Global Journal of Medical Research: F
Diseases
Volume 22 Issue 4 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4618 & Print ISSN: 0975-5888

Computational Analysis of Possibly Pathogenic Non-Synonymous Single Nucleotide Polymorphisms Variants in HGD Gene

By Nabaa Kamal Alshafei, Intisar Hassan Saeed
& Mona Abdelrahman Mohamed Khaier

University of Bahri

Abstract- Alkaptonuria (AKU) is an autosomal recessive disorder caused by mutations in the homogentisate-1,2-dioxygenase (HGD) gene leading to the deficiency of HGD enzyme activity. The aim of this study was to use some computational bioinformatics tools to predict the most pathogenic non-synonymous mutations in the HGD gene. The data was retrieved from the SNPs database of the National Center for Biotechnology Information (dbSNPs) (Oct. 2021). The primary sequence of the protein was obtained from the UniProt database (Oct. 2021). The pathogenic effect on the protein structure and function was predicted by GeneMANIA, SIFT, Provean, Polyphen-2, I-Mutant, and Project Hope software. The human HGD gene comprises a total of 423SNPs out of that 348 were found to be synonymous, 75 were missense SNPs (nsSNPs). Analysis of the nsSNPs by SIFT predicts 35 as deleterious and 40 as tolerated ones. Using Provean only 30 were deleterious while 5 SNPs were neutral. Taking the deleterious nsSNPSs to Polyphen-2, 25 nsSNPs were damaging (22 were probably damaging and 3 were possibly damaging), while 5 were benign.

Keywords: *Alkaptonuria (AKU); homogentisate-1,2-dioxygenase (HGD) gene; I-Mutant; Non-synonymous Single Nucleotide Polymorphisms (nsSNPs); Project Hope, and SIFT.*

GJMR-F Classification: *DDC Code: 724 LCC Code: NA500*

Strictly as per the compliance and regulations of:

Computational Analysis of Possibly Pathogenic Non-Synonymous Single Nucleotide Polymorphisms Variants in HGD Gene

Nabaa Kamal Alshafei [α], Intisar Hassan Saeed [σ] & Mona Abdelrahman Mohamed Khaier [ρ]

Abstract- Alkaptonuria (AKU) is an autosomal recessive disorder caused by mutations in the homogentisate-1,2-dioxygenase (HGD) gene leading to the deficiency of HGD enzyme activity. The aim of this study was to use some computational bioinformatics tools to predict the most pathogenic non-synonymous mutations in the HGD gene. The data was retrieved from the SNPs database of the National Center for Biotechnology Information (dbSNPs) (Oct. 2021). The primary sequence of the protein was obtained from the UniProt database (Oct. 2021). The pathogenic effect on the protein structure and function was predicted by GeneMANIA, SIFT, Provean, Polyphen-2, I-Mutant, and Project Hope software. The human HGD gene comprises a total of 423SNPs out of that 348 were found to be synonymous, 75 were missense SNPs (nsSNPs). Analysis of the nsSNPs by SIFT predicts 35 as deleterious and 40 as tolerated ones. Using Provean only 30 were deleterious while 5 SNPs were neutral. Taking the deleterious nsSNPSs to Polyphen-2, 25 nsSNPs were damaging (22 were probably damaging and 3 were possibly damaging), while 5 were benign. Using SNPs&GO 11 nsSNPs were predicted as disease-related while 14 were predicted to be neutral. Project Hope analysis the mutations according to their size, charge, hydrophobicity, and conservancy. In conclusion, 7 of the predicted mutations were not reported before according to the ClinVar database while the remaining 4 were reported from patients through DNA sequencing. More research is needed to confirm these new mutations in patients.

Keywords: Alkaptonuria (AKU); homogentisate-1,2-dioxygenase (HGD) gene; I-Mutant; Non-synonymous Single Nucleotide Polymorphisms (nsSNPs); Project Hope, and SIFT.

I. Introduction

The HGD gene provides instructions for making Homogentisate oxidase enzyme, which is active mainly in the liver and kidneys. This enzyme participates in a stepwise process that breaks down two amino acids, phenylalanine and tyrosine when they are no longer needed or are present in excess. These two amino acids also play a role in making certain hormones, pigments, and brain chemicals called neurotransmitters (Aliu et al., 2018). Homogentisate oxidase is responsible for a specific step in the breakdown of phenylalanine and tyrosine. Previous steps convert the two amino acids into a molecule called homogentisic acid. Homogentisate oxidase adds two oxygen atoms to homogentisic acid, converting it to another molecule called maleylacetoacetate. Other enzymes break down maleylacetoacetate into smaller molecules that are later used for energy or to make other products that can be used by the body (Berniniet al., 2021). Mutations in the HGD gene inactivate Homogentisate oxidase by changing its structure. Without a functional version of this enzyme, phenylalanine and tyrosine are not broken down properly and homogentisic acid builds up in the body. Excess homogentisic acid and related compounds are deposited in connective tissues such as cartilage and skin, which causes them to darken. Over time, a buildup of this substance in the joints leads to arthritis. Homogentisic acid is also excreted in the urine, making the urine turn dark when exposed to air(Wilson et al., 2021).

