Exploring the Structural and Functional Effects of Nonsynonymous SNPs in the Human Serotonin Transporter Gene Through *In Silico* Approaches

Md Arzo Mia[1*], Md Nasir Uddin[1*] (ID), Yasmin Akter[1], Jesmin[2] and Lolo Wal Marzan[1] (ID)

[1]Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong, Bangladesh. [2]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh.

ABSTRACT: The sodium-dependent serotonin transporter *SLC6A4* (solute carrier family 6 member 4) gene encodes an intrinsic membrane protein that transmits the serotonin neurotransmitter from synaptic clefts into presynaptic neurons. The product of the *SLC6A4* gene is related to the regulation of mood and social behavior, sleep, appetite, memory, digestion, and sexual desire. This protein is a target for antidepressant and psychostimulant drugs, thus prolonged neurotransmitter signaling remains blocked. In this study, the functional consequences of nsSNPs in the human *SLC6A4* gene were explored through computational tools: PhD-SNP, SIFT, Align GVGD, PROVEAN, PMut, nsSNP Analyzer, SNPs&GO, SNAP2, PolyPhen2, and PANTHER to identify the most deleterious and damaging nsSNPs. Then the mutant protein stabilities were assessed using I-Mutant, MUpro, and MutPred2; amino acid conservation using ConSurf, and posttranslational modification analysis using MusiteDEEP and PROSPER. Furthermore, the 3-dimensional (3D) model of the mutated proteins was predicted and validated using SPARKS-X, Verify3D, and PROCHECK. The protein–ligand binding sites were analyzed using the COACH meta-server. Results from this study predicted that T192M, G342E, R607C, W282S, R104C, P131L, P156L, and N351S were the most structurally and functionally significant nsSNPs in the human *SLC6A4* gene. Arg607 and Pro156 were the predicted sites for posttranslational modifications, and Thr192 and Try282 were the ligand-binding sites in the human *SLC6A4* gene. The analyzed data also suggested that R104C, P131L, P156L, T192M, G342E, and W282S mutants might affect the binding of sodium ions with this protein. Taken together, this study provided important information on structurally and functionally important nsSNPs of the human *SLC6A4* gene for further experimental validation. In the future, these damaging nsSNPs of the *SLC6A4* gene have the potential to be evaluated as prognostic biomarkers for *SLC6A4*-related disorder diagnosis and research.

KEYWORDS: Serotonin transporter, single nucleotide polymorphisms (SNPs), nonsynonymous, nsSNPs, *SLC6A4* gene, *in silico*

Introduction

Serotonin, known as the happy hormone, is a crucial neurotransmitter that modulates vital processes of the body and brain. The serotonin transporter gene (*SLC6A4*) is a very important target for the candidate genes involved in psychiatric disorders such as bipolar disorder (BP) and schizophrenia (SCZ), obsessive-compulsive disorder (OCD), anxiety disorder, depression, autism, seizure, eating disorder, attention-deficit hyperactivity disorder (ADHD), and substance abuse disorders.[1,2] The serotonin transporter protein, the *SLC6A4* gene product is located on chromosome 17q11.2.[3,4] The Human *SLC6A4* gene contains 15 exons spanning ~40 kb, while the human serotonin transporter protein contains 630 amino acids with 12 transmembrane domains. It has been seen that both normal and pathological association of the *SLC6A4* serotonin transporter gene variants was identified with human behaviors.[5] Usually, the SLC6A4 protein helps the cell in up taking the right amount of serotonin. However, variations due to different polymorphisms of this gene might affect the normal function of the gene product. In humans, the most common source of genetic variations is single nucleotide polymorphisms (SNPs).[6] Single nucleotide alterations can occur both in the intronic and exonic regions of a gene. However, SNPs in the coding region have a higher impact on the functional properties of the gene product and are known as the nonsynonymous SNPs (nsSNPs).[7,8] Evidences from genetic research show that more than 50% of the SNPs that are associated with genetic disorders are nonsynonymous (nsSNPs), also known as the missense variants.[8,9] The nsSNPs may affect protein functions by lowering the protein solubility or reducing the stability of the protein structure.[10-13] Thus, studying the association between different SNPs and their phenotypic impacts can help in understanding the molecular basis of many complex hereditary diseases.[5,14-18]

In this study, we aimed to identify the most deleterious and damaging nsSNPs of the human *SLC6A4* gene to unveil the structural–functional relationship between the genetic polymorphisms and their phenotypic effects using *in silico* approaches. Several open databases for SNPs, such as GWAS Central, dbSNP, and Swiss-Var6 were used to extract the missense SNPs data of the human *SLC6A4* gene.

