishi, F. and Kajii, T., 1990. Characterization of a 5′-flanking region of the human liver/bone/kidney alkaline phosphatase gene: two kinds of mRNA from a single gene. Biochemical and Biophysical Research Communications, 168 (3), pp. 993-1000.

[14] Weiss, M. J., Ray, K., Henthorn, P. S., Lamb, B., Kadesch, T. and Harris, H., 1988. Structure of the human liver/bone/kidney alkaline phosphatase gene. Journal of Biological Chemistry, 263 (24), pp. 12002-12010.

[15] Mornet, E. (Ed.), The Tissue Nonspecific Alkaline Phosphatase Gene Mutations database, University of Versailles-Saint Quentin, 2020, [updated March 31, 2020]. Available from http://www.sesep.uvsq.fr/03_hypo_mutations.php.

[16] Whyte, M. P., 2016. Hypophosphatasia—aetiology, nosology, pathogenesis, diagnosis and treatment. Nature Reviews Endocrinology, 12 (4), pp. 233-246.

[17] Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D. and Morris, Q., 2018. GeneMANIA update 2018. Nucleic acids research, 46 (W1), pp. W60-W64.

[18] Kumar, P., Henikoff, S. and Ng, P. C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols, 4 (7), pp. 1073-1081.

[19] Amberger, J., Bocchini, C. A., Scott, A. F. and Hamosh, A., 2009. McKusick's online Mendelian inheritance in man (OMIM®). Nucleic acids research, 37 (suppl_1), pp. D793-D796.

[20] Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. and Chan, A. P., 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 7 (10): e46688.

[21] Choi, Y., 2012, October. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. In Proceedings of the ACM conference on bioinformatics, computational biology and biomedicine (pp. 414-417).

[22] Ramensky, V., Bork, P. and Sunyaev, S., 2002. Human non-synonymous SNPs: server and survey. Nucleic acids research, 30 (17), pp. 3894-3900.

[23] Adzhubei, I., Jordan, D. M. and Sunyaev, S. R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, *76* (1), pp. 7-20.

[24] Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. and Casadio, R., 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. Human mutation, 30 (8), pp. 1237-1244.

[25] Capriotti, E., Calabrese, R. and Casadio, R., 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 22 (22), pp. 2729-2734.

[26] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25* (17), pp. 3389-3402.

[27] Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. and Wu, C. H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics, 23 (10), pp. 1282-1288.

[28] Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B. and Casadio, R., 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics*, *14* (3), pp. 1-7.

[29] Cheng, J., Randall, A. and Baldi, P., 2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*, *62* (4), pp. 1125-1132.

[30] Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010. Protein structure analysis of mutations causing inheritable diseases: an e-Science approach with life scientist friendly interfaces. BMC Bioinformatics, *11*, pp. 548-558.

[31] Schmidt, T., Mussawy, H., Rolvien, T., Hawellek, T., Hubert, J., Rüther, W., Amling, M. and Barvencik, F., 2017. Clinical, radiographic and biochemical characteristics of adult hypophosphatasia. Osteoporosis international, 28 (9), pp. 2653-2662.

[32] Hofmann, C., Girschick, H., Mornet, E., Schneider, D., Jakob, F. and Mentrup, B., 2014. Unexpected high intrafamilial phenotypic variability observed in hypophosphatasia. European Journal of Human Genetics, 22 (10), pp. 1160-1164.

[33] Högler, W., Langman, C., Gomes da Silva, H., Fang, S., Linglart, A., Ozono, K., Petryk, A., Rockman-Greenberg, C., Seefried, L. and Kishnani, P. S., 2019. Diagnostic delay is common among patients with hypophosphatasia: initial findings from a longitudinal, prospective, global registry. BMC Musculoskeletal Disorders, 20 (1), pp. 1-9.

[34] Horiike T. Invited Mini-Review an Introduction To Molecular. *Rev. Agric. Sci.* 2016; 4: 36–45.

[35] Mishra, S. K. and Jernigan, R. L., 2018. Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics. PLoS One, 13 (6), p. e0199225.

[36] Nielsen, S. V., Stein, A., Dinitzen, A. B., Papaleo, E., Tatham, M. H., Poulsen, E. G., Kassem, M. M., Rasmussen, L. J., Lindorff-Larsen, K. and Hartmann-Petersen, R., 2017. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. PLoS Genetics, 13 (4), p. e1006739.

[37] Zurutuza, L., Muller, F., Gibrat, J. F., Taillandier, A., Simon-Bouy, B., Serre, L. and Mornet, E., 1999. Correlations of genotype and phenotype in hypophosphatasia. Human Molecular Genetics, 8 (6), pp. 1039-1046.

[38] Fauvert, D., Brun-Heath, I., Lia-Baldini, A. S., Bellazi, L., Taillandier, A., Serre, J. L., De Mazancourt, P. and Mornet, E., 2009. Mild forms of hypophosphatasia mostly result from dominant negative effect of severe alleles or from compound heterozygosity for severe and moderate alleles. *BMC medical genetics*, *10* (1), pp. 1-8.

[39] Silvent, J., Gasse, B., Mornet, E. and Sire, J. Y., 2014. Molecular evolution of the tissue-nonspecific alkaline phosphatase allows prediction and validation of missense mutations responsible for hypophosphatasia. Journal of Biological Chemistry, 289 (35), pp. 24168-24179.

[40] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E. and Voelkerding, K., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine, 17 (5), pp. 405-423.

[41] Kato, M., Michigami, T., Tachikawa, K., Kato, M., Yabe, I., Shimizu, T., Asaka, T., Kitagawa, Y. and Atsumi, T., 2021. Novel mutation in the ALPL gene with a dominant negative effect in a Japanese family. *Journal of Bone and Mineral Metabolism*, *39* (5), pp. 804-809.

[42] Martins, L., de Almeida, A. B., Dos Santos, E. J. L., Foster, B. L., Machado, R. A., Kantovitz, K. R., Coletta, R. D. and Nociti Jr, F. H., 2019. A novel combination of biallelic ALPL mutations associated with adult hypophosphatasia: A phenotype-genotype association and computational analysis study. *Bone*, *125*, pp. 128-139.