Single Nucleotides Polymorphisms (SNPs) responsible for the maximum communal type of hereditary change in humans. Regarding throughout a coding areas of mammalian genomes, 500,000 SNPs fell into it (Shameem et al., 2021).The HGD protein protomer is composed of 445 amino acids (NP_000178.2) and is expressed in the prostate, small intestine, colon, kidney, and liver (Fernández et al., 1996), as well as in osteoarticular compartment cells (chondrocytes, synoviocytes, and osteoblasts) (Laschiet al., 2012).The enzymatic defect in AKU is caused by recessive mutations within the HGD gene (HGNC:4892), a single-copy gene that spans 54363bp of genomic sequence (3q13.33) and is split into 14 exons and codes for the HGD protomer (Zatkova and Nemethova, 2015.). The active form of the HGD protein is organized as a hexamer comprising two disc-like trimers. An intricate network of non-covalent interactions is required to maintain the spatial structure of the protomer, of the trimer andfinally of the hexamer, which can be easily disrupted by variants leading to effects on enzyme function (Titus et al., 2000). Compromising enzyme function, the missense variants are predicted to affect

Author α: Department of Biochemistry, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.
Author σ: Department of Physiology, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan.
Corresponding Author ρ: Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan. e-mail: munakhaier@gmail.com

the activity of the enzyme by three molecular mechanisms: decrease of stability of individual protomers, disruption of protomer–protomer interactions or modification of residues in the active site region (Nemethova et al., 2016).

The effects of SNPs on HGD protein structure and functions still remains elusive; therefore, in this present study, the deleterious effect of SNPs on HGD gene were analyzedby using various computational databases and bioinformatics tools. Instead of biological experiment confirmation, the study tries to provide a useful method for fast and cost effective screening for pathologic SNPs.

II. Material and Methods

a) Data retrieval

Data was retrieved from the SNP database of the National Center for Biotechnology Information (dbSNP) (http://www.ncbi.nlm.nih.gov/snp). The NCBI SNP database (https://www.ncbi.nlm.nih.gov/snp) was used to access the SNPs of the HGD gene (Oct 2021). The primary sequence of the protein (Uniprot accession number: Q93099) encoded by the HGD_HUMAN gene was obtained from the UniProt database (Oct 2021).

b) Gene MANIA software

Interaction of this gene with other genes was investigated using Gene MANIA (http://genemania.org). It is a flexible user-friendly website for generating hypotheses about gene function, analyzing gene lists, and prioritizing genes for functional assays. Given a query gene list, Gene MANIA finds functionally similar genes using a wealth of genomics and proteomics data. In this mode, it weights each functional genomic dataset according to its predictive value for the query. (Franz,et al., 2018).

c) Functional and structural analysis of the SNPs

Only missense SNPs were selected from the NCBI SNPs database as they can modify the sequence of the amino acid encoded by the protein and have 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-SNPs. The stability of the protein as the result of the mutation was studied using I- Mutant and MUPro, and finally the effect of the nsSNPs on the structure was predicted using Project Hope software.

i. *SIFT (Sorting Intolerant from Tolerant)*

This software was developed by Kumar et al., 2009. It predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT uses sequence homology among related genes and domains across species to predict the impact of all 20 possible amino acids at a given position, allowing users to determine which nsSNPs would be of most interest to study by sorting variants by this prediction score. It gives scores to each amino acid residue ranging from zero to one. 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 (Amberger et al., 2009).

ii. *Provean (Protein Variation Effect Analysis)*

Is a software tool that predicts whether an amino acid substitution has an impact on the biological function of a protein. 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 (Choi et al., 2012). 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 (Choi, 2012).

iii. *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 accession number together with sequence position with two amino acid variants. (Ramensky et al.,2002).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) (Adzhubei et al., 2013).

SNPs &GO (Single nucleotide polymorphism & Gene Ontology), PHD-SNP, (Predictor of Human Deleterious SNP)

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 (Calabrese et al., 2009).

d) Prediction of Protein stability

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

i. *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. (Capriotti et al., 2006).

ii. *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 (Cheng et al., 2006).

e) Prediction of protein modeling

This was achieved by using project Hope softwarehttps://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 PostgreSQL-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 (Das et al., 2022).