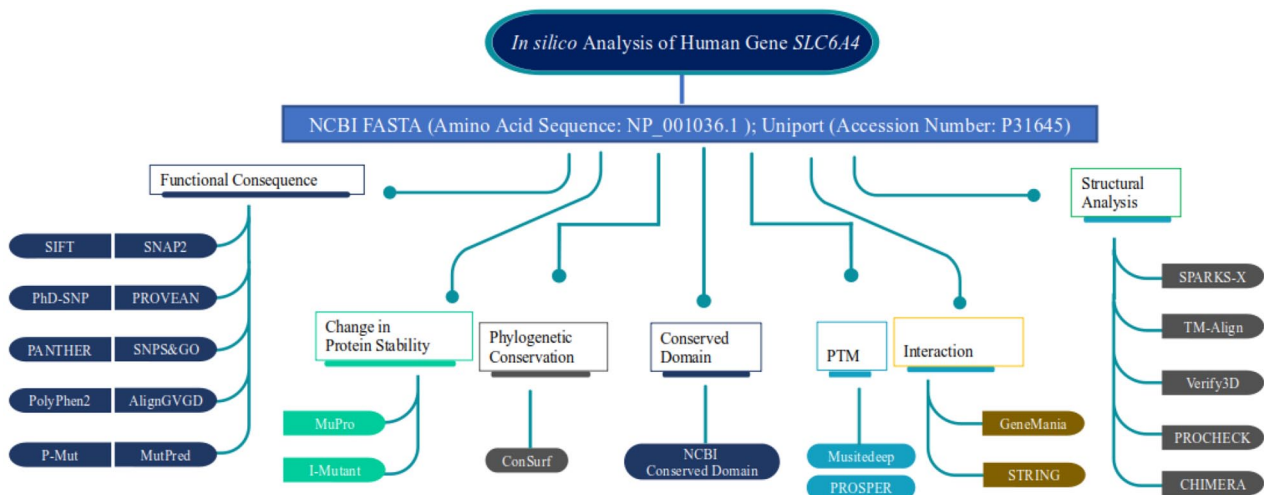* Md Arzo Mia and Md Nasir Uddin contributed equally to this work.

Figure 1. Schematic representation of whole work.
PTM indicates posttranslational modification.

We investigated the functional consequences of the missense SNPs: whether they are normal or disease-causing or effective by any chance using SIFT, Align GVGD, PolyPhen2, PROVEAN, SNAP2, P-Mut, PhD-SNP, SNPs&GO, and PANTHER.[6,17,19] The stability of the mutated proteins was analyzed using computational tools like MUpro and I-Mutant. Then most potential nsSNPs were further analyzed using MutPred2.[20] Conservation of the amino acid residues was predicted using ConSurf.[19] We also investigated the posttranslational modification (PTM) sites in the human SLC6A4 protein using Musite and PROSITE.[21,22] Mutated protein structures were generated by SPARKS-X, and the quality of the protein models was validated by Varify3D and PROCHECK. Furthermore, molecular characteristics and interactions of the predicted protein structures were investigated using UCSF Chimera. The ligand-binding sites were analyzed using COACH.[16,23-25] Together, this study conducted a thorough, in-depth computational analysis on all the nsSNPs of the *SLC6A4* gene to predict and identify the most damaging and deleterious nsSNPs in humans. The flow chart of the overall methodology is shown below (Figure 1).

