Heliyon 10 (2024) e37280

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Unraveling the potential effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on the Protein structure and function of the human *SLC30A8* gene on type 2 diabetes and colorectal cancer: An *In silico* approach

Md. Moin Uddin [a,1], Md. Tanvir Hossain [a,1], Md. Arju Hossain [b,c], Asif Ahsan [a], Kamrul Hasan Shamim [c], Md. Arif Hossen [d], Md. Shahinur Rahman [e], Md Habibur Rahman [f], Kawsar Ahmed [g,h,i,*], Francis M. Bui [g], Fahad Ahmed Al-Zahrani [j,**]

[a] *Department of Biotechnology, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh*
[b] *Department of Microbiology, Primeasia University, Banani, Dhaka 1213, Bangladesh*
[c] *Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh*
[d] *Department of Biochemistry and Molecular Biology, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh*
[e] *Department of Diabetes and Endocrinology, Pabna Diabetic Association Hospital, Pabna 6600, Bangladesh*
[f] *Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh*
[g] *Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada*
[h] *Group of Biophotomatiχ, Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh*
[i] *Health Informatics Research Lab, Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City (DSC), Birulia, Savar, Dhaka-1216, Bangladesh*
[j] *Department of Computer Engineering, Umm Al-Qura University, Mecca 24381, Saudi Arabia*

ARTICLE INFO

Keywords:
SLC30A8
T2D
Missense
Non-synonymous
Polymorphism
And colorectal cancer

ABSTRACT

Background and aims: The single nucleotide polymorphisms (SNPs) in *SLC30A8* gene have been recognized as contributing to type 2 diabetes (T2D) susceptibility and colorectal cancer. This study aims to predict the structural stability, and functional impacts on variations in non-synonymous SNPs (nsSNPs) in the human *SLC30A8* gene using various computational techniques.
Materials and methods: Several *in silico* tools, including SIFT, Predict-SNP, SNPs&GO, MAPP, SNAP2, PhD-SNP, PANTHER, PolyPhen-1,PolyPhen-2, I-Mutant 2.0, and MUpro, have been used in our study.
Results: After data analysis, out of 336 missenses, the eight nsSNPs, namely R138Q, I141N, W136G, I349N, L303R, E140A, W306C, and L308Q, were discovered by ConSurf to be in highly

* Corresponding author. Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada.
** Corresponding author.
E-mail addresses: moinbge12@gmail.com (Md.M. Uddin), tanvirhossain69744@gmail.com (Md.T. Hossain), arju.primer60@gmail.com (Md.A. Hossain), asifahsanbge@gmail.com (A. Ahsan), khs.shamim1@gmail.com (K.H. Shamim), arifh6528@gmail.com (Md.A. Hossen), dr.srahman90@yahoo.com (Md.S. Rahman), habib@iu.ac.bd (M.H. Rahman), kawsar.ict@mbstu.ac.bd, k.ahmed@usask.ca (K. Ahmed), francis.bui@usask.ca (F.M. Bui), fayzahrani@uqu.edu.sa (F.A. Al-Zahrani).
[1] Joint First Author.

https://doi.org/10.1016/j.heliyon.2024.e37280
Received 23 May 2024; Received in revised form 29 August 2024; Accepted 30 August 2024
Available online 31 August 2024

conserved regions, which could affect the stability of their proteins. Project HOPE determines any significant molecular effects on the structure and function of eight mutated proteins and the three-dimensional (3D) structures of these proteins. The two pharmacologically significant compounds, Luzonoid B and Roseoside demonstrate strong binding affinity to the mutant proteins, and they are more efficient in inhibiting them than the typical *SLC30A8* protein using Autodock Vina and Chimera. Increased binding affinity to mutant *SLC30A8* proteins has been determined not to influence drug resistance. Ultimately, the Kaplan-Meier plotter study revealed that alterations in *SLC30A8* gene expression notably affect the survival rates of patients with various cancer types.

Conclusion: Finally, the study found eight highly deleterious missense nsSNPs in the *SLC30A8* gene that can be helpful for further proteomic and genomic studies for T2D and colorectal cancer diagnosis. These findings also pave the way for personalized treatments using biomarkers and more effective healthcare strategies.

1. Introduction

Diabetes is a metabolic disease affecting populations globally, arising from insulin production or utilization disruptions. Notably, type 2 diabetes (T2D) serves as a well-established risk factor for colorectal cancer, a highly prevalent form of malignancy worldwide. Based on the most recent cancer data, colorectal cancer is a significant contributor to cancer-related deaths, posing a substantial threat to global health, with 1.9 million new cases recorded in 2020 [1]. A gene associated with insulin signaling, specifically *SLC30A8*, could potentially influence the connection between diabetes and the risk of colorectal cancer. This connection may offer new perspectives on the biological mechanisms that link diabetes and colorectal cancer [2]. Several cancers closely relate to the solute carrier 30 (SLC30) family of genes during their development and progression [3]. The presence of a specific variant in the *SLC30A8* gene was linked substantially to an increased risk of breast cancer in women [4]. The combined effects of uncontrolled blood sugar, hormone imbalances, and ongoing inflammation associated with T2D may linked to an increased risk of head and neck cancer development, metastasis, and severity [5–7]. Additionally, T2D is more prone to infection with Kaposi's sarcoma-associated herpesvirus (KSHV), the pathogen accountable for Kaposi's sarcoma [8].

SLC30A8, a gene responsible for encoding the zinc transporter 8 (ZnT8), is an influential variable in T2D risk [9]. ZnT8, which belongs to the ZnT family, is mostly found in pancreatic beta cells, and it is essential for transporting zinc from the cytoplasm to intracellular vesicles [10,11]. Residing inside insulin-storing granules, ZnT8 is essential for the synthesis, encapsulation, and discharge of insulin. Antibodies against ZnT8 are a potential breakthrough in diagnosing autoimmune diabetes [9].

Single nucleotide polymorphisms (SNPs) represent the predominant genetic variation within the DNA sequence. Their density varies across regions, yet they occur in approximately 100–300 bases of the human genome, constituting over 90 % of all genetic

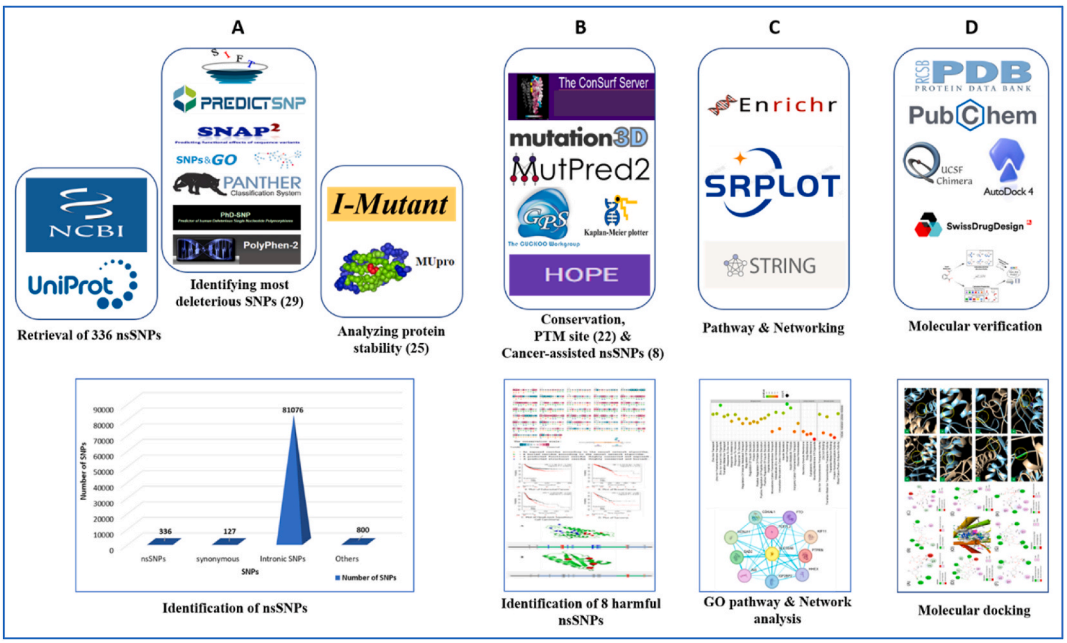

Fig. 1. Flow diagram of overall *in silico* SNPs analysis on the *SLC30A8* gene. A, B, C, and D periodically indicate the steps and their subsequent results.

variation in humans [12]. SNPs are present in the human genome's coding and non-coding areas, potentially influencing transcription-factor binding, splicing, or gene expression. Some SNPs do not impact cell function, while others can increase susceptibility to illness or alter medication response [13,14]. Approximately 500,000 SNPs have been identified within the coding region of the human genome [15]. Among these, non-synonymous SNPs (nsSNPs), hold significance compared to other proteins as they can modify amino acid residues [16]. Although not all structural and functional modifications induced by nsSNPs have the potential to be harmful. While some nsSNPs have functional effects, others impact structural features [17]. Modifications in function can have a favorable or detrimental impact on the structure or functionality of proteins. Potentially harmful effects include changes in gene regulation [18,19], instability of protein structures, effects on protein charge, geometry, and hydrophobicity [20,21] and changes in protein structure [22]. Therefore, it can be assumed that nsSNPs are associated with several human illnesses. The significant impact of nsSNPs on human health and disease has made this research more effective in recent years.

Researchers have found hundreds of SNPs in the *SLC30A8* gene's coding region because of advancements in genotyping methods. Nevertheless, more work needs to be done to determine which SNPs of the *SLC30A8* gene are the most harmful and disease-prone. Several studies conducted according to these techniques, using NCBI dbSNP, UniProtKB, SIFT, Predict-SNP, SNAP2, PANTHER, SNPs&GO, PolyPhen, PhD-SNP, I-Mutant, ConSurf, GPS-MSP, NetPhos, Mutpred2, MUpro, Mutation 3D, Project Hope, EnrichR, STRING, Chimera, AutoDock Vina, SwissADME, pkCSM, Kaplan-Meier Plotter, etc., have been employed to forecast the functional consequences of the most harmful nsSNPs in *SLC30A8* gene (Fig. 1(A-D)) [23,24] Additionally, we suggested novel treatment approaches that provide insightful information regarding their potential impact on the protein's functionality, stability, and structure. Our study also identified potentials biomarkers as drug targets for mutations linked to T2D and its progression to CRC. Therefore, it can be predicted how patients will respond to treatments, enabling more effective and developing personalized therapeutic strategies.

2. Materials and Methods

2.1. Retrieval of SLC30A8 nsSNPs dataset

The dataset of the human *SLC30A8* gene polymorphisms and protein sequence (Uniprot ID: Q8IWU4) was collected from NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp/) and UniProtKB (https://www.uniprot.org/) databases, respectively. The dataset comprised 82339 SNPs, including 336 nonsynonymous SNPs (nsSNPs) in the coding area and 945 in non-coding regions. Nonsynonymous mutations are those that result in a change in the amino acid sequence of a protein and can have a significant impact on the structure and function of the protein. In our dataset preparation, we excluded only missense variant of nsSNPs because, its impact can vary from benign to highly detrimental, depending on the role of the substituted amino acid in the protein's function or structure compared to nonsense variant [25]. As a result, coding region of 336 nsSNPs were prioritized for further investigation due to their recognized ability to significantly influence protein structure and function, potentially responding to the progression of several disease states. Subsequently, all the 336 nsSNPs were isolated and retrieved for further investigation.

2.2. Evaluation of the deleterious nsSNPs

In our study, we excluded deleterious nsSNPs that increase the stability of the protein rather than over-stabilization effects. Because the role of *SLC30A8* gene in insulin regulation and pancreatic beta-cell function, under-stabilization may be more immediately harmful. Over-stabilization, while potentially harmful in the long term, may not have as immediate an impact. Under-stabilization of the protein can quickly disrupt zinc homeostasis and insulin secretion, leading to metabolic disturbances and contributing to diabetes risk [26].

Besides, we utilized various bioinformatics tools, such as SIFT, Predict-SNP, MAPP, PANTHER, SNAP2, SNPs&GO, PhD-SNP, PolyPhen-1, and PolyPhen-2, to assess the functional consequences of the nsSNPs. The use of multiple methods ensured a rigorous and accurate screening process. Only modifications anticipated to exert harmful effects were chosen for additional scrutiny. SIFT (Sorting Intolerant From Tolerant) (https://sift.bii.a-star.edu.sg/) can forecast the functional outcomes of amino acid replacements by examining sequence homology and the physicochemical attributes of amino acids [27]. The PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html) is the Predictor of Human Negative Single Nucleotide Polymorphisms. PhD-SNP uses support vector machines (SVMs) to identify a point mutation in a human that is associated with a genetic disorder or polymorphism [28]. MAPP, SNAP 2, PANTHER, and PolyPhen-1 tools were combined into a consensus classifier Predict-SNP (https://loschmidt.chemi.muni.cz/predictsnp1/), offering much better prediction performance while also providing data for each mutation, showing that consensus prediction is a reliable and accurate substitute for the predictions made by different tools [29]. SNPs&GO (https://snps-and-go.biocomp.unibo.it/snps-and-go/) approach predicts harmful single amino acid polymorphisms (SAPs). For each protein variant, the server's output gives the likelihood that it will be linked to a disease in humans. When the likelihood score is > 0.5, it suggests that the variations are disease-associated mutations, and when it is 0.5, it shows that the variations have no effect [30]. The software PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) predicts the impact of amino acid alterations on the stability and functionality of human proteins, considering structural and evolutionary comparisons. Coding SNPs are mapped to gene transcripts, structural and functional information on proteins is collected, functional annotation of SNPs is completed, and conservation profiles are established. Subsequently, it computes the possibility of the missense mutation being harmful by integrating all these characteristics [31].