III. Results

Using Gene MANIA, the HGD gene was found to have an association with 20 other different genes. Among these is the HPD gene which provides instructions for making the 4-hydroxyphenylpyruvate dioxygenase enzyme. This gene is the second in a series of five enzymes that work to break down the amino acid tyrosine, a protein-building block found in many foods. Figure (1) and Table (1). The physical interaction and co-expression of this gene with other related genes is shown in Figure (1).

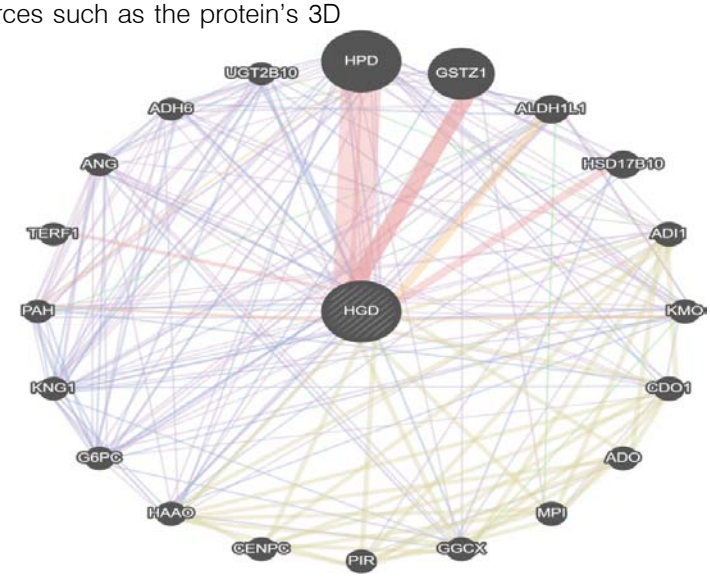


Figure 1: Gene MANIA result for HDG Gene

Table 1: Gene Description Rank Using GeneMANIA

Gene	Description
HGD	homogentisate 1,2-dioxygenase
HPD	4-hydroxyphenylpyruvate dioxygenase
GSTZ1	glutathione S-transferase zeta 1
ALDH1L1	aldehyde dehydrogenase 1 family member L1
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10
ADI1	acireductone dioxygenase 1
KMO	kynurenine 3-monooxygenase
CDO1	cysteine dioxygenase type 1
ADO	2-aminoethanethiol dioxygenase
MPI	mannose phosphate isomerase
GGCX	gamma-glutamyl carboxylase
PIR	Pirin
CENPC	centromere protein C
HAAO	3-hydroxyanthranilate 3,4-dioxygenase
G6PC	glucose-6-phosphatase catalytic subunit

KNG1	kininogen 1
PAH	phenylalanine hydroxylase
TERF1	telomeric repeat binding factor 1
ANG	Angiogenin
ADH6	alcohol dehydrogenase 6
UGT2B10	UDP glucuronosyltransferase family 2 member B10

The SNPs of the HGD gene systematically examined in this study were retrieved from the NCBI SNP database. The protein was retrieved from UniProtKB. The human HGD gene comprises a total of 423 SNPs out of that 348 were found to be synonymous, 75 were missense SNPs (nsSNPs). Analysis of the nsSNPs by SIFT predicts 35 as deleterious and 40 as tolerated ones. Using Provean only 30 were deleterious while 5 SNPs were neutral. Taking the deleterious nsSNPSs to Polyphen-2,25 nsSNPs were damaging (22were probably damaging and 3 were possibly damaging), while 5 were benign. Results were shown in Tables (2) and (3).

Table 2: The Results of Different Software

Software	Results
Retrieved SNPs	348 synonymous and 75 non- synonymous
SIFT	35 Deleterious and 40 Tolerated
Provean	30 deleterious 5 neutral
Polyphen-2	22 probably damaging 3 possibly damaging 5 benign
SNPs & GO and PHD SNPs	11 SNPs had a disease association 14 neutral