Materials and Methods

SNPs data

The nucleotide, SNPs, and protein data of the *SLC6A4* gene were retrieved from the following database: all SNPs (rs IDs) were extracted from the NCBI database of SNP (dbSNP) (http://www.ncbi.nlm.nih.gov/snp/). FASTA format of the nucleotide sequence (NC_000017.11) and amino acid sequence (NP_001036.1) were retrieved from NCBI (https://www.ncbi.nlm.nih.gov) and Uniprot ID (UniprotKB = P31645) from Uniprot database (https://www.uniprot.org) were retrieved for further computational analysis.

Prediction of the functional effects of nonsynonymous SNPs

The SIFT (Sorting Intolerance from Tolerance) tool employs an algorithm that determines if an amino acid substitution has an impact on protein function based on sequence homology and physicochemical qualities.[26,27] The substitution is considered deleterious if the SIFT score is between 0 and 0.05, and it is considered tolerant if the SIFT value is between 0.05 and 1.[28] The rs IDs of *SLC6A4* SNPs from the dbSNP data set used here as the input key for SIFT tool.

Align GVGD (http://agvgd.hci.utah.edu) is a freely available tool that predicts amino acid variants based on Grantham variation (GV) and Grantham deviation (GD) score.[29] Align GVGD produces a score with 7 classes (C0, C15, C25, C35, C45, C55, and C65), with C0 being neutral, C15 to C55 being less likely influenced, and C65 being the most likely affected.[30] The input key for Align GVGD was the FASTA format of the SLC6A4 protein and the position of amino acid substitutions.

Screening of non-acceptable Polymorphism2 (SNAP2) is a computational tool (https://www.rostlab.org/services/SNAP/). It predicts whether the amino acid variation is effective or neutral. The input query was the FASTA format of the SLC6A4 protein sequence.

Protein Variation Effect Analyzer (PROVEAN) is a technique that detects nonsynonymous variants that can alter protein function. PROVEAN (http://provean.jcvi.org/index.php) uses alignment-based ratings to determine whether an amino acid variation is deleterious or neutral. If the score range is less than –2.5, the variant is regarded detrimental, while a score range of more than –2.5 is considered neutral.[31] The input query was amino acid variants and FASTA format of the SLC6A4 protein sequence.

Polymorphism Phenotyping v2 (PolyPhen2) is an online tool (http://genetics.bwh.harvard.edu/pph2/) that predicts the effect

of amino acid substitutions on the structure and function of human proteins using the physical and evolutionary comparison.[32] This algorithm calculates PSIC (Position-Specific independent score). A score greater than 0.85 indicated probably damaging and >0.15 predicted possibly damaging otherwise designated as benign.[32] The input query was the FASTA format of the SLC6A4 protein sequence and amino acid variants.

PANTHER (http://pantherdb.org/tools/csnpScoreForm.jsp) predicts specific nonsynonymous SNP that affect protein function using the PSEP (position-specific evolutionary preservation) method. Through PSEP scores, it predicts whether the amino acid substitution is probably benign or damaging.[20,33] The input key was the SLC6A4 protein sequence and amino acid substitution.

Predictor of human harmful single nucleotide polymorphism (PhD-SNP) (http://snps.biofold.org/phd-snp/phd-snp.html) uses the support vector machine (SVM) method to discriminate between neutral and disease-related single-point amino acid polymorphisms.[34] The results were sequence- and profile-based, whereas reliability scores between 0 and 9 determined the amino acid substitution as disease-causing or neutral. The input query was the SLC6A4 protein sequence, residue position, and altered residue.

Single nucleotide polymorphism and gene ontology (SNPs&GO) is an online server (http://snps-and-go.biocomp.unibo.it/snpsand-go/) that predicts the effects of single amino acid change in protein sequence and function related to human diseases.[35] The input query was the UniProt accession number of the SLC6A4 protein (P31645) and its amino acids substitute variants.

P-Mut (http://mmb.pcb.ub.es/PMut/) is a free program that can predict pathogenic mutations with an accuracy of 80% and indicate users whether a single-point amino acid mutation is diseased or neutral.[36] The input query was the FASTA format of the SLC6A4 protein sequence and variations.

