INTERNATIONAL JOURNAL OF
MEDICAL BIOCHEMISTRY

DOI: 10.14744/ijmb.2025.02328
Int J Med Biochem 2025;8(3):212–221

Research Article

Assessing the pathogenicity of missense single nucleotide polymorphisms in the human *FUCA1* gene using multiple bioinformatics tools

Israa Fahim, Hanin Idris, Rama Almohamad Alyousef, Dima Joujeh

Department of Biotechnology Enginering, University of Aleppo, Aleppo, Syria

Abstract

Objectives: Fucosidosis is a rare lysosomal storage disorder caused by mutations in the *FUCA1* gene leading to a deficiency in α-L-fucosidase. This study aimed to investigate the pathogenic missense mutations in the *FUCA1* gene and their effects on protein stability using various bioinformatics tools.

Methods: Initially, 438 missense mutations were retrieved from the NCBI database, of which 43 mutations were identified by SIFT. The impact of these mutations on protein stability was assessed using I-MUTANT2.0 and MUPRO. Additionally, protein flexibility was analyzed using MEDUSA.

Results: Among the 43 mutations, SIFT predicted 20 mutations as "deleterious". PANTHER database predicted 21 mutations as "probably damaging" and 4 mutations as "possibly damaging", PolyPhen-2 tool identified 14 mutations as "probably damaging", and 6 mutations as "possibly damaging", PHD-SNP tool predicted 21 mutations as "disease-related", PROVEAN tool predicted 27 mutations as "deleterious", PMUT tool predicted 18 mutations as "disease-related", and SNP&GO tool predicted 24 mutations as "disease-related". Nine mutations (R173Q, W145R, W188C, R36C, R162Q, R308H, R162W, G425R, A368T) were commonly predicted by all the seven tools. The impact of these mutations on protein stability was assessed using I-MUTANT2.0 and MUPRO tools. The I-MUTANT2.0 tool indicated that all nine mutations result in a decrease in protein stability. Similarly, MUPRO tool showed that eight of the nine mutations decrease protein stability, and one mutation, G425R, was found to increase protein stability. Additionally, protein flexibility was analyzed using MEDUSA tool, which revealed that the positions of all 9 SNPs were rigid, except R36C and G425R which were flexible.

Conclusion: We hope that these findings could contribute to understanding the molecular basis of diseases associated with *FUCA1* gene mutations. However, Experimental validation is recommended to confirm these results and guide future therapeutic strategies.

Keywords: *FUCA1*, fucosidosis, missense, pathogenicity prediction, single nucleotide polymorphisms

How to cite this article: Fahim I, Idris H, Alyousef RA, Joujeh D. Assessing the pathogenicity of missense single nucleotide polymorphisms in the human *FUCA1* gene using multiple bioinformatics tools. Int J Med Biochem 2025;8(3):212–221.

Fucosidosis is a rare lysosomal storage disorder [1], caused by mutations in the *FUCA1* gene leading to a deficiency in α-L-fucosidase [2]. It is inherited in an autosomal recessive pattern [1].

The disease is classified into two primary types based on age and clinical severity. Type I typically manifests before the age of one year and progresses rapidly, with affected individuals usually dying between 5 and 10 years of age. In contrast, type II generally begins later than two years old and often allows patients to survive until adulthood [3]. Clinical features include growth retardation, dysostosis multiplex, recurrent upper respiratory infections, coarse facial features, and angiokeratoma corporis diffusum [2]. Other symptoms, frequently observed in fucosidosis patients include hepatosplenomegaly, epilepsy, and inguinal hernia [3].

Address for correspondence: Dima Joujeh, Dr. Department of Biotechnology Enginering, University of Aleppo, Aleppo, Syria
Phone: 963994046745 **E-mail:** dimajoujeh@gmail.com **ORCID:** 0000-0001-8240-9886
Submitted: February 14, 2025 **Revised:** April 25, 2025 **Accepted:** May 03, 2025 **Available Online:** June 17, 2025

The *FUCA1* gene is located on chromosome 1p34.11-1p36.11 and consists of eight exons and seven introns [4]. It encodes a homotetramer protein of 461 amino acids known as α-L-fucosidase [4]. Alpha-fucosidases are hydrolytic enzymes, that act on α-L-fucosyl linkages. They belong to the glycoside hydrolase families GH29 and GH95 [5].

The enzyme is made up of subunits with varying molecular weights (50 to 60 kDa), resulting from differences in N-glycosylation and proteolytic processing [6]. A deficiency or absence of this enzyme in fucosidosis disrupts the degradation of fucosylated glycoproteins and glycolipids within lysosomes, causing the accumulation of fucosylated substrates in various tissues [7]. Both homozygous mutations and heterozygous mutations in the *FUCA1* locus can result in reduced or complete loss of α-L-fucosidase activity [8].

Bioinformatics tools have been employed to predict the effects of novel alterations on the structure and function of protein [9]. This is the first in silico study of the human *FUCA1* gene, utilizing various bioinformatics tools to analyze the structural and functional effects of missense single nucleotide polymorphisms (SNPs) in the *FUCA1* gene, which may contribute to predicting pathogenic SNPs that may serve as diagnostic markers for fucosidosis.

Materials and Methods

Data set

SNPs associated with the *FUCA1* gene were retrieved from the single nucleotide polymorphism database (dbSNP). These SNPs are identified by their unique reference sequence IDs (rsID).

Prediction of deleterious missense SNPs

To predict the functional and pathogenic effects of missense SNPs in the coding region of the *FUCA1* gene, seven bioinformatics tools were employed (Fig. 1). All tools were utilized with their default settings unless specified otherwise.

SIFT: Sorting Intolerant from Tolerant (SIFT) is a bioinformatics tool that uses homology-based sequence analysis to assess the impact of missense SNPs. It calculates normalized probabilities for all possible substitutions from the alignment to differentiate between deleterious and tolerated missense SNPs. SNPs with scores ≤0.05 are classified as deleterious, while those >0.05 are considered tolerated [10].