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

SNP ID	Amino Acid Change	SIFT prediction	SIFT Score	PROVEAN Prediction (cutoff= -2.5)	PROVEAN score	POLYPHEN-2 Prediction	POLYPHEN-2 Score
rs138558042	P373L	Deleterious	0.002	Deleterious	-9.802	Probably Damaging	1
rs368717991	G360R	Deleterious	0	Deleterious	-7.572	Probably Damaging	1
rs139501220	M339I	Deleterious	0.002	Deleterious	-3.677	Probably Damaging	0.976
rs143396290	D326N	Deleterious	0.045	Neutral	-2.162		
rs199927284	V316F	Deleterious	0.001	Deleterious	-4.804	Probably Damaging	0.999
rs372084813	L122F	Deleterious	0.001	Deleterious	-5.44	Probably Damaging	1
rs201529624	P114R	Deleterious	0.186	Neutral	-0.592		
rs201529624	P308R	Deleterious	0.016	Deleterious	-7.053	Benign	0.311
rs143556739	R307C	Deleterious	0.047	Deleterious	-3.91	Possibly Damaging	0.618
rs143556739	R113C	Deleterious	0.047	Deleterious	-4.403	Probably Damaging	0.999
rs372420052	T105A	Deleterious	0.023	Neutral	0		
rs372420052	T299A	Deleterious	0.025	Deleterious	-4.913	Probably Damaging	0.996
rs148641817	A293E	Deleterious	0.014	Deleterious	-3.147	Possibly Damaging	0.933
rs148641817	A99E	Deleterious	0.014	Deleterious	-7.351	Possibly Damaging	0.618
rs200382812	L85M	Deleterious	0.048	Deleterious	-7.378	Probably Damaging	0.994
rs199536408	G41D	Deleterious	0	Deleterious	-5.638	Probably Damaging	1

rs199536408	G4D	Deleterious	0	Deleterious	-4.548	Probably Damaging	0.994
rs199536408	G198D	Deleterious	0.001	Deleterious	-6.729	Probably Damaging	1
rs368256121	V24I	Deleterious	0.042	Deleterious	-2.731	Probably Damaging	0.784
rs375283568	E168K	Deleterious	0.001	Deleterious	-3.978	Probably Damaging	1
rs375396766	H117L	Deleterious	0.019	Neutral	-1.931		
rs375283568	E11K	Deleterious	0.004	Deleterious	-6.639	Probably Damaging	1
rs140543217	L163F	Deleterious	0.009	Deleterious	-3.907	Probably Damaging	0.992
rs140543217	Y6F	Deleterious	0.037	Deleterious	-3.52	Probably Damaging	1
rs375396766	P158L	Deleterious	0.018	Deleterious	-9.94	Probably Damaging	0.998
rs375396766	H117L	Deleterious	0.019	Deleterious	-3.974	Benign	0.002
rs374473331	G123E	Deleterious	0	Deleterious	-7.547	Probably Damaging	1
rs374473331	G82E	Deleterious	0.001	Deleterious	-2.778	Benign	0.005
rs143267384	E101V	Deleterious	0.032	Deleterious	-3.128	Benign	0.017
rs143267384	E60V	Deleterious	0.044	Deleterious	-13.647	Probably Damaging	1
rs370003137	S67P	Deleterious	0.011	Deleterious	-4.446	Benign	0.374
rs370003137	S26P	Deleterious	0.036	Neutral	0		
rs200808744	R53Q	Deleterious	0.019	Deleterious	-3.724	Probably Damaging	1
rs373921680	E42A	Deleterious	0.001	Deleterious	-5.639	Probably Damaging	1
rs370453859	G11E	Deleterious	0.004	Deleterious	-6.573	Probably Damaging	1

Using additional software SNPs & GO showed that 11 SNPs had a disease effect and 14 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. The results were shown in Table (4).

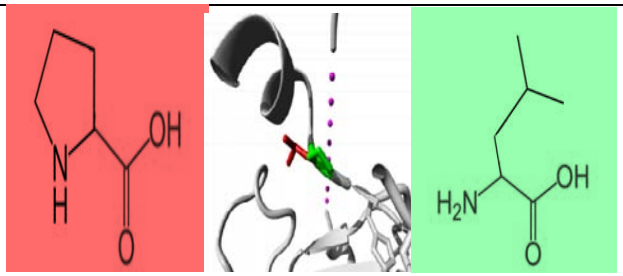
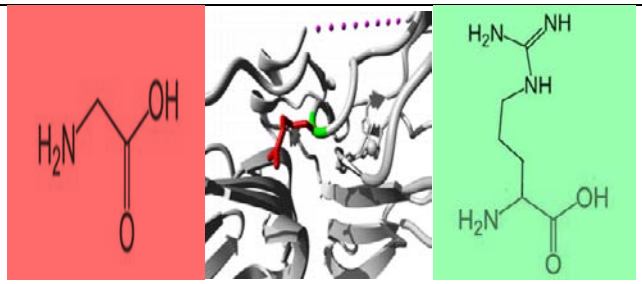
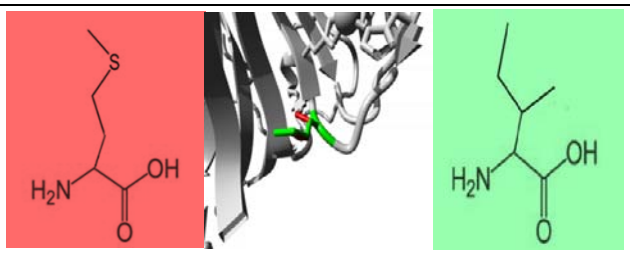
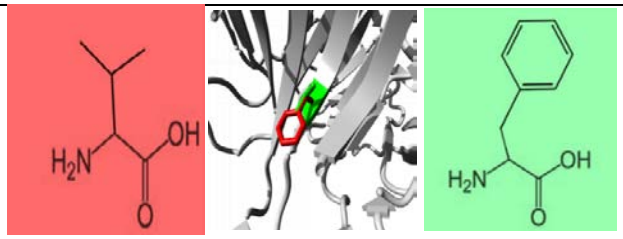
Table (4): Results of SNPs & GO, PHD SNP and I-Mutant software