2.3. Analyzing protein stability

The I-Mutant 2.0 tool is built on a support vector machine (SVM) (https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi). I-Mutant 2.0 is reasonably accurate, it is not infallible and can sometimes give erroneous predictions, especially for less common mutations or proteins not well represented in the training data. We employed the sequence-based edition of I-Mutant 2.0, which categorizes predictions into three distinct groups: substantial decrease (0.5 kcal/mol), notable increase (>0.5 kcal/mol), and neutral mutation (between −0.5 and 0.5 kcal/mol). The basis for the predicted free energy change (ΔΔG) by I-Mutant 2.0 is the variance between the unfolding Gibbs free energy change of the mutant and the native protein (in kcal/mol) [32,33].

Additionally, we utilized MUpro (http://mupro.proteomics.ics.uci.edu/), which has two machine-learning applications: neural networks and support vector machines. We used the program's sequence-based version and the default settings to run the SVM algorithm. The program's output indicates the energy change's sign (+ or -) [34]. With these methods, we could anticipate how nsSNPs could affect the stability of proteins and guide our subsequent investigation.

2.4. Evolutionary conservation analysis of the most damaging nsSNPs

A trade-off between the need to preserve the structural stability and functionalities of the macromolecule and the natural tendency of the molecular building blocks to mutate is indicated by the level of evolutionary conservation of nucleic acid in DNA/RNA or amino acid in proteins [35]. Compared to other popular approaches based on relative entropy and consensus techniques, ConSurf (https://consurf.tau.ac.il/) has a clear advantage. To estimate evolutionary rates, it considers the distinct dynamics of the sequences under study and the phylogenetic connections among the homologs using complex probabilistic evolutionary models [36]. The ConSurf server was utilized to evaluate the level of evolutionary conservation at each amino acid position in the *SLC30A8* protein and identify the high-risk nsSNPs [23].

2.5. Prediction of post-translational modification (PTM) sites

Proteins are fine-tuned by over 400 complex chemical modifications. These act as molecular control panels, meticulously regulating diverse cellular functions by altering protein structure and behaviour [37]. We relied on NetPhos 3.1 (https://services.healthtech.dtu.dk/services/NetPhos-3.1/) [38] to pinpoint potential areas where phosphate molecules could latch onto our protein. This software uses several neural networks to anticipate possible phosphorylation sites on tyrosine, threonine, and serine residues. A score exceeding 0.5 on the software's scale indicated a high phosphorylation probability at that location. Additionally, we employed GPS-MSP 1.0 (http://msp.biocuckoo.org/) [39] to predict possible methylation sites on the protein chain.

2.6. Identification of structural and functional alternations

We used Mutpred2 (http://mutpred.mutdb.org/), which integrates genetic and molecular data using machine learning methods and employs the amino acid sequence to predict structural and functional alteration caused by amino acid changes. Additionally, the instrument provides information on the precise biological mechanisms behind the onset of diseases [23]. We provided amino acid substitution and protein FASTA sequences as input for the Mutpred2 study. The g-score, derived from the collective scores of all neural networks within MutPred2, indicates pathogenicity. A g-score surpassing 0.50 ($g > 0.50$) is considered indicative of pathogenicity [40].

2.7. Evaluation of cancer-associated nsSNPs

Mutation 3D (http://www.mutation3d.org/) is employed to detect a group of amino acid substitutions arising from somatic cancer mutations. It proves invaluable in investigating the spatial arrangement of amino acid alterations within protein models and structures. This software utilizes a 3D clustering method to detect possible cancer-causing changes in amino acids inside a protein. It requires a target protein and its mutations as input [41].

2.8. Estimating the effects of high-risk nsSNPs on the structure of protein

An automated approach for analyzing mutants called Project HOPE (https://www3.cmbi.umcn.nl/hope/input/) helps examine the structural, functional and metabolic consequences of point mutations within protein sequences. It investigates how changes in amino acid composition affect native structures and the variations in hydrophobicity, charge, and size between residues of the wild-type and mutant-types. Project HOPE combines structural data with knowledge from literature and databases to predict the functional impact of mutations. It often compares the mutation in question to similar known mutations. The accuracy can vary depending on the extent and relevance of available comparative data [42]. Several computational studies utilizing the Project HOPE server to identify significance mutations in specific proteins including human LYZ C gene [43], CDK4 gene [44], and human RASSF5 gene [24]. Protein Data Bank (https://www.rcsb.org/) provides the *SLC30A8* protein sequence to Project HOPE for predicting the 3D structure of the mutated protein. It also provides a plethora of information on changing amino acids and estimates the composition and function of proteins after mutations [23,44].

2.9. Gene Ontology (GO) enrichment analysis

Genome-wide experiments generate massive datasets of genes, and enrichment analysis offers a powerful tool for extracting meaningful insights from these sets. Gene ontology analysis was performed using Enrichr (https://maayanlab.cloud/enrichr-kg) [45] to detect statistically noteworthy ($P < 0.05$) associations between the input gene set and curated databases of cellular components, molecular activities, and biological processes [46]. Employing SRplot (https://www.bioinformatics.com.cn/en), we constructed graphical summaries of the enriched analysis results [47].

2.10. Prediction of protein-protein interaction (PPI)

The online database STRING (Search Tool for Recurring Instances of Neighboring Genes) predicts the PPI network of the *SLC30A8* protein (https://string-db.org/). It retrieves the genes that are tangentially (via other genes) linked to the query gene using an evolutionary method [48]. The STRING output furnishes details regarding the protein's expression, localization, transcription, experimental evidence, and graphical depictions of the interactions in which the queried protein participates [49].

2.11. Superimposition and molecular docking of raw and mutated proteins

Protein structure determination and modeling involve determining and positioning protein structures in 3D space, known as protein structure superimposition. The best superimposition reflects dynamic changes over time and aligning structures to reveal common structural or functional patterns [50]. We used Chimera 1.16 [51] for the superimpositioning of native *SLC30A8* protein and its mutant variants.

Further, to assess the impact of deleterious point mutations on the binding affinity of *SLC30A8* (UniProt accession no. Q8IWU4.2), we employed the Autodock Vina (version 1.2.1) [52] to perform molecular docking analysis. We generated a mutated form of the target protein using the crystal structure complex of *SLC30A8* obtained from RCSB (https://www.rcsb.org/) [53] using phyr2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [54], and energy minimization employed by Swiss-PdbViewer (https://spdbv.unil.ch/) [55].

The mutated *SLC30A8* protein containing eight nsSNPs is the docking procedure's target protein along with normal protein. Ligands like Luzonoid B and Roseoside [56] obtain from pubchem (https://pubchem.ncbi.nlm.nih.gov/) [57]. The PDB format of these input ligands was transformed into pdbqt format using Autodock Vina. We tailored the grid box size for maximum efficiency. The docking outcome and the binding interactions between the ligand and receptor proteins were observed through BIOVIA Discovery Studio (version 21.1.0) [58].

2.12. ADME properties and toxicity analysis

Compounds such as Luzonoid B and Roseoside [56] exhibit notable physicochemical properties, pharmacokinetics, and drug-likeness while demonstrating no significant adverse effects according to the ADMET assay. These compounds could be assessed as potential therapeutic agents targeting *SLC30A8* in cancer and T2D [2]. Each compound was characterized *in silico* as a potential drug candidate to evaluate its pharmacokinetic properties, particularly absorption, distribution, metabolism, and excretion (ADME) [59]. Following the docking process, we assessed our chosen compounds' Pharmacokinetics and ADME characteristics using SwissADME

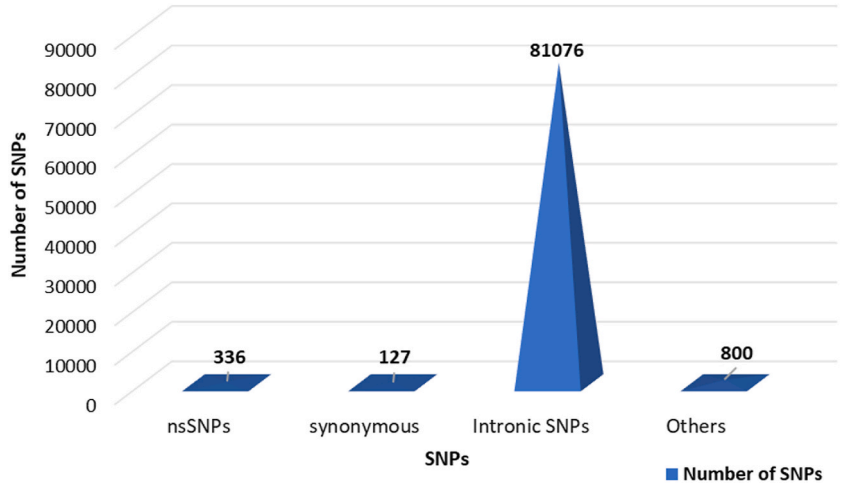

Fig. 2. The clustered pyramid illustrates the quantity and arrangement of SNPs within the human *SLC30A8* gene derived from the dbSNP database (nsSNPs: 336; synonymous SNPs: 127; intronic SNPs: 81076; others: 800).

(https://www.swissadme.ch/) [59] and the pkCSM (https://biosig.lab.uq.edu.au/pkcsm/) [60] online server.

2.13. Gene diseases Co-relation through survival analysis

The Kaplan-Meier technique (https://kmplot.com/analysis/) is a particularly effective way to evaluate changes associated with treatment interventions and forecast how they may affect cancer survival rates. This strategy makes use of overall survival (OS) and relapse-free data from the Cancer Genome Atlas (TCGA), European Genome Phenome Atlas (EGA), and Gene Expression Omnibus (GEO). It makes it possible to generate and assess cancer biomarkers through thorough meta-analyses. Once the survival job has been submitted, a Kaplan-Meier plot will demonstrate whether there is a substantial difference in the length of survival between the two patient groups [61].

3. Results

3.1. Retrieving missense variant of nsSNPs of SLC30A8 gene

The NCBI dbSNP database was employed to get polymorphism data for the *SLC30A8* gene. There were 82,339 SNPs, including 127 synonymous variants, 81,076 intron region variants, 336 missense variants of nsSNPs, and the remaining were different types (Fig. 2 and Supplementary Table S1). We excluded the other SNPs and only concentrated on 336 non-synonymous SNPs (missense variations) for our investigation. A missense variant is a single nucleotide change in the coding region for a different amino acid substitution, potentially affecting the protein's structure and function.

3.2. Prediction and analysis of deleterious nsSNPs

We used nine different harmful SNP prediction algorithms (SIFT, Predict-SNP, MAPP, SNAP 2, SNPs&GO, PANTHER, PhD-SNP, PolyPhen-1, and PolyPhen-2) to identify deleterious nsSNPs capable of modifying the overall arrangement or function of the *SLC30A8* protein. For analysis, a dataset of 336 polymorphic inputs was employed in total. All nine *in silico* tools predicted that 29 out of 336 nsSNPs were harmful (Table 1).

Table 1
A compilation of nsSNPs predicted to be highly deleterious by nine computational methods.