PolyPhen-2: Polymorphism phenotyping v2 (Polyphene-2) is a web-based tool integrates multiple sequence alignment and 3D structural analysis to predict the effects of amino acid substitutions on protein structure and function. The

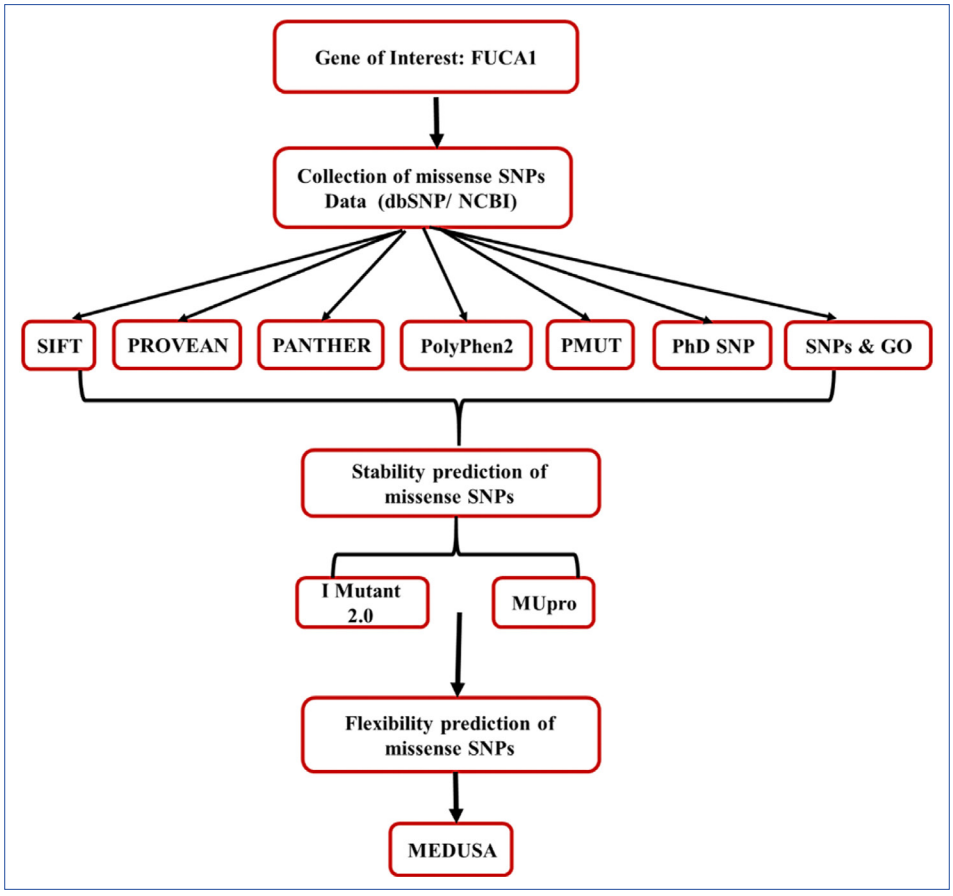

Figure 1. Workflow of in silico tools for computational analysis of *FUCA1* gene.

SNP: Single nucleotide polymorphisms; NCBI: National center for biotechnology information; SIFT: Sorting intolerant from tolerant; PROVEAN: Protein variation effect analyzer; PANTHER: Protein analysis through evolutionary relationship; PolyPhen-2: Polymorphism phenotyping v2; PMUT: Pathogenic mutation prediction.

tool calculates position-specific independent count (PSIC) scores for each of the two variants and determines the PSIC score difference between them [11]. SNPs are classified as benign (0.00–0.45), possibly damaging (0.45–0.95), or probably damaging (0.95–1) [10].

PhD-SNP: The PhD-SNP is a web server designed to assess whether a mutation occurs is benign or associated with inherited diseases, using support vector machines (SVMs). It provides an accuracy of 78% and a score range of 0–9 [12]. FASTA sequence of the corresponding protein and residue changes were submitted as inputs in the PhD-SNP server.

PANTHER: Protein Analysis Through Evolutionary Relationship (PANTHER) is a bioinformatics tool that predicts the likelihood of a mutation being harmful by estimating the evolutionary conservation of the amino acid position using a Hidden Markov Model and multiple sequence alignments [11]. The Substitution Position-Specific Evolutionary Preservation (PSEP) score indicates how long a site has been preserved during evolution:

- **Probably damaging:** >450 million years
- **Possibly damaging:** 200–450 million years
- **Probably benign:** <200 million years

PMUT: Utilizes neural networks' intelligence to predict whether SNPs are neutral or disease-causing, with an accuracy of 80% [13]. SNPs are classified as neutral (0–0.5) or pathological (0.5–1) [14].

PROVEAN: Is a web-based tool that determines the functional impact of amino acid substitutions based on sequence homology. Mutations with a score ≤-2.5 are predicted as "deleterious," while those with scores above -2.5 are considered "neutral" [13].

SNPs&GO: Single nucleotide polymorphism database and gene ontology (SNPs&GO) is a web server that combines sequence features, evolutionary data, and Gene Ontology (GO) annotations to predict disease relevance of mutations using an SVM model [14]. The result is based on the combination of Panther, PHD-SNP, and SNPs&GO results. SNPs with scores ≥0.5 are considered disease-causing, while those <0.5 are neutral.

Prediction of protein stability

Two different computational tools were used to predict the effects of single amino acid substitution on the stability of the human *FUCA1* protein.

I-mutant 2.0 is an automated web server based on SVM, that predicts changes in protein stability due to single-point mutations. It predicts the protein stability free energy change (Delta-DeltaG, or DDG) in kcal/mol. A negative DDG value indicates a decrease in stability, while a positive DDG value suggests an increase in stability [15]. DDG values for neutral mutations range from -0.5 ≤DDG ≤0.5, with values <-0.5 classified as a "large decrease" in stability and values >0.5 as a "large increase".

MUpro (mupro.proteomic.ics.uci.edu) is a web server that uses both neural network (NN) and support vector machine (SVM) models and generates confidence scores ranging between -1 and 1 [16]. A negative value indicates that the mutation is predicted to decrease protein stability, while a positive value suggests increased stability. The magnitude of the score reflects the confidence level of the prediction—the closer the score is to -1 or 1, the higher the confidence in the predicted effect.

Prediction of protein flexibility

MEDUSA is a web server used to predict the flexibility of proteins. It categorizes each residue as either flexible or rigid, considering evolutionary origin and physicochemical properties [10].