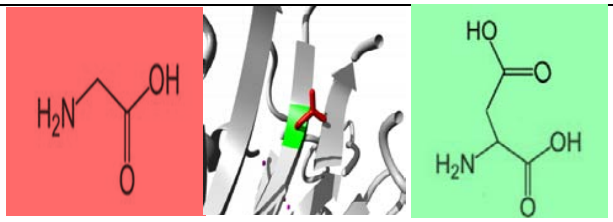
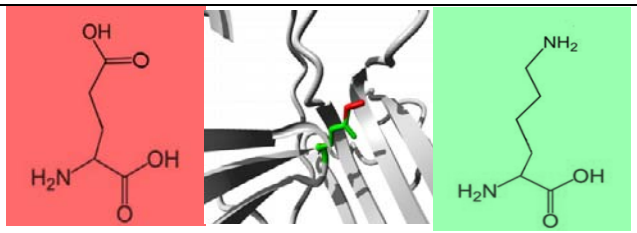
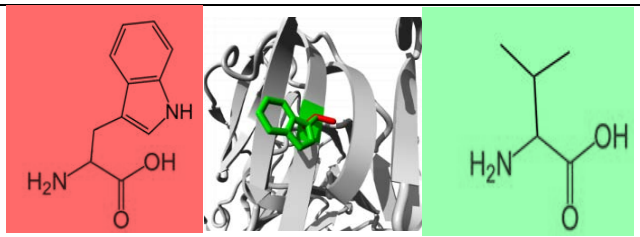
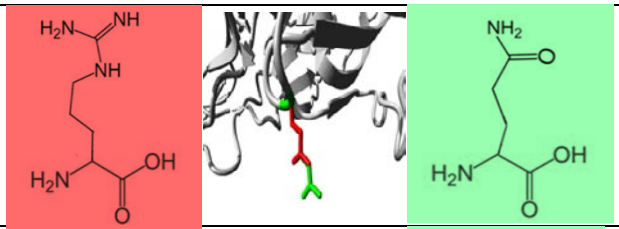
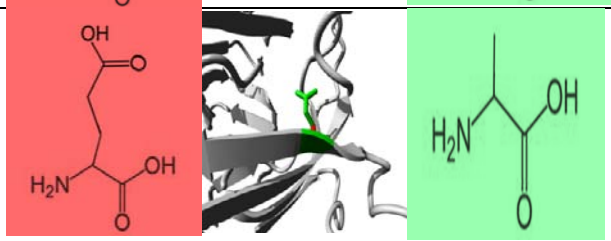
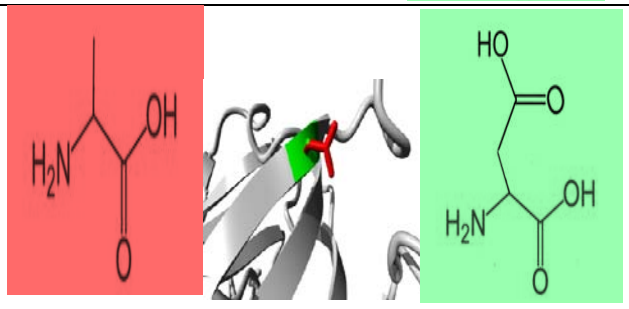
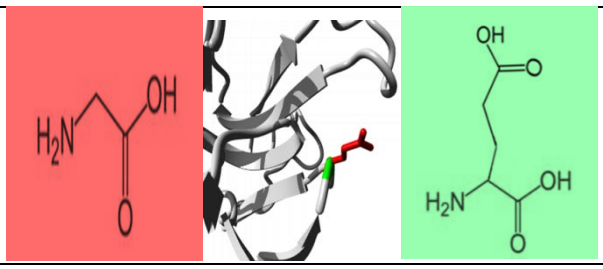
Mutation	SNP & GO Prediction	SNP & GO Probability	SNP & GORI	PHD Prediction	PHD Probability	PHD RI	I-Mutant Prediction	I-Mutant RI
L4D	Neutral	0.363	3	Neutral	0.439	4		
Y6F	Neutral	0.242	5	Disease	0.545	1		
G11E	Neutral	0.339	3	Disease	0.739	5		
G11K	Disease	0.509	0	Disease	0.782	6	Decrease	5
S24I	Neutral	0.036	9	Neutral	0,215	6		
A41D	Disease	0.527	1	Disease	0.876	8	Decrease	7
E42A	Disease	0.645	3	Disease	0.837	7	Decrease	3
R53Q	Disease	0.570	1	Disease	0.862	7	Decrease	9
W60V	Disease	0.623	2	Disease	0.843	7	Decrease	5
W85M	Neutral	0.233	8	Neutral	0.233	5		
P99E	Neutral	0.288	4	Disease	0.548	1		
V113C	Neutral	0.471	1	Disease	0.681	4		
A122F	Neutral	0.477	0	Disease	0.733	5		
G123E	Neutral	0.483	2	Disease	0.884	8		
P158L	Neutral	0.425	1	Disease	0.839	7		