MUpro (http://mupro.proteomics.ics.uci.edu/) is a web server that accurately predicts protein stability change (due to amino acid substitution) based on SVM and neural network >84% accuracy through 20-fold time cross-validation.[37] The input query was the plain sequence of SLC6A4 protein, mutation position, and original residue as well as substituted residue.

I-Mutant (http://folding.biofold.org/cgi-bin/i-mutant2.0) is a support vector tool used to determine protein stability change due to the substitution of an amino acid in a protein sequence. Prediction of the protein stability change is based on RI (Reliability Index) score from 0 to 10, where 0 shows the lowest, and 10 shows the highest reliability.[38] The input query was the SLC6A4 protein sequence, substitution position, and new residue.

Mutational association with the disease by MutPred

MutPred2 (http://mutpred.mutdb.org/) is a machine learning–based tool that predicts whether amino acid substitutions are pathogenic or not and their molecular mechanisms. It uses to screen functional and structural variations such as altered stability, loss catalytic site, and gain O-linked glycosylation. MudPred2 provides a result with a probability score where more than 0.5 is considered as deleterious and >0.75 is considered as most deleterious.[39-41] The input query was the FASTA amino acid sequence of the SLC6A4 protein.

Conservation analysis of deleterious nsSNP in SLC6A4

The ConSurf server (http://consurf.tau.ac.il) is a bioinformatics tool for predicting the evolutionary conservation of amino acid residues in protein sequence based on the phylogenetic association between similar sequences.[19] A conservation score (ranging from 1 to 9) of 1 to 3 indicates variable residues, 4 to 7 indicates average conserved residues, and 8 to 9 indicates the most conserved residues.[42,43] The input data were the FASTA format of the SLC6A4 protein sequence.

GnomAD

The genome Aggregation Database (gnomAD) is an open-source bioinformatics tool (https://gnomad.broadinstitute.org/) that provides MAF value to distinguish between common and rare variants in the population. The MAF value of rare variants is less than 0.05, whereas the common variants are greater than 0.0.[44] The input query was the *SLC6A4* gene name.

Prediction of the posttranslational site's modification

The MusiteDeep (https://www.musite.net) is an online tool that gives a general model for protein PTM site prediction and visualization within the protein sequence. Posttranslational modification, such as phosphorylation, glycosylation, ubiquitination, sumoylation, acetyl-lysine, methylation, pyrrolidone carboxylic acid, palmitoylation, and hydroxylation is identified by the MusiteDeep server.[45] PROSPER (https://prosper.erc.monash.edu.au/) is a web server for computer simulation and prediction of 24 different protease types of protease substrates and their cleavage sites, covering 4 leading protease families: aspartic acid (A) and cysteine (C), Metal (M), and serine (S). It is applied an algorithm-based approach to anticipate protease cleavage locales by using diverse but complementary sequence and structure characteristics.[22] The input query, both MusiteDeep, and PROSPER, was the FASTA format of the SLC6A4 protein sequence.

Prediction of nsSNPs positions in different protein domains

NCBI Conserved Domain Search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) is used to determine conserved domains and motifs in a particular protein. It is used to

predict nsSNPs location in different domains of SLC6A4 protein structure and provides functional analysis of proteins.[46,47] The input key was the FASTA format of the SLC6A4 protein.

Prediction and validation of mutant 3D protein structure

SPARKS-X (http://sparks-lab.org/yueyang/server/SPARKS-X/) server is used to generate the 3D structure of the mutant proteins. It uses template-based modeling, and the degree of similarity of templates was checked by BLASTp.[48] To create the SLC6A4 mutant protein structure, the changed amino acid has to be placed in the specified place, and this modified FASTA format amino acid sequence was an input query for SPARKS-X.