Results

Prediction of nsSNPs pathogenicity

A comprehensive study was conducted to identify pathogenic mutations in the *FUCA1* gene using seven bioinformatics tools. Initially, 438 missense mutations were collected from the dbSNP NCBI database. Of these, 43 mutations were identified by the SIFT server (Table 1), while the remaining mutations were classified as "not found." Among the 43 identified mutations, SIFT predicted 20 mutations as "deleterious" with a tolerance index score ≤0.05, and 23 mutations as 'tolerated'. These 43 mutations were subsequently analyzed using six additional bioinformatics tools. Panther predicted 21 mutations as "probably damaging" and 4 mutations as "possibly damaging" and 18 mutations were "probably benign". PolyPhen-2 identified 14 mutations as "probably damaging", 6 mutations as "possibly damaging" and 23 mutations were "benign". PHD-SNP classified 21 mutations as "disease-related" and 22 mutations as "neutral" (Table 2). PROVEAN indicated that 27 mutations were "deleterious" and 16 mutations were "neutral". PMUT categorized 18 mutations as "disease-related," 15 as "neutral" and 9 as "not found." SNP&GO predicted 24 mutations as "disease-related" and 19 as "neutral" (Table 3).

To refine the analysis, we performed an intersection of the pathogenic mutations identified across the different tools. The intersection revealed that nine missense SNPs (R173Q, W145R, W188C, R36C, R162Q, R308H, R162W, G425R, A368T) were common and were observed as deleterious, disease-associated, probably damaging, and possibly damaging (Table 4).

Prediction of protein stability

The impact of the nine mutations on protein stability was evaluated using I-MUTANT2.0 and MUPRO, based on the free energy change value (Delta Delta G, DDG) and the confidence score.

The I-MUTANT2.0 tool indicated that all nine mutations result in a decrease in protein stability. Similarly, MUPRO showed that eight of the nine mutations decrease protein stability, and one mutation, G425R, was found to increase protein stability according to MUPRO results based on stability and confidence score (Table 5).

Table 1. Tolerated and deleterious missense SNPs predicted by SIFT

No	SNP ID	CHR	REF allele	ALT allele	Amino acid change	Region	SIFT score	SIFT prediction
1	rs665	1	C	T	V260I	CDS	0.058	TOLERATED
2	rs13551	1	T	C	Q286R	CDS	0.074	TOLERATED
3	rs2070956	1	G	C	P10R	CDS	0.3	TOLERATED
4	rs61996282	1	C	G	A3P	CDS	0.084	TOLERATED
5	rs80358195	1	C	G	E380Q	CDS	0.134	TOLERATED
6	rs138378067	1	G	A	P87L	CDS	0.054	TOLERATED
7	rs143691289	1	C	G	L134F	CDS	0.159	TOLERATED
8	rs145537354	1	C	T	R178H	CDS	0.9	TOLERATED
9	rs145603001	1	C	T	R376Q	CDS	0.033	DELETERIOUS
10	rs148194937	1	C	T	R173Q	CDS	0.016	DELETERIOUS
11	rs149168482	1	G	A	S393L	CDS	0.2	TOLERATED
12	rs149540896	1	G	A	R178C	CDS	0.023	DELETERIOUS
13	rs150062050	1	A	G	W145R	CDS	0	DELETERIOUS
14	rs150422575	1	C	T	G425E	CDS	0.017	DELETERIOUS
15	rs189315801	1	C	A	W188C	CDS	0	DELETERIOUS
16	rs201209052	1	G	A	R36C	CDS	0.008	DELETERIOUS
17	rs201499886	1	C	T	E318K	CDS	0.869	TOLERATED
18	rs372045495	1	C	T	R162Q	CDS	0.021	DELETERIOUS
19	rs367641377	1	G	C	S46C	CDS	0.003	DELETERIOUS
20	rs372537257	1	C	T	R308H	CDS	0.011	DELETERIOUS
21	rs375270869	1	G	A	R162W	CDS	0	DELETERIOUS
22	rs377360830	1	C	T	A454T	CDS	0.314	TOLERATED
23	rs11549095	1	C	T	R113H	CDS	0.038	DELETERIOUS
24	rs75146878	1	C	T	D290N	CDS	0.442	TOLERATED
25	rs143545485	1	T	A	T38S	CDS	0.646	TOLERATED
26	rs149277420	1	G	A	S323L	CDS	0.336	TOLERATED
27	rs144919088	1	T	C	S287G	CDS	0.405	TOLERATED
28	rs199940255	1	G	A	R94C	CDS	0.012	DELETERIOUS
29	rs200493075	1	C	T	C235Y	CDS	0.354	TOLERATED
30	rs200678715	1	T	G	D309A	CDS	0.079	TOLERATED
31	rs202079642	1	T	C	I227V	CDS	0.255	TOLERATED
32	rs367783196	1	T	G	D344A	CDS	0.047	DELETERIOUS
33	rs202239236	1	G	A	R91C	CDS	0.033	DELETERIOUS
34	rs368369988	1	G	A	P34S	CDS	0.843	TOLERATED
35	rs368693347	1	T	C	S102G	CDS	0.082	TOLERATED
36	rs369435617	1	T	C	N144D	CDS	0.02	DELETERIOUS
37	rs371134787	1	G	A	P39L	CDS	0.001	DELETERIOUS
38	rs373007532	1	G	A	P10S	CDS	0.482	TOLERATED
39	rs373805274	1	T	C	N96S	CDS	0.079	TOLERATED
40	rs373999438	1	C	T	G425R	CDS	0.001	DELETERIOUS
41	rs374427540	1	C	T	A368T	CDS	0.003	DELETERIOUS
42	rs377191399	1	T	C	I442V	CDS	0.725	TOLERATED
43	rs2228424	1	G	A	P146L	CDS	0.009	DELETERIOUS

SNP: Single nucleotide polymorphism; SIFT: Sorting intolerant from tolerant; CHR: Chromosome; REF: Reference allele; ALT: Alternative allele; CDS: Coding DNA sequence.

Prediction of protein flexibility

Based on the flexibility prediction of *FUCA1* protein by MEDUSA (7=rigid, 2=flexible), the positions of all 9 SNPs were rigid, except R36C and G425R, which were flexible. The amino acid sequence positions W145R, W188C and A368T had a confidence score>0.75, while the positions R173Q, R36C and G425R had the confidence score of 0.6–0.75 and positions R162Q, R308H and R162W had a confidence score<0.6 (Fig. 2).