L163F	Neutral	0.376	2	Disease	0.739	5		
E168K	Disease	0.602	2	Disease	0.896	8	Decrease	9
G198D	Disease	0.877	8	Disease	0.948	9	Decrease	8
A293E	Neutral	0.411	2	Disease	0.675	4		
T299A	Neutral	0.298	4	Disease	0.633	3		
R307C	Neutral	0.372	3	Neutral	0.489	0		
V316F	Disease	0.821	6	Disease	0.916	8	Decrease	3
M339I	Disease	0.522	0	Disease	0.813	6	Decrease	7
G360R	Disease	0.734	6	Disease	0.759	5	Decrease	4
P373L	Disease	0.635	3	Disease	0.612	2	Decrease	7

The structural impact of the SNPs on protein structure and function was investigated using Project hope. Eleven which were damaging, disease related and affects the protein stability were analyzed using Project Hope the results were shown in Table (5):

Table (5): The effect mutation on protein sing Project Hope prediction

SNP ID	3D structure	Effect
rs138558042 **Proline** into a **Leucine** at position 373		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.
rs368717991 **Glycine** into a **Arginine at** position 360		The damaging effect is due to difference in charge the mutation introduces a charge, this can cause repulsion, the mutant residue is bigger, this might lead to bumps. The torsion angles for this residue are unusual mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.
rs139501220 **Methionine** into a **Isoleucine** at position 339		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.
rs199927284 **Valine** into a **Phenylalanine** at position 316		The damaging effect is due to, mutant amino acids increase in size leading to the loss of interactions.

rs199536408**Glycine** into a **Aspartic Acid** at position 198		The 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 and will disturb the local structure.
rs375283568**Glutamic Acid** into a **Lysine** at position 168		The damaging effect is due to mutant residue is bigger than the wild-type and is locatedin a domain that is important for the main activity of the protein this residue might disturb this function
rs143267384 **Tryptophan** into a **Valine** at position 60		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.
rs200808744**Arginine** into a **Glutamine at** position 53		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 this function
rs373921680**Glutamic Acid** into a **Alanine** at position 42		Only this residue type was found at this position. The damaging effect is due to decrease of wild-type residue size lead to loss of interactions with other molecules or residues. Decrease hydrophobicity of the mutant residue leading to loss of Hydrophobic interactions.
rs199536408 **Alanine** into a **Aspartic Acid** at position 41		The damage may come from the fact that the mutation is at a highly conserved region. The mutant type is bigger than the wild one. It is also negatively charged while the wild is neutral. The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein.
lycine into a **Glutamic Acid at** position 11		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. charge, this can disrupts the local structure.

**Note Grey colour: protein chains, red colour atoms are the wild amino acid residues, green are the mutated amino acids*

IV. Discussion

AKU is normally characterized through genetic changes in the HGD gene but the identification of variants likely affecting structure is not always straightforward. Evolutionary conservation (Shannon entropy) and population conservation (MTR) scores indicated that AKU variants were located at more conserved residue positions. This could provide insight into novel missense variants that have a high probability of being deleterious (Ascher et al., 2019).