Rs no	A.A. change	SIFT	MAPP	PANTHER	SNP&GO	PHD-SNP	Predict-SNP	SNAP2	PolyPhen1	PolyPhen2
Rs73317647	R165C	DL	DL	DL	D	D	DL	E	DL	PD
Rs139489847	G296R	DL	DL	DL	D	D	DL	E	DL	PD
Rs140404252	L74R	DL	DL	DL	D	D	DL	E	DL	PD
Rs141730422	H54R	DL	DL	DL	D	D	DL	E	DL	PD
Rs142407509	W152R	DL	DL	DL	D	D	DL	E	DL	PD
Rs145677283	R165H R165L	DL	DL	DL	D	D	DL	E	DL	PD
Rs149524118	A216T	DL	DL	DL	D	D	DL	E	DL	PD
Rs201697165	S182F	DL	DL	DL	D	D	DL	E	DL	PD
Rs746249658	R138L R138Q	DL	DL	DL	D	D	DL	E	DL	PD
Rs751350884	S182P	DL	DL	DL	D	D	DL	E	DL	PD
Rs762384069	H301N	DL	DL	DL	D	D	DL	E	DL	PD
Rs763719329	H304P H304L	DL	DL	DL	D	D	DL	E	DL	PD
Rs765852756	I141N	DL	DL	DL	D	D	DL	E	DL	PD
Rs771191320	W136G	DL	DL	DL	D	D	DL	E	DL	PD
Rs771330275	V219E	DL	DL	DL	D	D	DL	E	DL	PD
Rs776903420	S230R	DL	DL	DL	D	D	DL	E	DL	PD
Rs780551100	I349N	DL	DL	DL	D	D	DL	E	DL	PD
Rs866700996	L303R	DL	DL	DL	D	D	DL	E	DL	PD
Rs1035358113	R215G	DL	DL	DL	D	D	DL	E	DL	PD
Rs1255515664	W306C	DL	DL	DL	D	D	DL	E	DL	PD
Rs1264823419	H137P	DL	DL	DL	D	D	DL	E	DL	PD
Rs1271726667	L111R	DL	DL	DL	D	D	DL	E	DL	PD
Rs1300885615	E140A	DL	DL	DL	D	D	DL	E	DL	PD
Rs1305371427	S234R	DL	DL	DL	D	D	DL	E	DL	PD
Rs1325443204	L308Q	DL	DL	DL	D	D	DL	E	DL	PD
Rs1586608274	Q227R	DL	DL	DL	D	D	DL	E	DL	PD
Rs1822266670	D103V	DL	DL	DL	D	D	DL	E	DL	PD
Rs1822805541	L222R	DL	DL	DL	D	D	DL	E	DL	PD
Rs2130999340	L146R	DL	DL	DL	D	D	DL	E	DL	PD

DL: deleterious; D: disease; E: effect; PD: probably damaging.

3.3. Analyzing the protein stability of 29 identified mutant variants

To evaluate protein stability, 29 nsSNPs previously identified as harmful were further analyzed using two additional tools: I-Mutant 2.0 and MUpro. I-Mutant showed that 25 nsSNPs decreased protein stability, while four nsSNPs were found to increase the stability, using the reliability index (RI) value and changes in free energy (ΔΔG). MUpro also predicted that 28 nsSNPs destabilized the protein, while one (R138L) nsSNP stabilized it. For further analysis, we only focused on missense variants of nsSNPs predicted to destabilize by all *in silico* tools (Table 2).

3.4. Evolutionary conservation analysis of high-risk nsSNPs

It is crucial for a protein's function and biology that its amino acids remain conserved throughout evolution. Mutations in these conserved regions often have detrimental effects and can disrupt the protein's normal function. It has been discovered that among 25 nsSNPs, 22 are located within highly conserved regions, with nine of them being buried and eight exposed (Fig. 3). The remaining of five nsSNPs were classified as buried or exposed based on the network algorithm employed by the Consurf online database. Three SNPs were excluded from this analysis as they fall within the variable region (Supplementary Table S2).

3.5. Prediction of post-translation modification (PTM) site

Using the GPS-MSP 1.0 tool, we identified R325 as a potential methylation site on the *SLC30A8* protein. We turned to NetPhos 3.1 to predict phosphorylation sites (Fig. 4). It identified S307 in the native protein and S299 in the mutant as potential phosphorylation sites (Supplementary Table S3).

3.6. Evaluation of cancer-associated nsSNPs

Mutation 3D server classified our reported nsSNPs into two groups. One group, comprising SNPs W152R, R215G, A216T, V219E, L222R, S230R, and S234R, is designated as covered mutations (represented in blue) in Fig. 5. The "covered mutation group" in the context of the Mutation 3D server typically refers to a set of mutations that are analyzed for their potential impact on protein structure and function. Mutation 3D server provided several default exclusion criteria for the selection of covered mutation group such as (i)

Table 2
Changes in the stability of the *SLC30A8* protein structure predicted by I-mutant 2.0 and MUpro.

Rs no	Mutation	I-Mutant 2.0	MUpro
Rs73317647	R165C	D	D
Rs139489847	G296R	D	D
Rs140404252	L74R	D	D
Rs141730422	H54R	D	D
Rs142407509	W152R	D	D
Rs145677283	R165H	D	D
	R165L	D	D
Rs149524118	A216T	D	D
Rs201697165	S182F	**I**	D
Rs746249658	R138L	D	**I**
	R138Q	D	D
Rs751350884	S182P	**I**	D
Rs762384069	H301N	D	D
Rs763719329	H304P	**I**	D
	H304L	**I**	D
Rs765852756	I141N	D	D
Rs771191320	W136G	D	D
Rs771330275	V219E	D	D
Rs776903420	S230R	D	D
Rs780551100	I349N	D	D
Rs866700996	L303R	D	D
Rs1035358113	R215G	D	D
Rs1255515664	W306C	D	D
Rs1264823419	H137P	**I**	D
Rs1271726667	L111R	D	D
Rs1300885615	E140A	D	D
Rs1305371427	S234R	D	D
Rs1325443204	L308Q	D	D
Rs1586608274	Q227R	D	D
Rs1822266670	D103V	D	D
Rs1822805541	L222R	D	D
Rs2130999340	L146R	D	D

Five SNPs including one isoform (H304P) of rs746249658 have been excluded for further analysis (Bold).

Fig. 3. The ConSurf server conducted an analysis of evolutionary conservation for *SLC30A8*. Here, 'f' represents a predicted functional residue (highly conserved and exposed), while 's' denotes a predicted structural residue (highly conserved and buried).

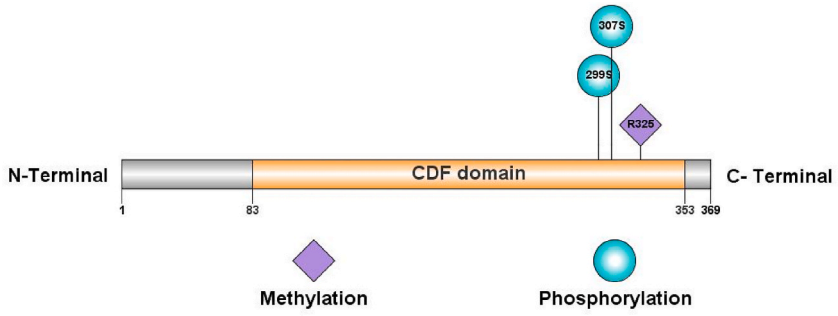

Fig. 4. PTM sites (methylation and phosphorylation) predicted by GPS-MSP 1.0, and NetPhos 3.1 (using IBS software).

mutations are typically annotated with their exact position in the protein sequence and structure, (ii) the specific substitution of amino acid such as missense at the given position, (iii) mutations in highly conserved regions indicating potential functional importance, (iv) finally, mutations that are predicted to affect the stability or folding of the protein structure.

Another group is the clustered mutations are categorized consisting of the eight nsSNPs (shown in red) and divided into two clusters. According to this research, our identified eight nsSNPs including W136G, R138Q, E140A, I141N, L303R, W306C, L308Q, and

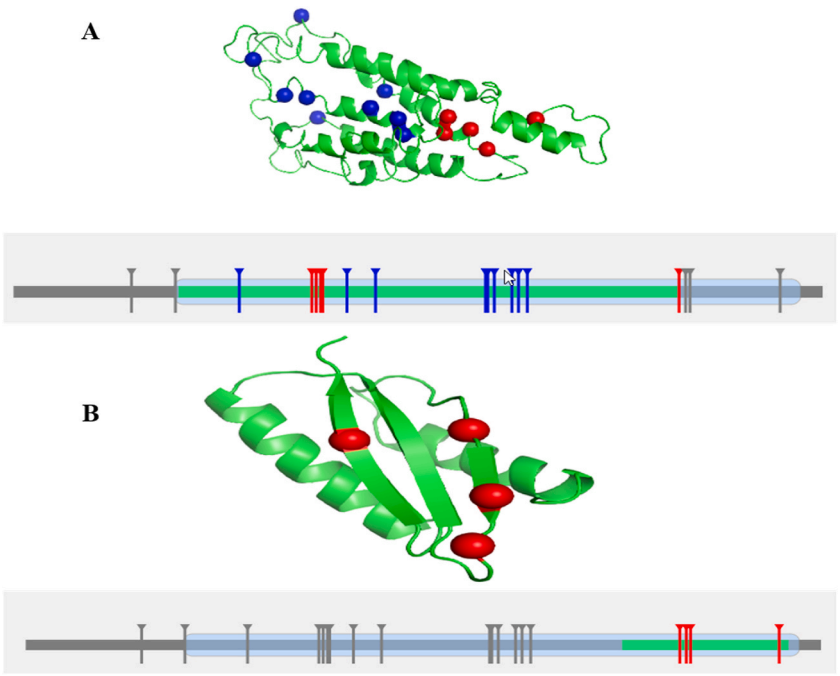

Fig. 5. The mutation 3D server identified certain nsSNPs as potential cancer-causing mutations (red mark). Red indicates clustered mutations, whereas blue signifies covered mutations. A represents cluster-1, where five nsSNPs (W136G, R138Q, E140A, I141N, and L303R) are present, and B represents cluster-2 where four nsSNPs are (L303R, W306C, L308Q, and I349N) present related with cancer (red mark).

Table 3
MutPred2 predicts structural and functional modifications of *SLC30A8* protein. The term "g-value" refers to the MutPred2 general score.

Mutation	Actionable/confident hypothesis	g-value	p-value	Probability	Affected PROSITE and ELM Motifs
R138Q	Altered Transmembrane protein	0.690	0.03	0.28	None
	Altered Ordered interface		0.03	0.25	
I141N	Altered Transmembrane protein	0.768	4.9e-03	0.34	ELME000045
	Altered Ordered interface		0.04	0.28	
	Gain of Relative solvent accessibility		0.02	0.28	
	Gain of Allosteric site at R138		0.01	0.24	
W136G	Gain of Intrinsic disorder	0.899	4.8e-03	0.44	ELME000008, ELME000062, PS00008
	Altered Ordered interface		2.8e-03	0.34	
	Altered Transmembrane protein		0.01	0.31	
	Gain of Relative solvent accessibility		0.01	0.29	
	Gain of Catalytic site at R138		0.05	0.08	
I349N	Gain of Intrinsic disorder	0.810	6.5e-03	0.41	ELME000173
	Altered Stability		5.5e-03	0.30	
L303R	Altered Metal binding	0.888	4.5e-03	0.29	ELME000051, ELME000063
	Altered Ordered interface		0.04	0.24	
	Loss of Catalytic site at H304		7.7e-03	0.23	
	Gain of Allosteric site at W306		0.03	0.21	
	Altered Stability		0.04	0.11	
E140A	Altered Transmembrane protein	0.823	3.8e-04	0.28	None
	Altered Ordered interface		0.03	0.25	
	Loss of Relative solvent accessibility		0.04	0.24	
W306C	Altered Ordered interface	0.839	3.3e-03	0.32	None
	Altered Metal binding		4.5e-03	0.30	
	Loss of Allosteric site at W306		0.02	o.25	
	Loss of Catalytic site at H304		0.01	0.21	
L308Q	Altered Metal binding	0.707	8.3e-03	0.26	ELME000202
	Altered Ordered interface		0.04	0.24	
	Loss of Catalytic site at H304		9.0e-03	0.22	
	Gain of Allosteric site at W306		0.03	0.21	
	Altered Stability		0.04	0.11	

I349N are linked to cancer and being examined further (Fig. 5(A-B)).

3.7. Identification of functional and structural modifications

Eight nsSNPs were identified as potentially harmful and further analyzed using the MutPred2 tool. The results, including g-scores and p-values, are summarized in Table 3. Notably, all eight nsSNPs (W136G, R138Q, E140A, I141N, L303R, W306C, L308Q, and I349N) exhibited strong pathogenic potential, with g-scores exceeding 0.60 and p-values below 0.05.

3.8. Estimating the effects of high risk nsSNPs on the structure of protein

Project HOPE utilizes computational methods to analyze the structural alterations induced by amino acid changed in native protein. The study revealed that wild-type and mutant-type amino acids exhibit distinct physicochemical attributes, including size, charge, and hydrophobicity. All eight nsSNPs resulted in changes in amino acid size. Four mutant amino acids (I141N, L303R, L308Q, I349N) exhibited increased size compared to their wild-type counterparts, while the remaining four mutations displayed reduced size. Three nsSNPs (R138Q, E140A, L303R) were found to alter the amino acid charge. Additionally, E140A exhibited increased hydrophobicity in the mutant form. R138Q did not affect hydrophobicity, while all other six mutants showed reduced hydrophobicity compared to the wild type (Table 4).

Project HOPE identified notable distinctions between wild-type and mutant-type amino acids regarding structural alterations, mutations within conserved domains, and properties of amino acids (Supplementary Table S4). The study employed computational methods to predict the physicochemical attributes, including structural features and conserved domains, which were then used to assess the possible molecular impact of high-risk nsSNPs on protein structure (Supplementary Table S5). Furthermore, Project HOPE generated 3D model structures for the eight mutated *SLC30A8* proteins. Fig. 6 illustrates the ribbon presentation and resulting 3D models following the introduction of mutations. These pictures show how mutations in the *SLC30A8* protein change its structure, which helps us understand valuable insights.