Table 2. Prediction of the pathogenicity of the missense SNPs using PANTHER, PolyPhen-2, and PHD-SNP servers

No	SNP ID	PANTHER		PolyPhen-2		PHD-SNP	
		Score (Pdel)	Prediction	Score	Prediction	Score	Prediction
1	rs665	0.86	PROBABLY DAMAGING	0.538	POSSIBLY DAMAGING	6	NEUTRAL
2	rs13551	0.19	PROBABLY BENIGN	0.001	BENIGN	6	NEUTRA
3	rs2070956	0.27	PROBABLY BENIGN	0.272	BENIGN	7	NEUTRAL
4	rs61996282	0.13	PROBABLY BENIGN	0.000	BENIGN	9	NEUTRAL
5	rs80358195	0.57	PROBABLY DAMAGING	0.007	BENIGN	3	DISEASE
6	rs138378067	0.85	PROBABLY DAMAGING	0.029	BENIGN	2	DISEASE
7	rs143691289	0.86	PROBABLY DAMAGING	0.112	BENIGN	2	NEUTRAL
8	rs145537354	0.27	PROBABLY BENIGN	0.003	BENIGN	2	NEUTRAL
9	rs145603001	0.86	PROBABLY DAMAGING	0.351	BENIGN	6	DISEASE
10	rs148194937	0.86	PROBABLY DAMAGING	0.956	POSSIBLY DAMAGING	8	DISEASE
11	rs149168482	0.5	POSSIBLY DAMAGING	0.004	BENIGN	8	DISEASE
12	rs149540896	0.27	PROBABLY BENIGN	0.998	PROBABLY DAMAGING	8	DISEASE
13	rs150062050	0.95	PROBABLY DAMAGING	0.999	PROBABLY DAMAGING	5	DISEASE
14	rs150422575	0.86	PROBABLY DAMAGING	0.992	PROBABLY DAMAGING	7	DISEASE
15	rs189315801	0.86	PROBABLY DAMAGING	1.000	PROBABLY DAMAGING	7	DISEASE
16	rs201209052	0.74	PROBABLY DAMAGING	0.954	POSSIBLY DAMAGING	2	DISEASE
17	rs201499886	0.5	POSSIBLY DAMAGING	0.307	BENIGN	8	NEUTRAL
18	rs372045495	0.86	PROBABLY DAMAGING	1.000	PROBABLY DAMAGING	7	DISEASE
19	rs367641377	0.27	PROBABLY BENIGN	0.988	PROBABLY DAMAGING	6	DISEASE
20	rs372537257	0.86	PROBABLY DAMAGING	0.976	PROBABLY DAMAGING	2	DISEASE
21	rs375270869	0.86	PROBABLY DAMAGING	1.000	PROBABLY DAMAGING	8	DISEASE
22	rs377360830	0.19	PROBABLY BENIGN	0.000	BENIGN	5	NEUTRAL
23	rs11549095	0.86	PROBABLY DAMAGING	0.008	BENIGN		NEUTRAL
24	rs75146878	0.27	PROBABLY BENIGN	0.001	BENIGN	8	NEUTRAL
25	rs143545485	0.57	PROBABLY DAMAGING	0.97	BENIGN	2	NEUTRAL
26	rs149277420	0.27	PROBABLY BENIGN	0.014	BENIGN	7	NEUTRAL
27	rs144919088	0.19	PROBABLY BENIGN	0.001	BENIGN	8	NEUTRAL
28	rs199940255	0.19	PROBABLY BENIGN	0.981	PROBABLY DAMAGING	5	DISEASE
29	rs200493075	0.27	PROBABLY BENIGN	0.000	BENIGN	3	DISEASE
30	rs200678715	0.19	PROBABLY BENIGN	0.006	BENIGN	4	DISEASE
31	rs202079642	0.57	PROBABLY DAMAGING	0.476	POSSIBLY DAMAGING	5	NEUTRAL
32	rs367783196	0.5	POSSIBLY DAMAGING	0.124	BENIGN	7	DISEASE
33	rs202239236	0.27	PROBABLY BENIGN	0.989	PROBABLY DAMAGING	3	DISEASE
34	rs368369988	0.27	PROBABLY BENIGN	0.000	BENIGN	5	NEUTRAL
35	rs368693347	0.27	PROBABLY BENIGN	0.005	BENIGN	7	NEUTRAL
36	rs369435617	0.86	PROBABLY DAMAGING	0.992	PROBABLY DAMAGING	1	NEUTRAL
37	rs371134787	0.86	PROBABLY DAMAGING	0.997	PROBABLY DAMAGING	0	NEUTRAL
38	rs373007532	0.27	PROBABLY BENIGN	0.013	BENIGN	9	NEUTRAL
39	rs373805274	0.86	PROBABLY DAMAGING	0.576	POSSIBLY DAMAGING	1	NEUTRAL
40	rs373999438	0.86	PROBABLY DAMAGING	1.000	PROBABLY DAMAGING	6	DISEASE
41	rs374427540	0.86	PROBABLY DAMAGING	0.628	POSSIBLY DAMAGING	8	DISEASE
42	rs377191399	0.27	PROBABLY BENIGN	0.000	BENIGN	8	NEUTRAL
43	rs2228424	0.5	POSSIBLY DAMAGING	0.996	PROBABLY DAMAGING	2	NEUTRAL

SNP: Single nucleotide polymorphisms; PANTHER: Protein analysis through evolutionary relationship; PolyPhen: Polymorphism phenotyping; PHD-SNP: Predictor of human deleterious single nucleotide polymorphisms.