In this study a total of 11 SNPs were shown to be damaging, disease related and affecting the protein stability using 6 different software. Seven of them were novel not reported in ClinVar database. Namely, rs370453859 (G11K), rs199536408 (A41D), rs143267384 (W60V), rs199536408 (G198D), rs199927284 (V316F), rs139501220 (M339I), rs138558042 (P373L). The effect of the mutation on the protein function was due to the location if it is in a conserved region the protein will be highly affected. Most of these mutations were in a conserved region. Also the difference in size between the wild and mutant residue affects the protein function, if the mutant residue is bigger in size (G11K, A41D, G198D, V316F and P373L) it cannot fit and might lead to bumps, if it is smaller (W60V and M339I) this will cause an empty space in the core of the protein. The difference in hydrophobicity will leads to loss of hydrophobic interactions with other molecules on the surface of the protein. The physical properties between the wild and mutant amino acid also affects the protein function (the flexibility of Glycine and the rigidity of Proline).

Compromising the Homogentisate oxidase enzyme function Nemethova et al., 2016 showed that the missense variants are predicted to affect the activity of the enzyme by three molecular mechanisms: decrease of stability of individual protomers, disruption of protomer–protomer interactions or modfication of residues in the active site region. In agreement with our results fournsSNPs namely rs373921680 (E42A), rs200808744 (R53Q), rs375283568 (E168K), and rs368717991 (G360R), have already been previously reported as mutation in HDG gene in patients with AKU through direct DNA sequencing (de Bernabe et al., 1998; Nemethova et al, 2016; Ascher et al., 2019; Vilboux et al., 2009; Higashino, 1998; and Porfirio et al., 2000). According to, de Bernabe et al., 1998,who mentioned that rs373921680 (E42A) is pathogenic and clustered within exon 03, the variant remarks is missense, predicted mutation resulting in the amino acid substitutions affect the protomer destabilization, hexamer disruptionis crucial for the enzymatic activity of HGD. In the study by, Nemethova et al, 2015, they recorded that, SNPsrs200808744 (R53Q) was remarkably changing the amino acid residues and found to be pathogenic, and this mutation has recently been reported as one of the important mutations in this HDG gene, predicted the mutation to be highly destabilize the formation of the hexamer, because of the loss of the interactions made by the arginine. Higashino et al., 1998 approved that, rs375283568 (E168K) as a pathogenic mutation and changed a glutamic acid residue at position 168 to a lysine residue. Predicted mutation affect substitution hexamer disruption. Porfirio et al., 2000, found that, rs368717991 (G360R) affect the protomer destabilization and hexamer disruption due to substitution of wide amino Glycine into an Arginine at position 360. The mutated residue is located in a domain that is important for the main activity of the protein. Mutation of the residue might disturb this function.

These nsSNPs, rs375396766 (P158L) and rs143556739 (R307C) were predicted by other researchers to be pathogenic, in this study they were predicted to be damaging but not disease related.

The structural analysis of the identified variants allowed their classfication based on the predicted effects into three classes: (i) those that alter the active site, reducing activity; (ii) those that destabilize the protein, reducing activity; and (iii) those that prevent formation of the homohexamer, disrupting activity (Nemethova et al, 2016). Stabilizing amino acids can be predicted based on long-range interactions in protein structures and hydrophobicity and conservation of amino acid residues. Mutations found at stability centers were considered by us to be destabilizing and thus deleterious. SRide combines several methods to identify residues expected to play key roles in stabilization. It analyzes tertiary structures, rather than primary structures, and the evolutionary conserved residues contained within. A residue is predicted to be stabilizing if it is surrounded by hydrophobic residues, exhibits long-range order, has a high conservation score, and is part of a stability center (Magyar et al., 2005).

V. Conclusion

The data presented in this study represent extensive computational account of AKU nsSNPs, to filter out deleterious substitutions that are unlikely to affect protein function and can offer a more feasible means for phenotype prediction based on the biochemical severity of the amino acid substitution and the protein sequence and structural information. A total of 423 SNPs were found to be associated with mutations in HGD gene and we identified 7 novel HGD gene variants and associated intragenic polymorphisms, and they provide a general understanding of the variability at the HGD gene locus in both AKU and normal individuals, population genetics and clinical studies are important to confirm the outcomes of such study.