3.9. Gene Ontology (GO) analysis

We assessed the biological properties of *SLC30A8* by functionally annotating the primary targets on GO enrichment. The three components of the GO enrichment analysis are biological process (BP), cellular component (CC), and molecular function (MF), depicted by green, orange, and blue bubbles, respectively. About 34 GO words in all were produced, of which BP provided 24, CC provided 4, and MF provided 6. Node sizes indicated linked target genes, and green to red colors represented high to low p-values (Fig. 7).

3.10. Prediction of protein-protein interaction (PPI)

Changes in a protein's structure caused by mutations can affect its function. An analysis of the *SLC30A8* protein's interaction network using the STRING server revealed interactions with PTPRN, HHEX, CDKAL1, INS, TCF7L2, GAD2, IGF2BP2, KCNJ11, FTO, and KIF11 (Fig. 8 and Supplementary Table S6). The *SLC30A8* protein exhibits high connectivity within this network, interacting with ten other proteins to form a highly interconnected network with 39 edges and 11 nodes. The mean node degree stands at 7.09, while the PPI enrichment value is 6.27e-10. This network suggests these proteins have strong functional associations and potential regulatory relationships.

3.11. Superimposition and molecular docking of raw and mutated proteins

A molecular docking simulation predicted the binding affinities of eight mutated *SLC30A8* for two target compounds. After matchmaking (superimposed) through chimera tools, the *SLC30A8* protein and eight mutant proteins are structurally similar (Fig. 9(A-H)). There is no significant difference between the wild-type and mutant-type *SLC30A8* protein structures after superposition. In the

Table 4
Properties of wild-type and mutant-type amino acids regarding size, charge, and hydrophobicity analyzed by Project HOPE.

AA change	Wild type AA			Mutant type AA		
	Size	Charge	Hydrophobicity	Size	Charge	Hydrophobicity
R138Q	Larger	Positive	None	Smaller	Neutral	None
I141N	Smaller	_	More hydrophobic	Larger	_	Less hydrophobic
W136G	Larger	_	More hydrophobic	Smaller	_	Less hydrophobic
I349N	Smaller	_	More hydrophobic	Larger	_	Less hydrophobic
L303R	Smaller	Neutral	More hydrophobic	Larger	Positive	Less hydrophobic
E140A	Larger	Negative	Less hydrophobic	Smaller	Neutral	More hydrophobic
W306C	Larger	_	More hydrophobic	Smaller	_	Less hydrophobic
L308Q	Smaller	_	More hydrophobic	Larger	_	Less hydrophobic

'_' indicates neutral effect.

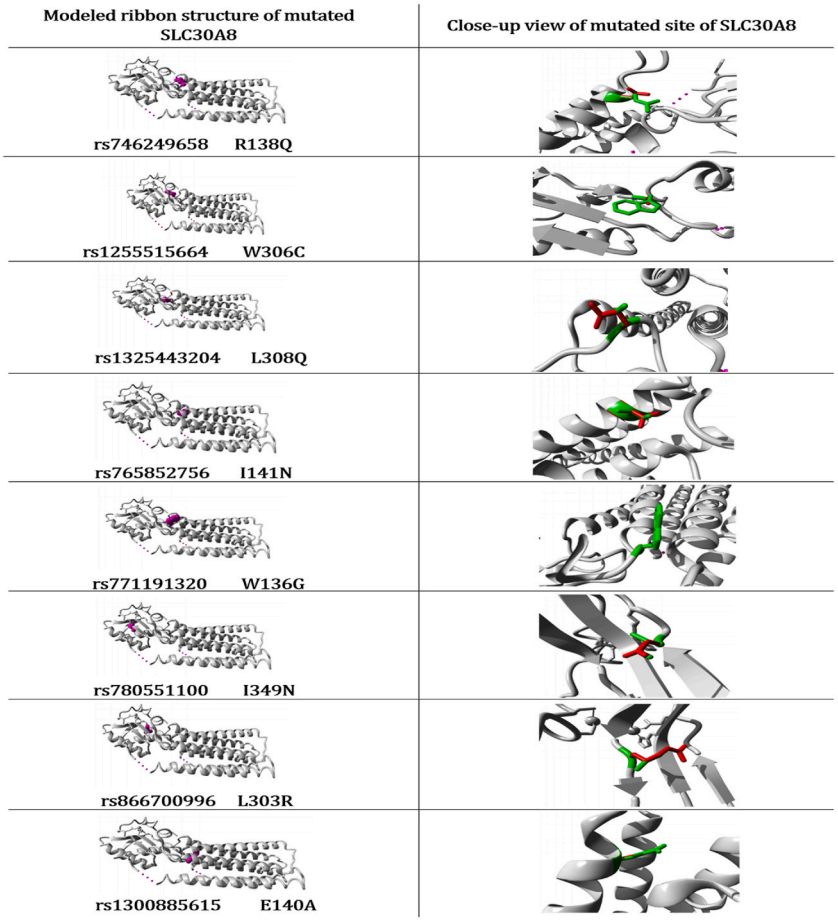

Fig. 6. The 3D model structure of the *SLC30A8* protein was predicted using Project HOPE. The location of the mutation is depicted in violet on the ribbon diagram. The native and mutated amino acids are represented by green and red colors, respectively.

Fig. 9, we have only illustrated the substitution of amino acids. While validating our docking tools using RMSD analysis, we identified eight complexes with co-crystallized unique ligands to those complexes: luzonoid B and Roseoside, along with normal protein. The 6xpd exhibited a high binding affinity for luzonoid B (−6.9 kcal/mol) and Roseoside (−6.5 kcal/mol). Eight mutated proteins also show promising affinities with luzonoid B and Roseoside, as shown in Table 5. Among their hydrogen and hydrophobic bonds, the residues like Glu_{66}, Tyr_{67}, Asn_{311}, Met_{344}, and His_{345} share between mutated proteins (Fig. 10(A-H) and Fig. 11(A-H)).

3.12. ADME properties and toxicity analysis

The tools SwissADME and pkCSM analyze the molecular properties of selected compounds to ensure compliance with Lipinski's rule of five, a crucial factor in drug design. The chosen compounds obtained from PubChem were assessed for their drug-likeness and physiochemical characteristics. To determine the most promising candidates' drug-likeness, we evaluate selected compounds' ADME properties. This evaluation included Lipinski's rule of five compliance, bioactivity rating, and ADME parameters such as Total Solvent Accessible Surface Area (TPSA), Predicted Aqueous Solubility (PSA), Human Oral Absorption (HIA), Molecular Weight (MW), Blood-Brain Barrier (BBB) permeability, Hydrogen Bond Donor (HBD) count, Hydrogen Bond Acceptor (HBA) count, Central Nervous System (CNS) permeability, and Partition Coefficient (LogP). These factors are crucial for determining a compound's potential to effectively interact with its target site and be safely metabolized and eliminated from the body (Table 6). The toxicity evaluation of the selected compounds revealed encouraging results, suggesting their potential as drug candidates. The AMES test, Hepatotoxicity, Oral Rat Acute Toxicity (LD50), and Skin Sensitization tests all demonstrated low toxicity and minimal adverse effects, supporting the compounds' safety and suitability for further development (Table 6). These ligands' favorable physiochemical properties and ADME profiles suggest their potential as promising drug candidates for future development.

3.13. Gene diseases Co-relation through survival analysis

A research investigation was carried out to examine the correlation between the *SLC30A8* gene and the survival rates of patients

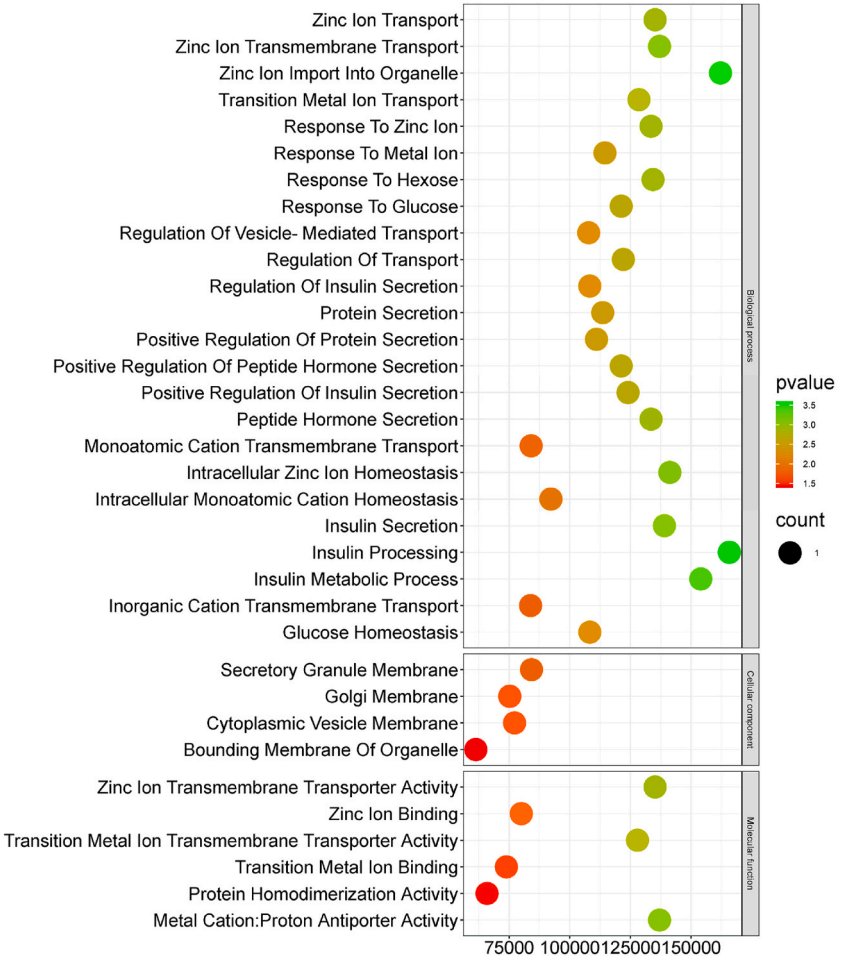

Fig. 7. Gene Ontology (GO) pathway study of *SLC30A8* gene using Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). The pathway term outcomes (based on P-value) were sorted using the combined score. The color gradient green to red color indicates the higher to lower p-value.

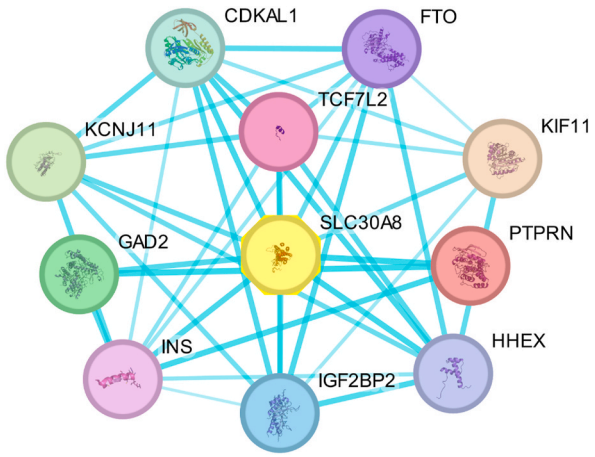

Fig. 8. PPI network of *SLC30A8* protein predicted by STRING server and Designed through Cytoscape software. The circular shape indicates the neighboring proteins and solid line indicates the interaction between those proteins.