Discussion

The human genome shares approximately 99.9% identical DNA sequences, with the remaining 0.1% accounting for individual variations [17]. SNPs are the most common type of DNA alterations, involving changes at a single nucleotide within the genomic sequence [12]. These genomic variations can affect protein structure and function, potentially

Table 3. Prediction of the pathogenicity of the missense SNPs using PROVEAN, PMUT, SNP&GO servers

No	SNP ID	PROVEAN		PMUT		SNP&GO		
		Score	Prediction	Score	Prediction	RI	Probability	Prediction
1	rs665	-0.79	NEUTRAL	-	-	6	0.191	NEUTRAL
2	rs13551	-	NEUTRAL	-	-	6	0.206	NEUTRAL
3	rs2070956	-0.15	NEUTRAL	-	-	3	0.331	NEUTRAL
4	rs61996282	-0.34	NEUTRAL	0.43 (85%)	NEUTRAL	8	0.116	NEUTRAL
5	rs80358195	-2.23	NEUTRAL	0.64 (84%)	DISEASE	1	0.428	NEUTRAL
6	rs138378067	-6.08	DELETERIOUS	-	-	1	0.539	DISEASE
7	rs143691289	-3.27	DELETERIOUS	0.30 (89%)	NEUTRAL	2	0.386	NEUTRAL
8	rs145537354	2.10	NEUTRAL	-	-	2	0.416	NEUTRAL
9	rs145603001	-2.71	DELETERIOUS	0.85 (82%)	DISEASE	0	0.492	NEUTRAL
10	rs148194937	-3.94	DELETERIOUS	0.39 (86%)	NEUTRAL	7	0.867	DISEASE
11	rs149168482	-2.99	DELETERIOUS	0.88 (92%)	DISEASE	6	0.777	DISEASE
12	rs149540896	-4.13	DELETERIOUS	0.90 (93%)	DISEASE	6	0.781	DISEASE
13	rs150062050	-13.98	DELETERIOUS	-	-	7	0.861	DISEASE
14	rs150422575	-7.46	DELETERIOUS	0.89 (92%)	DISEASE	5	0.760	DISEASE
15	rs189315801	-12.21	DELETERIOUS	0.55 (81%)	DISEASE	8	0.921	DISEASE
16	rs201209052	-4.01	DELETERIOUS	0.39 (86%)	NEUTRAL	2	0.602	DISEASE
17	rs201499886	-0.83	NEUTRAL	0.70 (86%)	DISEASE	8	0.122	NEUTRAL
18	rs372045495	-3.99	DELETERIOUS	0.86 (91%)	DISEASE	7	0.826	DISEASE
19	rs367641377	-3.62	DELETERIOUS	0.51 (79%)	DISEASE	2	0.578	DISEASE
20	rs372537257	-4.32	DELETERIOUS	0.90 (93%)	DISEASE	2	0.604	DISEASE
21	rs375270869	-7.99	DELETERIOUS	0.02 (98%)	NEUTRAL	8	0.895	DISEASE
22	rs377360830	-0.72	NEUTRAL	0.10 (96%)	NEUTRAL	6	0.177	NEUTRAL
23	rs11549095	-0.83	NEUTRAL	0.14 (94%)	NEUTRAL	6	0.212	NEUTRAL
24	rs75146878	-0.09	NEUTRAL	-	-	7	0.153	NEUTRAL
25	rs143545485	-1.14	NEUTRAL	-	-	7	0.127	NEUTRAL
26	rs149277420	-3.27	DELETERIOUS	0.18 (93%)	NEUTRAL	7	0.152	NEUTRAL
27	rs144919088	-2.10	NEUTRAL	0.11 (95%)	NEUTRAL	8	0.089	NEUTRAL
28	rs199940255	-4.18	DELETERIOUS	0.53 (80%)	DISEASE	3	0.645	DISEASE
29	rs200493075	-2.59	DELETERIOUS	0.13 (94%)	NEUTRAL	2	0.600	DISEASE
30	rs200678715	-0.51	NEUTRAL	0.42 (85%)	NEUTRAL	2	0.618	DISEASE
31	rs202079642	-6.43	DELETERIOUS	0.31 (89%)	NEUTRAL	3	0.361	NEUTRAL
32	rs367783196	-3.02	DELETERIOUS	0.64 (84%)	DISEASE	7	0.840	DISEASE
33	rs202239236	0.41	NEUTRAL	0.81 (89%)	DISEASE	1	0.567	DISEASE
34	rs368369988	-3.01	DELETERIOUS	0.07 (97%)	NEUTRAL	7	0.163	NEUTRAL
35	rs368693347	-3.93	DELETERIOUS	0.11 (95%)	NEUTRAL	6	0.210	NEUTRAL
36	rs369435617	-9.14	DELETERIOUS	0.82 (90%)	DISEASE	2	0.417	NEUTRAL
37	rs371134787	0.31	NEUTRAL	0.80 (89%)	DISEASE	1	0.454	NEUTRAL
38	rs373007532	-4.39	DELETERIOUS	0.05 (97%)	NEUTRAL	7	0.159	NEUTRAL
39	rs373805274	-7.95	DELETERIOUS	0.72 (87%)	DISEASE	3	0.362	NEUTRAL
40	rs373999438	-3.73	DELETERIOUS	0.76 (88%)	DISEASE	6	0.782	DISEASE
41	rs374427540	0.13	NEUTRAL	0.60 (83%)	DISEASE	5	0.728	DISEASE
42	rs377191399	-6.72	DELETERIOUS	0.29 (90%)	NEUTRAL	9	0.048	NEUTRAL
43	rs2228424	-3.93	DELETERIOUS	-	-	0	0.505	DISEASE

SNP: Single nucleotide polymorphisms; PROVEAN: Protein variation effect analyzer; PMUT: Pathogenic mutation prediction; SNP&GO: Single nucleotide polymorphism&gene ontology.

altering the normal characteristics of an organism. A single nucleotide substitution can result in either a missense or nonsense mutation. In a missense mutation, one amino acid is replaced by another, whereas nonsense mutations replace a coding codon with a stop codon, leading to protein truncation [18].

Table 4. Common deleterious SNPs in *FUCA1* gene as predicted by seven in silico tools

No	SNP ID	Amino acid	SIFT	PMUT	PANTHER	PROVEAN	PHD_SNP	SNP_GO	POLYPHEN-2
1	rs148194937	R173Q	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
2	rs150062050	W145R	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
3	rs189315801	W188C	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
4	rs201209052	R36C	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
5	rs372045495	R162Q	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
6	rs372537257	R308H	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
7	rs375270869	R162W	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
8	rs373999438	G425R	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging
9	rs374427540	A368T	DELETERIOUS	DISEASE	Probably damaging	DELETERIOUS	DISEASE	DISEASE	Possibly damaging

SNP: Single nucleotide polymorphisms; SIFT: Sorting intolerant from tolerant; PMUT: Pathogenic mutation prediction; PANTHER: Protein analysis through evolutionary relationship; PROVEAN: Protein variation effect analyzer; PHD_SNP: Predictor of human deleterious single nucleotide polymorphisms; SNP&GO: Single nucleotide polymorphism&gene ontology.