References Références Referencias

1. 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.
2. Aliu, E., Kanungo, S. and Arnold, G.L., 2018. Amino acid disorders. Annals of translational medicine, 6(24).
3. 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.
4. Ascher, D.B., Spiga, O., Sekelska, M., Pires, D.E., Bernini, A., Tiezzi, M., Kralovicova, J., Borovska, I., Soltysova, A., Olsson, B. and Galderisi, S., 2019. Homogentisate 1, 2-dioxygenase (HGD) gene variants, their analysis and genotype–phenotype correlations in the largest cohort of patients with AKU. European Journal of Human Genetics, 27(6), pp.888-902.
5. Bernini, A., Petricci, E., Atrei, A., Baratto, M.C., Manetti, F. and Santucci, A., 2021. A molecular spectroscopy approach for the investigation of early phase ochronotic pigment development in Alkaptonuria. Scientific reports, 11(1), pp.1-14.
6. 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.
7. 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.
8. 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.
9. 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).
10. 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.
11. Das, S.C., Rahman, M. and Das Gupta, S., 2022. In-silico analysis unravels the structural and functional consequences of non-synonymous SNPs in the human IL-10 gene. Egyptian Journal of Medical Human Genetics, 23(1), pp.1-14.
12. de Bernabé, D.B.V., Granadino, B., Chiarelli, I., Porfirio, B., Mayatepek, E., Aquaron, R., Moore, M.M., Festen, J.J.M., Sanmartí, R., Penalva, M.A. and de Córdoba, S.R., 1998. Mutation and polymorphism analysis of the human homogentisate 1, 2-dioxygenase gene in alkaptonuria patients. The American Journal of Human Genetics, 62(4), pp.776-784.
13. Fernández-Cañón, J.M., Granadino, B., De Bernabé, D.B.V., Renedo, M., Fernández-Ruiz, E., Peñalva, M.A. and De Córdoba, S.R., 1996. The molecular basis of alkaptonuria. Nature genetics, 14(1), pp.19-24.
14. 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., Web Server issue. DOI: 10.1093/var/gky311.
15. Higashino, K., 1998. A novel point mutation associated with alkaptonuria. Clin Genet, 53, pp. 228-229.. doi: 10.1111/j.1399-0004.1998.tb02684.x.
16. 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.
17. Laschi, M., Tinti, L., Braconi, D., Millucci, L., Ghezzi, L., Amato, L., Selvi, E., Spreafico, A., Bernardini, G. and Santucci, A., 2012. Homogentisate 1, 2 dioxygenase is expressed in human osteoarticular cells: implications in alkaptonuria. Journal of Cellular Physiology, 227(9), pp.3254-3257.
18. Magyar, C., Gromiha, M.M., Pujadas, G., Tusnady, G.E. and Simon, I., 2005. SRide: a server for identifying stabilizing residues in proteins. Nucleic acids research, 33(suppl_2), pp.W303-W305.
19. Nemethova, M., Radvanszky, J., Kadasi, L., Ascher, D.B., Pires, D.E., Blundell, T.L., Porfirio, B., Mannoni, A., Santucci, A., Milucci, L. and Sestini, S., 2016. Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. European Journal of Human Genetics, 24(1), pp.66-72.
20. Porfirio, B., Chiarelli, I., Graziano, C., Mannoni, A., MORRONE, A., ZAMMARCHI, E., DE BERNABÉ, D.B.V. and DE CÓRDOBA, S.R., 2000. Alkaptonuria in Italy: polymorphic haplotype background, mutational profile, and description of four novel mutations in the homogentisate 1, 2-dioxygenase gene. Journal of Medical Genetics, 37(4), pp. 309-312.
21. Ramensky, V., Bork, P. and Sunyaev, S., 2002. Human non-synonymous SNPs: server and survey. Nucleic acids research, 30(17), pp. 3894-3900.
22. Shameem, M.M., Sasikanth, S.M., Annamalai, R. and Raman, R.G., 2021. A brief review on polymer nanocomposites and its applications. Materials Today: Proceedings, 45, pp. 2536-2539.

23. Titus, G.P., Mueller, H.A., Burgner, J., Rodríguez de Córdoba, S., Peñalva, M.A. and Timm, D.E., 2000. Crystal structure of human homogentisate dioxygenase. Nature structural biology, 7(7), pp. 542-546.
24. Vilboux, T., Kayser, M., Introne, W., Suwannarat, P., Bernardini, I., Fischer, R., O'Brien, K., Kleta, R., Huizing, M. and Gahl, W.A., 2009. Mutation spectrum of homogentisic acid oxidase (HGD) in alkaptonuria. Human mutation, 30(12), pp. 1611-1619.
25. Wilson, P.J., Ranganath, L.R., Bou-Gharios, G., Gallagher, J.A. and Hughes, J.H., 2021. Expression of tyrosine pathway enzymes in mice demonstrates that homogentisate 1, 2-dioxygenase deficiency in the liver is responsible for homogentisic acid-derived ochronotic pigmentation. JIMD reports, 58(1), pp. 52-60.
26. Zatkova, A. and Nemethova, M., 2015. Genetics of alkaptonuria–an overview Genetika alkaptonúrie–prehľad. Acta Fac. Pharm. Univ. Comen. LXII, 11, pp. 27-32.