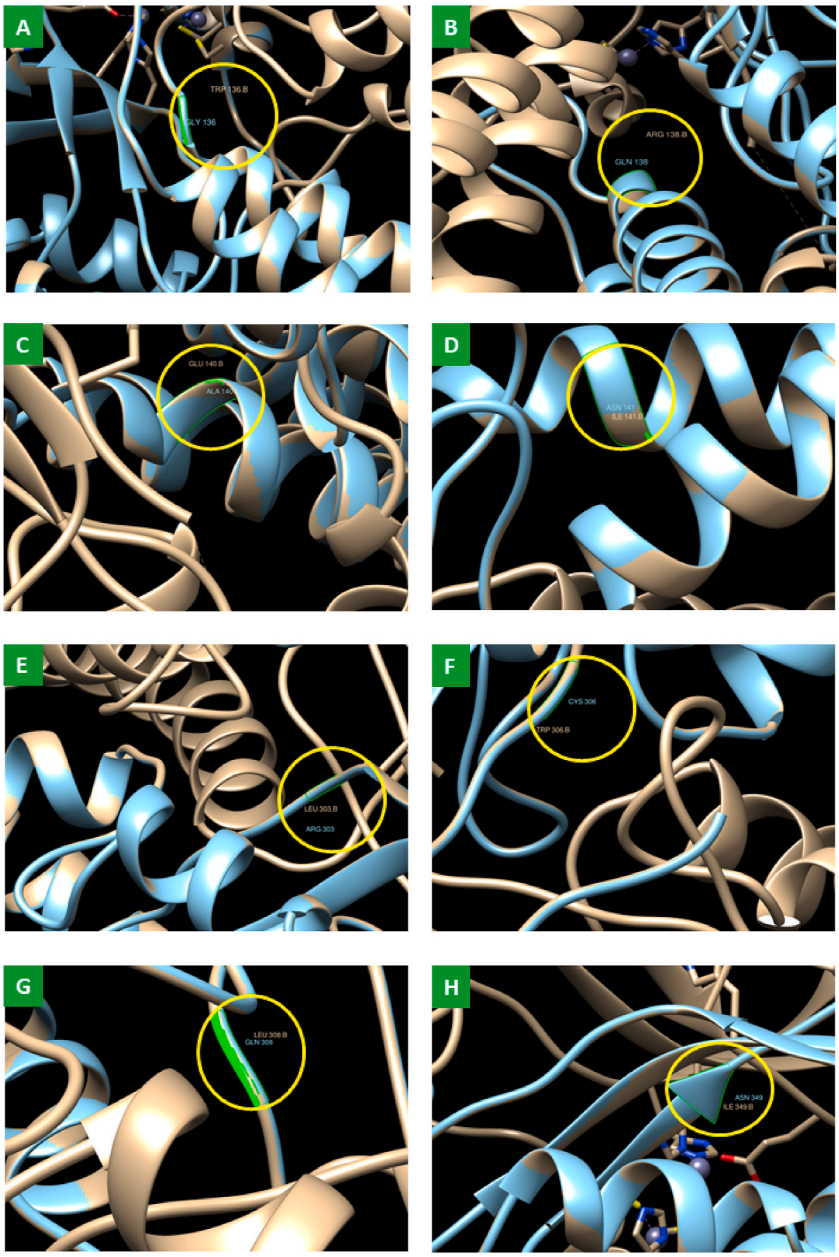

Fig. 9. Superimposition of Wild type of *SLC30A8* protein with eight mutated proteins. Superimposed structure of wild-type *SLC30A8* protein and mutant-type having mutations from (A) Tryptophan to Glycine at position 136, (B) Arginine to Glutamine at position 138, (C) Glutamic acid to Alanine at position 140, (D) Isoleucine to Asparagine at position 141, (E) Leucine to Arginine at position 303, (F) Tryptophan to Cysteine at position 306, (G) Leucine to Glutamine at position 308 and (H) Isoleucine to Asparagine at position 349. Ash color indicates wild-type protein, sky blue for mutant protein, and the yellow box for mutations between a wild-type protein and eight mutated proteins, along with their positions.

with breast cancer, colorectal cancer, head-neck squamous cell carcinoma, and sarcoma. The findings indicated that individuals exhibiting elevated levels of *SLC30A8* expression faced reduced mortality risk across all four types of cancer. For colorectal and breast cancer, the HR ratio and P-value were (HR 1.38 [1.1–1.74], P = 0.0057) and (HR 1.33 [0.92–1.91], P = 0.13), respectively. Similarly, for head-neck squamous cell carcinoma and sarcoma, the HR ratio and P-value stood at (HR 1.28 [0.96–1.71], P = 0.092) and (HR 1.53 [1.02–2.29], P = 0.037), respectively (Fig. 12(A-D)). These findings imply that the *SLC30A8* gene could be a prognostic indicator for these cancers. Furthermore, eight nsSNPs were detected within the *SLC30A8* gene. These nsSNPs are anticipated to impair the function of the *SLC30A8* protein, potentially fostering cancer progression.

Table 5
Interaction of ligands showing high docking scores for mutated *SLC30A8* protein.

Receptor type	Receptor	Binding Affinity of Luzonoid B	Binding Affinity of Roseoside	Interacting residues			
				Hydrogen bond		Hydrophobic bond	
				Luzonoid B	Roseoside	Luzonoid B	Roseoside
Normal *SLC30A8* protein	6xpd	−6.9	−6.5	His_{106}, Glu_{164}, Arg_{165}, Glu_{227}, Ser_{228}	Thr_{102}, Ser_{182}	Leu_{158}, Leu_{161}, Met_{178}, Val_{181}, Ala_{185}	Ala_{105}, His_{106}, Ile_{109}, Val_{154}, Leu_{158}, Leu_{161}
Mutated *SLC30A8* protein	W136G	−6.6	−6.6	Lys_{73}, Asn_{311}	Tyr_{67}, Asn_{311}	Glu_{66}, Tyr_{67}, Ala_{210}	Thr_{343}, Met_{344}, His_{345}
	R138Q	−7.9	−7.2	Glu_{66}, Lys_{73}, Glu_{209}, Asn_{211}, Arg_{215}	Gly_{56}, His_{137}, Asn_{311}, Met_{344}, His_{345}	Tyr_{67}, His_{137}, Ala_{210}, Leu_{196}, Thr_{309}	Tyr_{67}
	E140A	−7.2	−8.0	Gln_{209}	Glu_{66}, Asn_{311}	Tyr_{67}, Ala_{70}, Leu_{74}, Leu_{121}, Ala_{140}, Arg_{215}	Tyr_{67}, Ala_{70}, Leu_{74}, Leu_{121}, Ala_{140}, Met_{310}
	I141N	−7.4	−6.5	Tyr_{67}, Lys_{73}, Ser_{213}, Thr_{309}	Pro_{59}, Lys_{126}, His_{137}, His_{345}	His_{137}, Leu_{196}, Ala_{210}	Ala_{64}, Tyr_{67}, Ser_{124}
	L303R	−7.3	−6.3	Gln_{209}, Asn_{211}, Arg_{215}, Gln_{312}	Asn_{211}, Ser_{213}, Ser_{281}	Tyr_{67}, Tyr_{69}, Thr_{309}	Leu_{273}, Pro_{279}, Met_{310}, Phe_{342}
	W306C	−7.0	−7.5	Cys_{53}, Ser_{55}, Glu_{66}, Ser_{339}, Met_{344}, His_{345}	Ser_{55}, Met_{344}	Lys_{62}, Asn_{311}	Gly_{56}, Pro_{59}, Lys_{62}, His_{345}
	L308Q	−7.3	−7.2	Lys_{73}, His_{137}, Gln_{209}, Ser_{213}, Arg_{215}	Ser_{55}, Asn_{311}, Thr_{343}, Met_{344}, His_{345}	Tyr_{67}, Ala_{70}, Thr_{309}	Pro_{59}, His_{137}
	I349N	−7.2	−7.4	Cys_{53}, Thr_{343}, Met_{344}, His_{345}, Leu_{347}	Ser_{55}, Gly_{63}, Met_{344}	Pro_{59}, Lys_{62}, Tyr_{67}	His_{54}, Pro_{59}, Lys_{62}, His_{345}

3.14. Biological validation of reported SNP

According to our study, nsSNPs of the *SLC30A8* gene may have a detrimental impact on several fatal diseases like T2D, Colorectal cancer, and so on. The *SLC30A8* gene polymorphism rs13266634 raises the probability of developing T2D in Jordanian population (Mashal et al. [62]). A different study detected the *SLC30A8* rs13266634 and rs16889462 genetic variations in Chinese individuals diagnosed with T2D, which were connected with susceptibility to the disease and the effectiveness of repaglinide as a treatment (Huang et al. [63]). Also, other research groups investigated and examined the correlation between the rs13266634 polymorphism and T2D in Iranian populations (Faghih et al. [64]) and the rs13266634 *SLC30A8* gene variant in Han Chinese and minority ethnic groups in China (Wang et al. [65]).

The association between diabetes and colorectal cancer risk was modified by variants on chromosome 8q24.11 within the *SLC30A8* gene, with rs3802177 being the genetic variant showing the most significant effect (Dimou et al. [2]). Four single-nucleotide polymorphisms (SNPs) in T2D were found to increase the risk of colorectal cancer: rs7578597 (THADA), rs864745 (JAZF1), rs5219 (KCNJ11), Rs7961581 (TSPAN8, LGR5) (Cheng et al. [66]). The patients who possessed the TCF7L2_rs7903146_T gene variant were more likely to develop colorectal cancer (Sainz et al. [67]).

Above all, *the SLC30A8* gene with different rs IDs potentially impacts T2D and Colorectal cancer in humans. We evaluated eight rs IDs of nsSNPs of *the SLC30A8* gene, possibly linked to several diseases, including T2D, colorectal cancer, etc.

4. Discussion

Single nucleotide polymorphisms (SNPs) can change the structure and function of proteins, thereby influencing human diseases, traits, drug responses, and therapeutic outcomes [68,69]). However, characterizing these SNPs is a laborious and expensive technique due to their complex relationship with specific phenotypes. *In silico* approaches, which utilize computer simulations to analyze biological data, can provide insights into genetic differences in disease susceptibility and their phenotypic impacts, hence lowering the number of candidates for molecular screening [70].

This research determines genetic variations within the *SLC30A8* gene to assess their potential impact on gene function. Out of the 336 documented missense SNPs, eight high-risk SNPs (rs746249658, rs765852756, rs771191320, rs780551100, rs866700996, rs1300885615, rs1255515664, and rs1325443204) were determined to be harmful by nine distinct computational techniques (Table 1). Two computational approaches, MuPro and I-Mutant 2.0, were employed to assess the impact of 29 identified nsSNPs on protein stability. According to I-Mutant, the S182F, S182P, H304P, H304L, and H137P SNPs are anticipated to enhance protein stability. Meanwhile, MuPro suggested that the R138L SNP would also increase protein stability. Since mutations stabilizing the protein were not the focus of this study, they were excluded from this study. Only the 25 SNPs, which were consistently forecasted by both tools to instabilize the protein, were kept for further studies (Table 2).

Utilizing ConSurf, a study evaluated how the residues in the *SLC30A8* protein have evolved according to their conservation. This

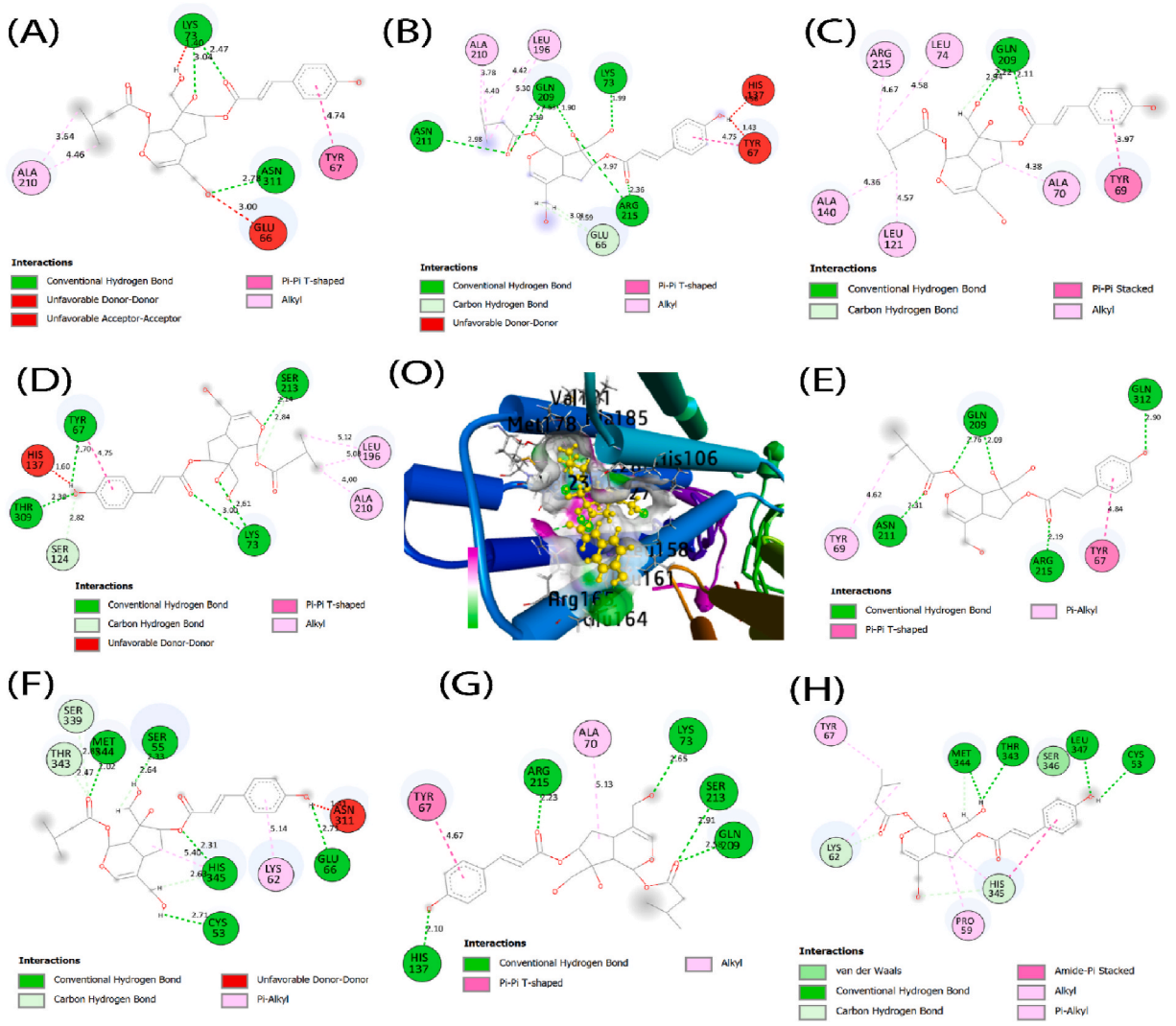

Fig. 10. Analysis of binding interactions of (O) raw and mutated *SLC30A8* proteins using Autodock Vina. From the right side, (A) depicts the interaction between W136G and Luzonoid B, (B) R138Q and Luzonoid B, (C) E140A and Luzonoid B, (D) I141N and Luzonoid B, (E) L303R and Luzonoid B, (F) W306C and Luzonoid B, (G) L308Q and Luzonoid B, (H) I349N and Luzonoid B.