Table 5. The effect of missense SNPs on protein stability

No	SNP ID	I-Mutant 2.0		MUpro		MUpro (SVM)		MUpro (neural network)	
		Stability	DDG	Stability	DDG	Stability	Confidence score	Stability	Confidence score
1	rs148194937	Decrease	-0.42	Decrease	-0.92804456	Decrease	-0.53267036	Decrease	**-0.81992480**
2	rs150062050	Decrease	-1.11	Decrease	-0.79418806	Decrease	-0.71590402	Decrease	**-0.72355086**
3	rs189315801	Decrease	-1.94	Decrease	-1.2967016	Decrease	-1	Decrease	**-0.96074062**
4	rs201209052	Decrease	-0.79	Decrease	-0.93061864	Decrease	-1	Decrease	**-0.90201448**
5	rs372045495	Decrease	-1.27	Decrease	-1.0057917	Decrease	-0.4656619	Decrease	**-0.86043601**
6	rs372537257	Decrease	-2.02	Decrease	-1.403344	Decrease	-0.7006115	Decrease	**-0.85234457**
7	rs375270869	Decrease	-0.72	Decrease	-0.82555437	Decrease	-0.12250558	Decrease	**-0.77772743**
8	rs373999438	Decrease	-1.82	Decrease	-0.41414478	Decrease	0.22525227	Decrease	**0.6359362**
9	rs374427540	Decrease	-0.71	Decrease	-0.54392681	Decrease	-0.25394459	Decrease	**-0.78975651**

SNP: Single nucleotide polymorphisms; SVM: Support vector machine; DDG: Delta delta G.

Identifying SNPs that impact morphology is essential for understanding the genetic basis of diseases and phenotypic diversity. Furthermore, this knowledge aids in selecting markers for use in population-based association studies [12].

In the current study, the pathogenic missense mutations in the *FUCA1* gene and their effects on protein stability and flexibility were assessed using various bioinformatics tools.

The findings illustrate significant variability in the predictions among different tools, emphasizing the importance of cross-referencing outputs to improve the reliability of pathogenic mutation identification. We performed an intersection of the pathogenic mutations identified across the different tools. This approach was employed to ensure greater accuracy by focusing on mutations consistently predicted as pathogenic by multiple tools. Nine mutations were commonly identified as pathogenic across all tools. This consistency in results between different independent tools, which evaluate various aspects of protein structure, function, and evolutionary conservation, adds strength to the reliability of these predictions. Such predictions suggest that these mutations may disrupt the normal folding and stability of the *FUCA1* protein, potentially leading to the pathogenic phenotypes observed in diseases like fucosidosis. Furthermore, these mutations may interfere with the protein's function by altering critical functional sites, which could include regions essential for substrate binding or catalysis, ultimately impairing protein activity.

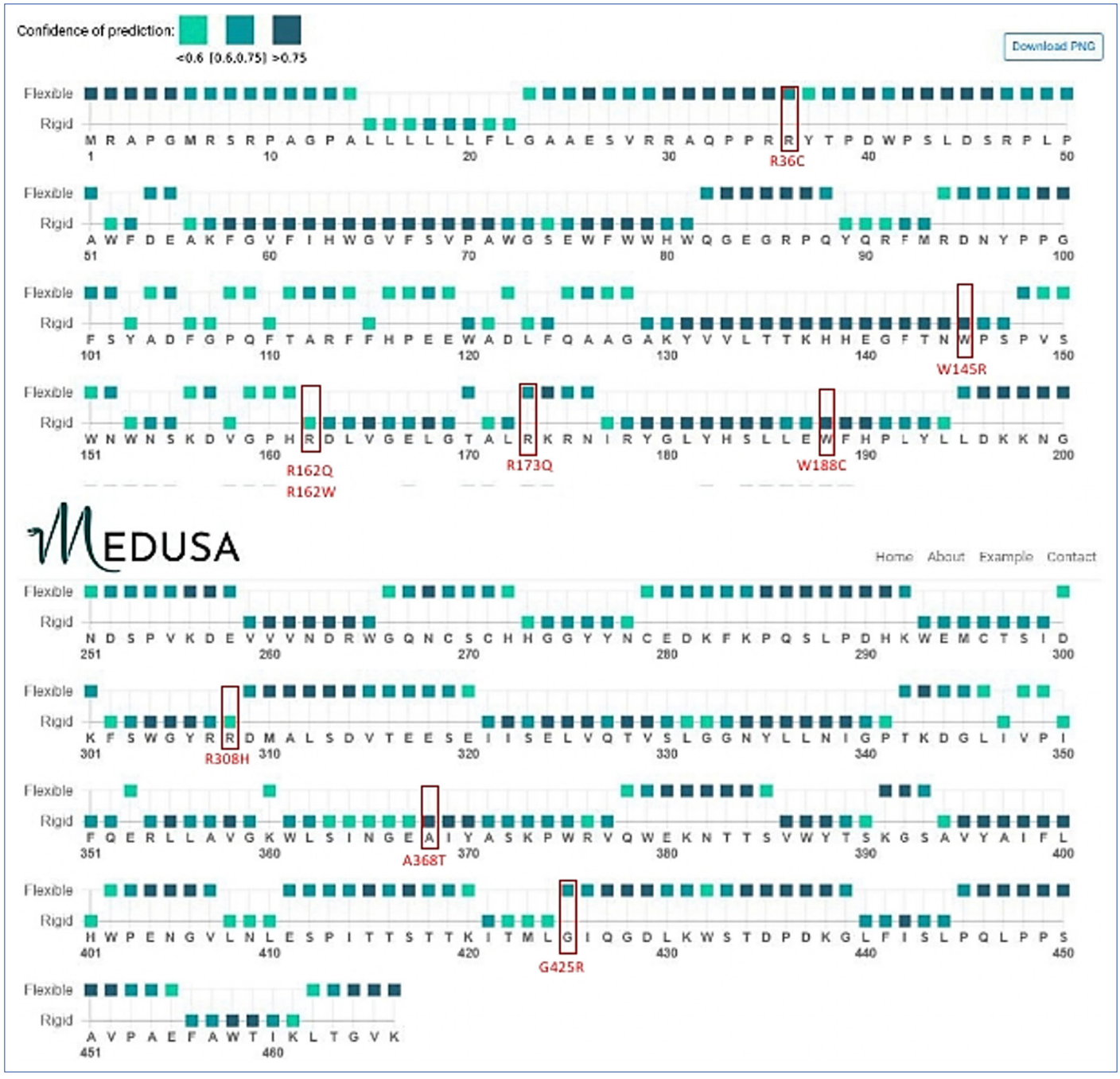

Figure 2. Flexibility prediction of missense SNPs on α-L-fucosidase protein.