analysis revealed that mutations are more likely to occur in conserved regions, where critical amino acid positions change less frequently due to their importance for protein function. An amino acid's evolutionary conservation indicates its mutability, with highly conserved and exposed residues most prone to harm (Ashkenazy et al. [36]). The Evolutionary Trace approach and its variations are the most well-known substitute for ConSurf for discovering functional areas [71,72] Among 25 nsSNPs, ConSurf revealed 22 nsSNPs with a conservation score of 8 or 9, indicating they were highly conserved (Supplementary Table S2, and Fig. 3). Afterward, the study of mutation 3D servers revealed the correlation of W136G, R138Q, E140A, I141N, L303R, W306C, L308Q, and I349N nsSNPs with cancer (Fig. 4). GPS-MSP 1.0 anticipated R325's methylation location. The NetPhos 3.1 program predicted the phosphorylation sites to be wild-type residues (S307) and mutant residues (S299) (Fig. 5(A-B)).

Protein stability alterations influence the conformational structure, dictating protein function [73]. In this research, the MutPred2 web server was employed to investigate potential molecular alterations in the structure or function of the *SLC30A8* protein caused by mutations. According to their g and p scores, every detected deleterious SNP was categorized as "pathogenic" (Table 3). Among the eight analyzed nsSNPs, W136G possesses the highest probability score, indicating its significant potential to affect protein structure and function. Conversely, R138Q exhibits the lowest g-value, suggesting a comparatively less impact. This analysis reveals G47R as the most impactful mutation, with the remaining seven potentially influencing protein stability and possibly affecting structure and function. As demonstrated by previous studies, reduced protein stability can result in altered folding mechanisms, accelerated breakdown, and erroneous aggregation [74,75]).

Project HOPE provides new insights into the adverse effects of point mutations on protein structure. It demonstrates notable

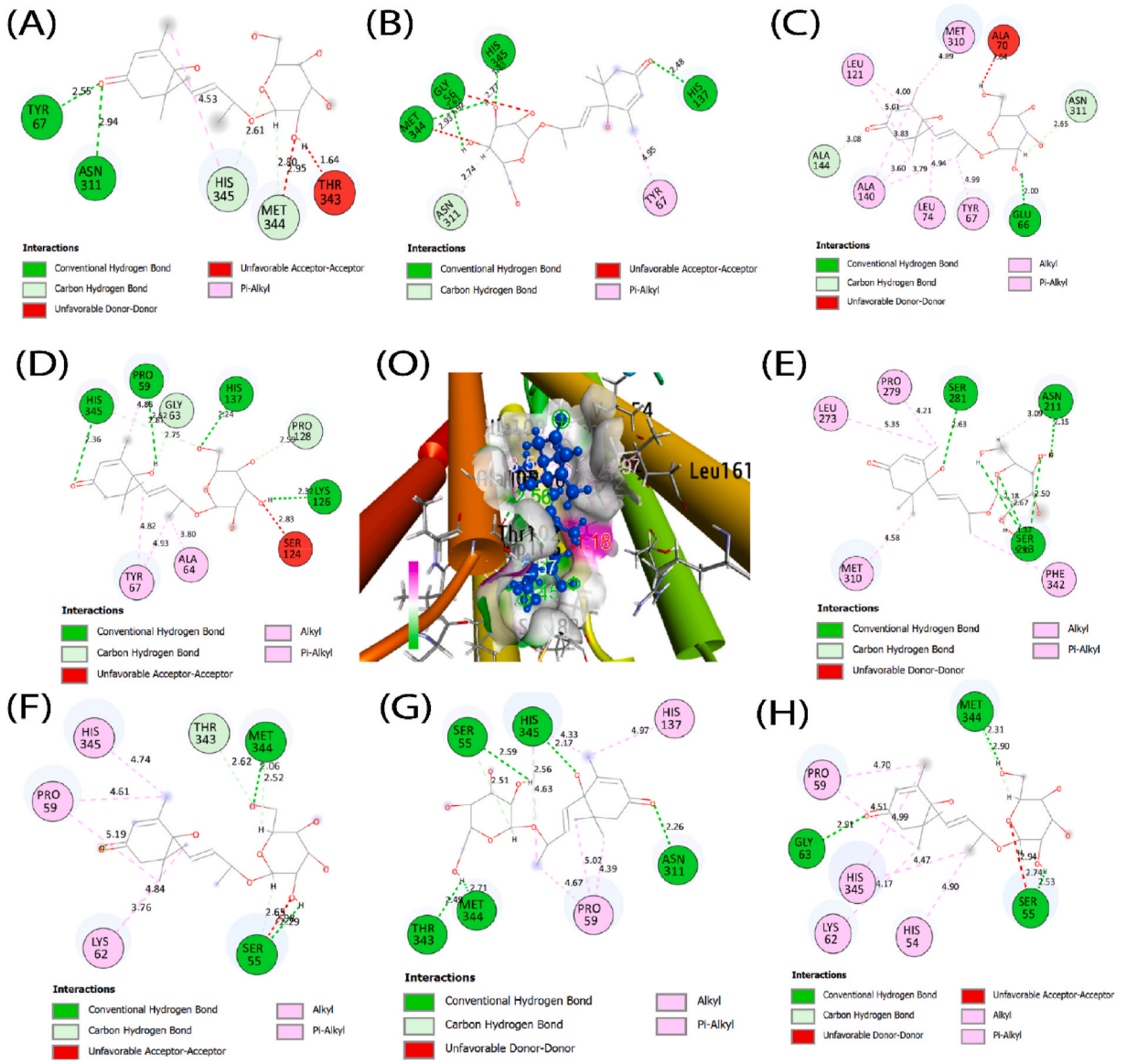

Fig. 11. Analysis of binding interactions of (O) raw and mutated *SLC30A8* proteins using Autodock Vina. From the right side, (A) depicts the interaction between W136G and Roseoside, (B) R138Q and Roseoside, (C) E140A and Roseoside, (D) I141N and Roseoside, (E) L303R and Roseoside, (F) W306C and Roseoside, (G) L308Q and Roseoside, (H) I349N and Roseoside.

variances in physicochemical characteristics between wild and mutant amino acids, encompassing size, charge, and hydrophobicity. Notably, the mutations I141N, L303R, L308Q, and I349N result in increased amino acid size compared to the wild-type counterpart, whereas W136G, R138Q, E140A, and W306C mutations lead to a size decrease. Moreover, three mutations (R138Q, E140A, and L303R) exhibit altered amino acid charges in the mutant variant. Regarding hydrophobicity, W136G, I141N, L303R, W306C, L308Q, and I349N mutations render the protein less water-repellent than the wild-type residue, while E140A exhibits increased hydrophobicity, potentially impacting hydrophobic interactions. Notably, R138Q does not contribute to hydrophobic interactions (Table 4).

STRING analysis shows that the *SLC30A8* protein plays crucial roles in several vital processes. It works with other proteins like *PTPRN, INS*, and *TCF7L2* to maintain normal levels of necessary brain chemicals like norepinephrine, dopamine, and serotonin; boost critical metabolic processes in the liver, including glycolysis, the pentose phosphate cycle, and glycogen synthesis; and additionally, involves in the important Wnt signalling pathway (Fig. 8). *ZnT8,* a protein encoded by *SLC30A8*, may transport zinc ions from the cytoplasm to the granules that secrete insulin. While high zinc concentrations within these granules are well-documented, the precise physiological function of secreted zinc remains unclear. It validated three loci in a noncoding region close to *CDKN2A* and *CDKN2B*, in an intron of *IGF2BP2*, and in an intron of *CDKAL1*, which was shown to be related to T2D. Additionally, it replicated associations close to *HHEX* and in *SLC30A8* [76,77].

Table 6
Analysis of reported compounds, considering their ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity), physicochemical properties, lipophilicity, and drug-likeness.

Properties	Model Name	luzonoid B	Roseoside
Absorption	Intestinal absorption (human)	96.016	47.786
	Skin Permeability	−2.781	−2.859
	caco-2	0.176	0.379
Distribution	VDss (human)	−0.113	−0.131
	Fraction unbound (human)	0.228	0.601
	BBB permeability	−1.227	−1.067
	CNS permeability	−3.526	−3.632
Metabolism	CYP2D6 substrate	No	No
	CYP3A4 substrate	Yes	No
	CYP2D6 inhibitor	No	No
	CYP3A4 inhibitor	No	No
Excretion	Total Clearence	0.873	1.389
Toxicity	AMES toxicity	No	No
	Oral Rat Acute Toxicity (LD5)	2.808	2.2
	Hepatotoxicity	No	No
	Skin Sensitization	No	No
Physicochemical	Molecular weight	462.49 g/mol	386.44 g/mol
	LogP	1.4985	−0.576
	Num. H-bond acceptors	9	8
	Num. H-bond donors	4	5
	Molar Refractivity	117.94	96.23
	TSPA	142.75 $Å^2$	136.68 $Å^2$
Lipophilicity	Consensus Log Po/w	1.61	0.01
Water Solubility	Log S (ESOL)	−2.91	−1.24
	Solubility class	Soluble	Soluble
Drug-likeness	Lipinski violation	Yes; 0 violation	Yes; 0 violation
	Bioavailability Score	0.55	0.55
Medicinal Chemistry	PAINs	0 alert	0 alert

We employed molecular docking to explore the possibility that luzonoid B and roseoside bind to a mutated *SLC30A8* protein, which plays a role in both T2D and colorectal cancer [2,9] Based on their exceptional binding affinity to the eight mutated *SLC30A8* proteins along with normal protein, luzonoid B and roseoside were identified as the most promising compounds with promising binding scores (Figs. 10 and 11 and Table 5). This suggests that the two compounds exhibit the strongest attraction to the mutant proteins and are more effective at inhibiting them compared to the normal *SLC30A8* protein. The increasing binding affinity towards mutant *SLC30A8* proteins has been demonstrated not to contribute to drug resistance [78,79].

We evaluated the pharmacokinetics of the selected compounds to ensure they can be effectively absorbed, distributed, metabolized, and excreted, allowing them to reach their target sites and exert their desired effects [80]. Lower molecular weight (<500) enhances permeability, facilitating transport, diffusion, and absorption through the cell membrane [81]. Selected compounds displayed molecular weights under 500 g/mol, including Luzonoid B (462.49 g/mol) and Roseoside (386.44 g/mol). Higher positive LogP values suggest enhanced permeation of compounds across biological membranes, with a recommended threshold below 5 considered acceptable [82]. LogS indicates drug solubility, with lower values representing higher solubility [83]. These results demonstrate the compounds as promising candidates for further evaluation as drugs for T2D and colorectal cancer in living organisms (*in vivo*). Optimal membrane permeation for drugs requires a specific range of hydrogen bonds (H-donors: <5 H-acceptors: <10). Optimal oral bioavailability necessitates limiting the number of rotatable bonds in a drug molecule to a maximum of 10. ADMET prediction suggests that all two docked compounds possess properties within the allowed range for predicted drug development (Table 6).

The *SLC30A8* gene dysregulation was found to have a prognostic significance and to impact the total survival rate for individuals who have suffering from breast cancer, colorectal cancer, sarcoma, and head-neck squamous cell carcinoma using the Kaplan-Meier plotter. All four malignancies combined posed a decreased death risk to those with greater expression of *SLC30A8*. These results suggest any dysregulation of *SLC30A8* expression may significantly impact patient survival in these four cancers (Fig. 12). While the *in-silico* approach is powerful, the study would be strengthened by experimental validation including *in vitro* or *in vivo* of the predicted effects in near future. This approach allows researchers to efficiently predict the impact of genetic mutations on protein function and disease pathophysiology, thereby prioritizing variants for further experimental validation and saving time and resources in the research process [84]. Therefore, we will be emphasized the need for experimental validation to elucidate how *SLC30A8* affects cancer progression and investigate its viability as a potential drug in the future of the predicted impacts of identified mutations.