SNP: Single nucleotide polymorphisms.

Rs148194937 (C/T mutation) leads to the conversion of arginine (R) to glutamine (Q) at position 173 (R173Q). The molecular weight of glutamine is smaller than that of arginine, and unlike arginine, which is positively charged, glutamine is neutral at physiological pH. This mutation can result in the loss of protein function due to changes in charge and potential alterations in hydrogen bonding.

Rs150062050 (A/G mutation) results in the conversion of tryptophan (W) to arginine (R) at position 145 (W145R). The molecular weight of arginine is smaller than that of tryptophan, and arginine is positively charged. The elongated side chain of arginine can form multiple hydrogen bonds, leading to incorrect folding and aggregation of the protein, thus impairing its function.

Rs189315801 (C/A mutation) causes the conversion of tryptophan (W) to cysteine (C) at position 188 (W188C). Tryptophan has a higher molecular weight than cysteine, and the thiol group (-SH) of cysteine is a nucleophile, making it reactive and prone to forming disulfide bonds. This mutation can disrupt the protein's stability and lead to a loss of function.

Rs201209052 (G/A mutation) leads to the substitution of arginine (R) with cysteine (C) at position 36 (R36C). Cysteine has a lower molecular weight than arginine and contains a thiol group. This mutation is known to cause disease due to its impact on protein folding and stability.

Rs372045495 (C/T mutation) causes the conversion of arginine (R) to glutamine (Q) at position 162 (R162Q). Glutamine has a smaller molecular weight than arginine and is neutral at physiological pH. This mutation has been established as deleterious, likely due to the loss of positive charge.

Rs372537257 (C/T mutation) results in the substitution of arginine (R) with histidine (H) at position 308 (R308H). Histidine has a smaller molecular weight than arginine but retains a positive charge under physiological conditions. This mutation is considered deleterious due to its impact on the protein's functional and structural properties.

Rs375270869 (G/A mutation) causes the conversion of arginine (R) to tryptophan (W) at position 162 (R162W). Tryptophan has a higher molecular weight than arginine and is neutral. This substitution disrupts the charge balance and is predicted to be damaging to the protein due to altered hydrophobicity and structural interactions.

Rs373999438 (C/T mutation) leads to the replacement of glycine (G) with arginine (R) at position 425 (G425R). Arginine has a higher molecular weight and is polar, whereas glycine is non-polar and lacks a substantial side chain. This mutation introduces a charged residue, which disrupts the local structural conformation, making it damaging to the protein.

Rs374427540 (C/T mutation) results in the substitution of alanine (A) with threonine (T) at position 368 (A368T). Threonine has a higher molecular weight than alanine and is polar due to its hydroxyl group, allowing it to form hydrogen bonds. Alanine is non-polar and cannot form such interactions. This mutation is deleterious because it alters the hydrophobicity and hydrogen bonding potential of the protein [19].

After identifying the nine mutations common across the seven tools, their impact on protein stability was evaluated using I-MUTANT2.0 and MUPRO. The I-MUTANT2.0 tool indicated that all nine mutations result in a decrease in protein stability. This aligns with the general understanding that missense mutations in conserved regions of proteins often destabilize their native structures, rendering them nonfunctional. MUPRO showed that eight of the nine mutations decrease protein stability, and one mutation, G425R, was found to increase protein stability, suggesting that this mutation might lead to a more compact protein structure. However, this does not necessarily indicate a functional gain, as increased stability could potentially hinder the protein's conformational flexibility or its ability to interact with substrates or other proteins, further complicating the interpretation of its biological effect. The effect of G425R warrants further experimental investigation, as in silico predictions alone may not fully capture the nuances of its impact on protein function.

However, the discrepancy observed in the G425R mutation may be attributed to the computational methods used by MUPRO, which include SVM and neural network approaches that incorporate additional structural or physicochemical features into their predictions. These methods may consider certain compensatory effects or unique interactions within the protein structure that stabilize the protein despite the mutation. Neural networks are designed to recognize complex patterns and may account for subtle, non-linear interactions in the protein structure that are missed by SVM or other simpler algorithms. For instance, the G425R mutation might induce specific structural changes, such as enhanced intramolecular interactions or improved packing within the protein, that offset the destabilizing effects typically caused by mutations. These stabilizing effects could be captured by the neural network's ability to analyze a broader range of protein features, leading to the prediction of increased stability for this mutation. However, this variation underscores the complexity of protein stability predictions and highlights the importance of using multiple tools to gain a more comprehensive understanding of mutation impacts. Further experimental studies, such as molecular dynamics simulations or thermal stability assays, are recommended to validate these computational predictions and elucidate the precise effect of the G425R mutation on protein stability.

The MEDUSA analysis was conducted to evaluate the flexibility of the protein, and the results revealed that the positions of all 9 SNPs were rigid, except R36C and G425R, which were flexible.

Rigid regions in a protein are often critical for maintaining structural integrity and ensuring proper folding and stability. Mutations in these regions are likely to disrupt the protein's core structure, leading to a loss of functional efficiency or complete misfolding. Such structural disturbances may explain the pathogenic nature of these mutations, as *FUCA1* plays a key enzymatic role, and any alteration in its stable regions could hinder its catalytic activity.

On the other hand, flexible regions are typically involved in dynamic processes, such as substrate binding, enzyme activation, or interactions with other biomolecules. Mutations in these regions might alter the dynamic behavior of the protein, leading to improper substrate accommodation or destabilization of transient conformations necessary for enzymatic function.