5. Conclusion

Analyzing genetic variations (SNPs) in *SLC30A8* suggests it might be an essential gene for impacting the onset of type 2 diabetes, a recognized risk factor for colorectal cancer. Eight potentially harmful nsSNPs of *SLC30A8* were found in this study using various *in silico* methods. Finding these nsSNPs should facilitate quick and effective screening to identify diseases linked to *SLC30A8* expression.

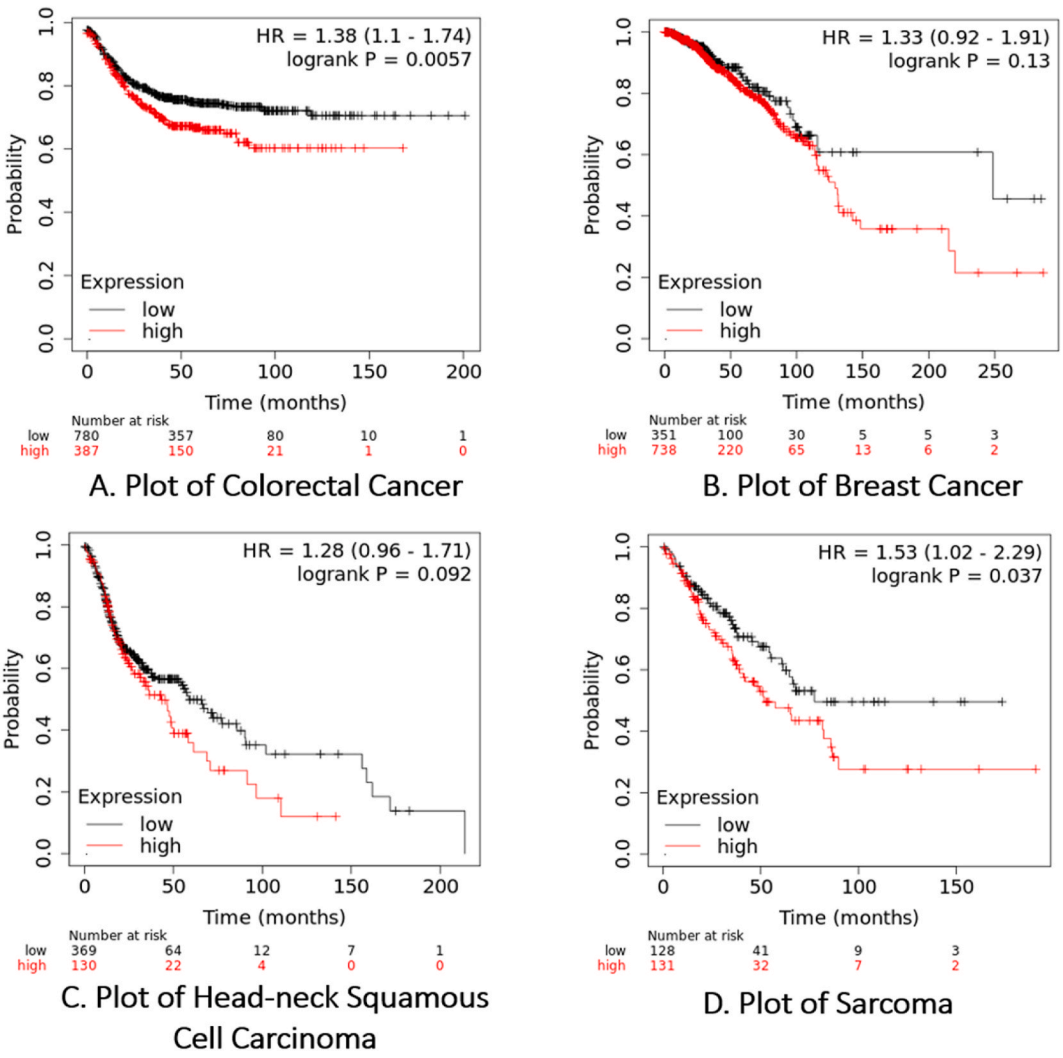

Fig. 12. *SLC30A8* gene expression levels and survival rates in patients with various cancer types using microarray data by Kaplan–Meier Plotter.

After identifying nsSNPs of *SLC30A8*, we developed a novel drug against its harmful effects; this suggests that the two compounds exhibit the strongest attraction to the mutant proteins and are more effective at inhibiting them compared to the normal *SLC30A8* protein. The high binding affinity towards mutant *SLC30A8* proteins has been exhibited not to contribute to drug resistance, which might be detrimental to curing these fatal diseases. Moreover, it will greatly simplify the process of creating experiments for upcoming lab-based research.

Clinical significance

In our study, identification of specific genetic polymorphisms can serve as biomarkers for early detection of diseases progression, helping to tailor treatment plans based on individual risk profiles. In addition, studying such types of polymorphisms can provide insights into the molecular pathways as well as prognostic indicators allowing for more proactive management of T2D and CRC. In the area of pharmacogenomics, understanding how genetic variations affect drug metabolism can optimize drug selection and dosing, minimizing side effects and maximizing efficacy. Besides, advances in gene editing technologies, such as CRISPR, might allow for the correction of our identified genetic variants that contribute to T2D and CRC.

Limitations

In silico methods can predict SNP functional effects, although they are only sometimes accurate. Database accuracy and completeness are crucial to *in silico* SNP analysis. Reference data errors or gaps can affect SNP identification and interpretation. The SNP identification of *SLC30A8* is just the beginning. Interpreting their functional significance or relationship with other genetic traits/

diseases and environmental factors related to glucose homeostasis and insulin secretion is complicated and requires further experimental validation and *in vitro or in vivo* research.

Even though *SLC30A8* has been linked to insulin secretion, diabetes risk, and many malignancies, our knowledge of its genetic diversity and the functional effects of its variants continues growing. This suggests that some crucial genetic variations and their impacts may be insufficiently identified using computer-based SNP analysis. Since SNPs in genes' coding regions are more likely to affect gene function, most *in silico* techniques concentrate on these regions. On the other hand, SNPs in non-coding areas should be more appropriately characterized for their potential importance. Finally, SNP impacts on *SLC30A8* function typically necessitate experimental validation.

Data availability

The datasets supporting the conclusions of the study are included within the article and its additional file as a supplementary file. Correspondence and requests for materials should be addressed to Kawsar Ahmed and Md. Arju Hossain. All supporting data are available on a publicly available respiratory (Github) https://github.com/Md-Arju-Hossain/In-silico-Single-neucleotide-Polymorphisms.

Funding Statement

The Funding number is 4170008.

Ethical Statement

We declared that, our manuscript submitted to the Heliyon Journal has been done in accordance to Cell Press guidelines of publication ethics in a responsible way.

CRediT authorship contribution statement

Md. Moin Uddin: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Md. Tanvir Hossain:** Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Md. Arju Hossain:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Asif Ahsan:** Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation. **Kamrul Hasan Shamim:** Writing – original draft, Visualization, Resources, Methodology, Investigation. **Md. Arif Hossen:** Writing – original draft, Visualization. **Md. Shahinur Rahman:** Writing – original draft, Visualization. **Md Habibur Rahman:** Writing – original draft, Validation, Supervision, Project administration, Methodology. **Kawsar Ahmed:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Conceptualization. **Francis M. Bui:** Writing – review & editing, Validation, Project administration, Funding acquisition. **Fahad Ahmed Al-Zahrani:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37280.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[2] N. Dimou, A.E. Kim, O. Flanagan, et al., Probing the diabetes and colorectal cancer relationship using gene–environment interaction analyses, Br. J. Cancer (2023) 1–10, https://doi.org/10.1038/s41416-023-02312-z.

[3] Y. Guo, Y. He, Comprehensive analysis of the expression of SLC30A family genes and prognosis in human gastric cancer, Sci. Rep. 10 (2020) 18352, https://doi.org/10.1038/s41598-020-75012-w.

[4] L.T.D. Nguyen, M. Gunathilake, J. Lee, et al., Zinc intake, SLC30A8 rs3802177 polymorphism, and colorectal cancer risk in a Korean population: a case–control study, J. Cancer Res. Clin. Oncol. 149 (2023) 16429–16440, https://doi.org/10.1007/s00432-023-05381-y.
[5] M. Ujpal, O. Matos, G. Bibok, et al., Diabetes and oral tumors in Hungary: epidemiological correlations, Diabetes Care 27 (2004) 770–774, https://doi.org/10.2337/diacare.27.3.770.
[6] C.H. Wu, T.Y. Wu, C.C. Li, et al., Impact of diabetes mellitus on the prognosis of patients with oral squamous cell carcinoma: a retrospective cohort study, Ann. Surg Oncol. 17 (2010) 2175–2183, https://doi.org/10.1245/s10434-010-0996-1.
[7] C.H. Tseng, Pioglitazone and oral cancer risk in patients with type 2 diabetes, Oral Oncol. 50 (2014) 98–103, https://doi.org/10.1016/j.oraloncology.2013.10.015.
[8] M. Cui, Q. Fang, J. Zheng, et al., Kaposi's sarcoma-associated herpesvirus seropositivity is associated with type 2 diabetes mellitus: a case–control study in Xinjiang, China, Int. J. Infect. Dis. 80 (2019) 73–79, https://doi.org/10.1016/j.ijid.2019.01.003.
[9] L. Cheng, D. Zhang, L. Zhou, et al., Association between SLC30A8 rs13266634 polymorphism and type 2 diabetes risk: a meta-analysis, Med Sci Monit Int Med J Exp Clin Res 21 (2015) 2178, https://doi.org/10.12659/MSM.894052.
[10] F. Chimienti, S. Devergnas, A. Favier, et al., Identification and cloning of a β-cell–specific zinc transporter, ZnT-8, localized into insulin secretory granules, Diabetes 53 (2004) 2330–2337, https://doi.org/10.2337/diabetes.53.9.2330.
[11] F. Chimienti, S. Devergnas, F. Pattou, et al., In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion, J. Cell Sci. 119 (2006) 4199–4206, https://doi.org/10.1242/jcs.03164.
[12] J.E. Lee, J.H. Choi, J.H. Lee, et al., Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis, Mutat Res Mol Mech Mutagen 573 (2005) 195–204, https://doi.org/10.1016/j.mrfmmm.2004.08.018.
[13] L. Prokunina, M.E. Alarcón-Riquelme, Regulatory SNPs in complex diseases: their identification and functional validation, Expert Rev Mol Med 6 (2004) 1–15, https://doi.org/10.1017/S1462399404007690.
[14] P.D. Stenson, M. Mort, E.V. Ball, et al., The human gene mutation database: 2008 update, Genome Med. 1 (2009) 1–6, https://doi.org/10.1186/gm13.
[15] F.S. Collins, L.D. Brooks, A. Chakravarti, A DNA polymorphism discovery resource for research on human genetic variation, Genome Res. 8 (1998) 1229–1231, https://doi.org/10.1101/gr.8.12.1229.
[16] E.S. Lander, The new genomics: global views of biology, Science 80 (274) (1996) 536–539, https://doi.org/10.1126/science.274.5287.536.
[17] R.J. Dobson, P.B. Munroe, M.J. Caulfield, et al., Predicting deleterious nsSNPs: an analysis of sequence and structural attributes, BMC Bioinf. 7 (2006) 1–9, https://doi.org/10.1186/1471-2105-7-217.
[18] I. Barroso, M. Gurnell, V.E.F. Crowley, et al., Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension, Nature 402 (1999) 880–883, https://doi.org/10.1038/47254.
[19] M. Petukh, T.G. Kucukkal, E. Alexov, On human disease-causing amino acid variants: statistical study of sequence and structural patterns, Hum. Mutat. 36 (2015) 524–534, https://doi.org/10.1002/humu.22770.
[20] D. Chasman, R.M. Adams, Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation, J. Mol. Biol. 307 (2001) 683–706, https://doi.org/10.1006/jmbi.2001.4510.
[21] T.G. Kucukkal, M. Petukh, L. Li, et al., Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins, Curr. Opin. Struct. Biol. 32 (2015) 18–24, https://doi.org/10.1016/j.sbi.2015.01.003.
[22] R. Thomas, R. McConnell, J. Whittacker, et al., Sandford, Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR, Am. J. Hum. Genet. 65 (1999) 39–49, https://doi.org/10.1086/302460.
[23] S.C. Das, M.A. Rahman, S. Das Gupta, In-silico analysis unravels the structural and functional conseque'nces of non-synonymous SNPs in the human IL-10 gene, Egypt J Med Hum Genet 23 (2022) 10, https://doi.org/10.1186/s43042-022-00223-x.
[24] M.S. Hossain, A.S. Roy, M.S. Islam, In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions, Sci. Rep. 10 (2020) 14542, https://doi.org/10.1038/s41598-020-71457-1.
[25] M. Petrosino, L. Novak, A. Pasquo, R. Chiaraluce, P. Turina, E. Capriotti, et al., Analysis and interpretation of the impact of missense variants in cancer, Int. J. Mol. Sci. 22 (11) (2021) 5416, https://doi.org/10.3390/ijms22115416.
[26] Q. Zeng, B. Tan, F. Han, X. Huang, J. Huang, Y. Wei, et al., Association of solute carrier family 30 A8 zinc transporter gene variations with gestational diabetes mellitus risk in a Chinese population, Front. Endocrinol. 14 (2023) 1159714, https://doi.org/10.3389/fendo.2023.1159714.
[27] P.C. Ng, S. Henikoff, Predicting the effects of amino acid substitutions on protein function, Annu. Rev. Genom. Hum. Genet. 7 (2006) 61–80, https://doi.org/10.1146/annurev.genom.7.080505.115630.
[28] Subbiah H. Venkata, P. Ramesh Babu, U. Subbiah, In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene, Egypt J Med Hum Genet 21 (2020) 1–9, https://doi.org/10.1186/s43042-020-00110-3.
[29] J. Bendl, J. Stourac, O. Salanda, et al., PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations, PLoS Comput. Biol. 10 (2014) e1003440, https://doi.org/10.1371/journal.pcbi.1003440.
[30] E. Capriotti, R. Calabrese, P. Fariselli, et al., WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation, BMC Genom. 14 (2013) 1–7, https://doi.org/10.1186/1471-2164-14-S3-S6.
[31] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, Curr Protoc Hum Genet 76 (2013) 7–20, https://doi.org/10.1002/0471142905.hg0720s76.
[32] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005) W306–W310, https://doi.org/10.1093/nar/gki375.
[33] B. Kamaraj, R. Purohit, In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA32013, BioMed Res. Int. 2013 (2013) 697051.
[34] S. Khan, M. Vihinen, Performance of protein stability predictors, Hum. Mutat. 31 (2010) 675–684.
[35] G. Celniker, G. Nimrod, H. Ashkenazy, et al., ConSurf: using evolutionary data to raise testable hypotheses about protein function, Isr. J. Chem. 53 (2013) 199–206.
[36] H. Ashkenazy, S. Abadi, E. Martz, et al., ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, Nucleic Acids Res. 44 (2016) W344–W350, https://doi.org/10.1093/nar/gkw408.
[37] S. Ramazi, J. Zahiri, Post-translational modifications in proteins: resources, tools and prediction methods, Database 2021 (2021) baab012, https://doi.org/10.1093/database/baab012.
[38] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362, https://doi.org/10.1006/jmbi.1999.3310.
[39] Y. Xue, F. Zhou, M. Zhu, et al., GPS: a comprehensive www server for phosphorylation sites prediction, Nucleic Acids Res. 33 (2005) W184–W187, https://doi.org/10.1093/nar/gki393.
[40] V. Pejaver, J. Urresti, J. Lugo-Martinez, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat. Commun. 11 (2020) 5918, https://doi.org/10.1038/s41467-020-19669-x.
[41] M.J. Meyer, R. Lapcevic, A.E. Romero, et al., mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome, Hum. Mutat. 37 (2016) 447–456, https://doi.org/10.1002/humu.22963.
[42] H. Venselaar, T.A.H. Te Beek, R.K.P. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinf. 11 (2010) 548, https://doi.org/10.1186/1471-2105-11-548.
[43] Subbiah H. Venkata, P. Ramesh Babu, U. Subbiah, Determination of deleterious single-nucleotide polymorphisms of human LYZ C gene: an in silico study, J. Genet. Eng. Biotechnol. 20 (1) (2022) 92, https://doi.org/10.1186/s43141-022-00383-8.
[44] R. Islam, M. Rahaman, H. Hoque, et al., Computational and structural based approach to identify malignant nonsynonymous single nucleotide polymorphisms associated with CDK4 gene, PLoS One 16 (2021) e0259691, https://doi.org/10.1371/journal.pone.0259691.

[45] E.Y. Chen, C.M. Tan, Y. Kou, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (2013) 1–14, https://doi.org/10.1186/1471-2105-14-128.

[46] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (2016) W90–W97, https://doi.org/10.1093/nar/gkw377.

[47] D. Tang, M. Chen, X. Huang, et al., SRplot: a free online platform for data visualization and graphing, PLoS One 18 (2023) e0294236, https://doi.org/10.1371/journal.pone.0294236.

[48] B. Snel, G. Lehmann, P. Bork, et al., STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene, Nucleic Acids Res. 28 (2000) 3442–3444, https://doi.org/10.1093/nar/28.18.3442.

[49] D. Szklarczyk, A. Franceschini, S etal Wyder, STRING v10: protein–protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) D447–D452, https://doi.org/10.1093/nar/gku1003.

[50] D. Wu, Z. Wu, Superimposition of protein structures with dynamically weighted RMSD, J. Mol. Model. 16 (2010) 211–222, https://doi.org/10.1007/s00894-009-0538-6.

[51] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al., UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.

[52] J. Eberhardt, D. Santos-Martins, A.F. Tillack, et al., AutoDock Vina 1.2. 0: new docking methods, expanded force field, and python bindings, J. Chem. Inf. Model. 61 (2021) 3891–3898, https://doi.org/10.1021/acs.jcim.1c00203.

[53] P.W. Rose, A. Prlić, A. Altunkaya, et al., The RCSB protein data bank: integrative view of protein, gene and 3D structural information, Nucleic Acids Res. (2016), https://doi.org/10.1093/nar/gkw1000 gkw1000.

[54] L.A. Kelley, S. Mezulis, C.M. Yates, et al., The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858, https://doi.org/10.1038/nprot.2015.053.

[55] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, Electrophoresis 18 (15) (1997) 2714–2723, https://doi.org/10.1002/elps.1150181505.

[56] M.K. Gupta, R. Vadde, Insights into the structure–function relationship of both wild and mutant zinc transporter ZnT8 in human: a computational structural biology approach, J. Biomol. Struct. Dyn. (2019), https://doi.org/10.1080/07391102.2019.1567391.

[57] S. Kim, J. Chen, T. Cheng, et al., PubChem 2023 update, Nucleic Acids Res. 51 (D1) (2023) D1373–D1380, https://doi.org/10.1093/nar/gkac956.

[58] D.S. Biovia, Discovery Studio Visualizer V21. 1.0. 20298, Dassault Systèmes, San Diego, 2021.

[59] J.A. Pradeepkiran, K.K. Kumar, Y.N. Kumar, et al., Modeling, molecular dynamics, and docking assessment of transcription factor rho: a potential drug target in Brucella melitensis 16M, Drug Des Devel Ther 2015 (2015) 1897–1912, https://doi.org/10.2147/DDDT.S77020.

[60] D.E.V. Pires, T.L. Blundell, D.B. Ascher, pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures, J. Med. Chem. 58 (9) (2015) 4066–4072, https://doi.org/10.1021/acs.jmedchem.5b00104.

[61] X. Chen, Z. Miao, M. Divate, et al., KM-express: an integrated online patient survival and gene expression analysis tool for the identification and functional characterization of prognostic markers in breast and prostate cancers, Database 2018 (2018) bay069, https://doi.org/10.1093/database/bay069.

[62] S. Mashal, M. Khanfar, S. Al-Khalayfa, et al., SLC30A8 gene polymorphism rs13266634 associated with increased risk for developing type 2 diabetes mellitus in Jordanian population, Gene 768 (2021) 145279, https://doi.org/10.1016/j.gene.2020.145279.

[63] Q. Huang, J.Y. Yin, X.P. Dai, et al., Association analysis of SLC30A8 rs13266634 and rs16889462 polymorphisms with type 2 diabetes mellitus and repaglinide response in Chinese patients, Eur. J. Clin. Pharmacol. 66 (2010) 1207–1215, https://doi.org/10.1007/s00228-010-0882-6.

[64] H. Faghih, S.R. Khatami, N. Azarpira, A.M. Foroughmand, SLC30A8 gene polymorphism (rs13266634 C/T) and type 2 diabetes mellitus in south Iranian population, Mol. Biol. Rep. 41 (2014) 2709–2715, https://doi.org/10.1007/s11033-014-3158-x.

[65] Y. Wang, L. Duan, S. Yu, et al., Association between" solute carrier family 30 member 8"(SLC30A8) gene polymorphism and susceptibility to type 2 diabetes mellitus in Chinese Han and minority populations: an updated meta-analysis, Asia Pac. J. Clin. Nutr. 27 (6) (2018) 1374–1390, https://doi.org/10.6133/apjcn.201811_27(6).0025.

[66] I. Cheng, C.P. Caberto, A. Lum-Jones, et al., Type 2 diabetes risk variants and colorectal cancer risk: the Multiethnic Cohort and PAGE studies, Gut 60 (12) (2011) 1703–1711, https://doi.org/10.1136/gut.2011.237727.

[67] J. Sainz, A. Rudolph, M. Hoffmeister, et al., Effect of type 2 diabetes predisposing genetic variants on colorectal cancer risk, J. Clin. Endocrinol. 97 (5) (2012) E845–E851, https://doi.org/10.1210/jc.2011-2565.

[68] M.P. Miller, S. Kumar, Understanding human disease mutations through the use of interspecific genetic variation, Hum. Mol. Genet. 10 (21) (2001) 2319–2328, https://doi.org/10.1093/hmg/10.21.2319.

[69] A. Vignal, D. Milan, M. SanCristobal, et al., A review on SNP and other types of molecular markers and their use in animal genetics, Genet. Sel. Evol. 34 (3) (2002) 275–305, https://doi.org/10.1186/1297-9686-34-3-275.

[70] T.C. Dakal, D. Kala, G. Dhiman, et al., Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene, Sci. Rep. 7 (1) (2017) 6525, https://doi.org/10.1038/s41598-017-06575-4.

[71] I. Mihalek, I. Reš, O. Lichtarge, A family of evolution–entropy hybrid methods for ranking protein residues by importance, J. Mol. Biol. 336 (5) (2004) 1265–1282, https://doi.org/10.1016/j.jmb.2003.12.078.

[72] S. Sankararaman, B. Kolaczkowski, K. Sjölander, INTREPID: a web server for prediction of functionally important residues by evolutionary analysis, Nucleic Acids Res. 37 (suppl_2) (2009) W390–W395, https://doi.org/10.1093/nar/gkp339.

[73] M.C. Deller, L. Kong, B. Rupp, Protein stability: a crystallographer's perspective, Acta Crystallogr Sect F, Struct Biol Commun 72 (Pt 2) (2016) 72–95, https://doi.org/10.1107/S2053230X15024619.

[74] S.M. Singh, N. Kongari, J. Cabello-Villegas, et al., Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-β aggregates, Proc Natl Acad Sci 107 (34) (2010) 15069–15074, https://doi.org/10.1073/pnas.1008818107.

[75] S. Witham, K. Takano, C. Schwartz, et al., A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics, Proteins 79 (8) (2011) 2444–2454, https://doi.org/10.1002/prot.23065.

[76] R. Saxena, B.F. Voight, V. Lyssenko, et al., Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels, Science 316 (80) (2007) 1331–1336, https://doi.org/10.1126/science.1142358, 5829.

[77] L.J. Scott, K.L. Mohlke, L.L. Bonnycastle, et al., A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants, Science (80-) 316 (5829) (2007) 1341–1345, https://doi.org/10.1126/science.1142382.

[78] J. Flannick, G. Thorleifsson, N.L. Beer, et al., Loss-of-function mutations in SLC30A8 protect against type 2 diabetes, Nat. Genet. 46 (4) (2014) 357–363, https://doi.org/10.1038/ng.2915.

[79] L. Sui, Q. Du, A. Romer, et al., ZnT8 loss of function mutation increases resistance of human embryonic stem cell-derived beta cells to apoptosis in low zinc condition, Cells 12 (6) (2023) 903, https://doi.org/10.3390/cells12060903.

[80] M. Yang, X. Liu, T. Hou, et al., Synthesis and luminescent properties of GdNbO4: Bi3+ phosphors via high temperature high pressure, J. Alloys Compd. 723 (2017) 1–8, https://doi.org/10.1016/j.jallcom.2017.06.204.

[81] W.P. Walters, Going further than Lipinski's rule in drug design, Expert Opin Drug Discov 7 (2) (2012) 99–107, https://doi.org/10.1517/17460441.2012.648612.

[82] N.A. Al-Shabib, J.M. Khan, A. Malik, et al., Molecular insight into binding behavior of polyphenol (rutin) with beta lactoglobulin: spectroscopic, molecular docking and MD simulation studies, J. Mol. Liq. 269 (2018) 511–520, https://doi.org/10.1016/j.molliq.2018.07.122.

[83] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (1) (2017) 42717, https://doi.org/10.1038/srep42717.

[84] A. Hoda, M. Lika, V. Kolaneci, Identification of deleterious nsSNPs in human HGF gene: in silico approach, J. Biomol. Struct. Dyn. 41 (21) (2023) 11889–11903, https://doi.org/10.1080/07391102.2022.2164060.