Conclusion

This study analyzed the effects of amino acid variations on α-L-fucosidase protein structure, function, and disease association. Among 438 mutations initially retrieved, nine were identified as commonly pathogenic across seven computational tools. Most mutations decrease protein stability, with one exception (G425R), which showed an increase in stability under specific conditions. Additionally, flexibility analysis using MEDUSA demonstrated that seven mutations were located in rigid regions of the protein, while two were found in flexible regions. These findings highlight the importance of computational approaches to understand the molecular basis of *FUCA1*-related diseases. Experimental validation is recommended to confirm the precise effects of these mutations and their contributions to disease phenotypes and guide future therapeutic strategies.

Informed Consent: The study is purely bioinformatics-based and does not involve any human subjects or patient data.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Funding: This manuscript is supported by University of Aleppo, Syria.

Use of AI for Writing Assistance: No AI technologies utilized.

Authorship Contributions: Concept – D.J.; Design – D.J.; Supervision – D.J.; Data collection and/or processing – I.F., R.A.A., H.I., D.J.; Data analysis and/or interpretation – D.J., I.F.; Literature search – I.F., R.A.A., H.I., D.J.; Writing – I.F., R.A.A., H.I., D.J.; Critical review – D.J.

Peer-review: Externally peer-reviewed.

References

1. Zhang X, Zhao S, Liu H, Wang X, Wang X, Du N, et al. Identification of a novel homozygous loss-of-function mutation in FUCA1 gene causing severe fucosidosis: A case report. J Int Med Res 2021;49(4):3000605211005975. [CrossRef]
2. Johnson WG. Disorders of glycoprotein degradation: Sialidosis, fucosidosis, α-mannosidosis, β-mannosidosis, and aspartylglycosaminuria. In Rosenberg RN, editor. Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease. 6th ed. Academic Press; 2020. p. 519–34. [CrossRef]
3. Mao SJ, Zhao J, Shen Z, Zou CC. An unusual presentation of fucosidosis in a Chinese boy: A case report and literature review (childhood fucosidosis). BMC Pediatr 2022;22(1):403. [CrossRef]
4. Puente-Ruiz N, Ellis I, Bregu M, Chen C, Church HJ, Tylee KL, et al. Long-term outcomes in two adult siblings with Fucosidosis - Diagnostic odyssey and clinical manifestations. Mol Genet Metab Rep 2023;37:101009. [CrossRef]
5. Klontz EH, Li C, Kihn K, Fields JK, Beckett D, Snyder GA, et al. Structure and dynamics of an α-fucosidase reveal a mechanism for highly efficient IgG transfucosylation. Nat Commun 2020;11(1):6204. [CrossRef]
6. Stepien KM, Ciara E, Jezela-Stanek A. Fucosidosis-clinical manifestation, long-term outcomes, and genetic profile-review and case series. Genes (Basel) 2020;11(11):1383. [CrossRef]
7. Bhattacherjee A, Desa E, Lone KA, Jaiswal A, Tyagi S, Dalal A. Genotype first approach & familial segregation analysis help in the elucidation of disease-causing variant for fucosidosis. Indian J Med Res 2023;157(4):363–6. [CrossRef]
8. Domin A, Zabek T, Kwiatkowska A, Szmatola T, Deregowska A, Lewinska A, et al. The identification of a novel fucosidosis-associated FUCA1 mutation: A case of a 5-year-old Polish girl with two additional rare chromosomal aberrations and affected DNA methylation patterns. Genes (Basel) 2021;12(1):74. [CrossRef]
9. Chkioua L, Amri Y, Chaima S, Fenni F, Boudabous H, Ben Turkia H, et al. Fucosidosis in Tunisian patients: Mutational analysis and homology-based modeling of FUCA1 enzyme. BMC Med Genomics 2021;14(1):208. Erratum in: BMC Med Genomics. 2021;14(1):293. [CrossRef]
10. Kamal MM, Mia MS, Faruque MO, Rabby MG, Islam MN, Talukder MEK, et al. In silico functional, structural and pathogenicity analysis of missense single nucleotide polymorphisms in human MCM6 gene. Sci Rep 2024;14(1):11607. [CrossRef]
11. Abuzaid O, Idris AB, Yılmaz S, Idris EB, Idris LB, Hassan MA. Prediction of the most deleterious non-synonymous SNPs in the human IL1B gene: Evidence from bioinformatics analyses. BMC Genom Data 2024;25(1):56. [CrossRef]
12. Waheed S, Ramzan K, Ahmad S, Khan MS, Wajid M, Ullah H, et al. Identification and In-silico study of non-synonymous functional SNPs in the human SCN9A gene. PLoS One 2024;19(2):e0297367. Erratum in: PLoS One 2024;19(3):e0301489. [CrossRef]
13. Mustafa MI, Murshed NS, Abdelmoneim AH, Makhawi AM. In silico analysis of the functional and structural consequences of SNPs in human ARX gene associated with EIEE1. Inform Med Unlocked 2020;21:100447. [CrossRef]
14. Azmi MB, Khan W, Azim MK, Nisar MI, Jehan F. Identification of potential therapeutic intervening targets by in-silico analysis of nsSNPs in preterm birth-related genes. PLoS One 2023;18(3):e0280305. [CrossRef]
15. Pavithran H, Kumavath R. In silico analysis of nsSNPs in CYP19A1 gene affecting breast cancer associated aromatase enzyme. J Genet 2021;100:23. [CrossRef]
16. Sattar MMK, Anjum AA, Chang YF, Yaqub T, Aslam A. In silico analysis of non-synonymous single nucleotide polymorphisms in α-toxin of Clostridium perfringens toxinotype B isolated from lamb dysentery cases in Pakistan. Pakistan J Zool 2024;1–9. [CrossRef]
17. Hossain MS, Roy AS, Islam MS. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Sci Rep 2020;10(1):14542. [CrossRef]
18. Ali MZ, Farid A, Ahmad S, Muzammal M, Mohaini MA, Alsalman AJ, et al. In silico analysis identified putative pathogenic missense nsSNPs in human SLITRK1 gene. Genes (Basel) 2022;13(4):672. [CrossRef]
19. National Library of Medicine. PubChem Compound. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Glutamine. Accessed May 20, 2025.