TM-align server (https://zhanglab.ccmb.med.umich.edu/TM-align/) checks the similarity between wild-type and mutant models. It is a structural alignment program for comparing 2 proteins whose sequences can be different. The output of TM-align is TM-score (template modeling score), RMSD (root-mean-square deviation), and structural superposition. TM-score ranges from 0 to 1, where 1 indicates a perfect match between 2 structures, scores less than 0.2 determine unrelated proteins, and more than 0.5 generally assume a similar fold in SCOP/CATH. Root-mean-square deviation values also determine variations between wild- and mutant-type structures, whereas higher RMSD value assumes more significant variation.[49,50] The input query was a wild and mutant protein structure. The input query for Varify-3D and PROCHECK was the protein structure generated by SPARKS-X. Protein structures generated by SPARKS-X were checked using both Varify-3D (http://servicesn.mbi.ucla.edu/Verify3D) and PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/).[24] Finally, Chimera V1.14 was used to study the features and interactive visualization of the predicted protein structure at the molecular level.[25]

Ligand–binding site prediction

COACH (http://zhanglab.ccmb.med.umich.edu/COACH/) is a web tool used for protein–ligand binding site prediction. COACH provides a C-score (confidence score) that determines binding site reliability of the protein-ligand interaction. C-score ranges (0-1), where a higher score indicates a higher reliable prediction. Cluster size is the whole number of templates in a cluster. Ligand lists provide all ligands in a cluster.[51] The input key was the SLC6A4 protein structure.

Results
SNPs data

This study investigated the *SLC6A4* gene, and SNP data are taken from the dbSNP database (dbSNPNCBI: https://www.ncbi.nlm.nih.gov/snp/?term=SLC6A4). It contains 10 593

SNPs, out of which 360 are missense (nsSNPs), 198 are synonymous, 72 are noncoding transcript variants, 2 are inframe deletion, 2 are inframe insertion, and 8572 are intronic sequence (Figure 2). Only nsSNPs of *SLC6A4* were selected for this study.

Identification of deleterious nsSNPs

All the nsSNPs retrieved from the dbSNP database were subjected to various bioinformatics tools for the prediction of functional nsSNPs in the *SLC6A4* gene. Through SIFT analysis, 89 SNPs were predicted to be tolerated or deleterious out of a total of 360 nsSNPs. From these 89 SNPs, SIFT classified 67 as tolerated and 22 as deleterious. All the 89 SNPs predicted in SIFT were further validated by Align GVGD, SNAP2, PROVEAN, PolyPhen2, and PANTHER, PhD-SNP, SNPs&GO, P-Mut, MUpro, I-Mutant tools to increase the accuracy of computational techniques (Table 1).

Out of 89 nsSNPs, Align GVGD anticipated 38 SNPs as the most likely affected and 50 nsSNPs as less likely involved, and 1 predicted neutral. SNAP2 exhibited, 28 had an effect on protein function and 61 anticipated as neutral. PROVEAN analysis anticipated 31 SNPs were as deleterious, whereas 58 SNPs were neutral. PolyPhen-2 server, predicted 34 SNPs as probably damaging, 54 SNPs were determined as benign, and 1 SNP was not predicted by PolyPhen2. Out of 89 nsSNPs, 33 nsSNPs were predicted probably damaging, 36 nsSNPs predicted possibly damaging, and 20 nsSNPs predicted probably benign using PANTHER (Table 1).

Total 35 SNPs showed disease association, and the rest of 54 predicted neutral by PhD-SNP server. SNPs&GO predicted 16 as diseased and 73 as neutral, whereas the P-Mut predicted 29 SNPs disease-causing, and 60 SNPs predicted neutral. The SNPs were further analyzed for their impact on protein stability using MUpro and I-Mutant. MUpro predicted 81 nsSNPs, with decreased SLC6A4 protein stability, and 8 nsSNPs showed increased protein stability. I-Mutant predicted 13 nsSNPs that increased SLC6A4 protein stability and 76 nsSNPs with decreased protein stability (Table 1).

From all these analyses, we identified the 15 nsSNPs that met the criteria and predicted by all 11 different algorithms as harmful SNPs. We selected these 15-high risk nsSNPs for further analysis using MutPred and ConSurf (Table 2). MutPred results showed that many nsSNPs may cause protein alteration and may affect their function or structure (Supplementary File 3). The ConSurf server predicted Gly342, Trp282, Arg104, Pro131, Pro156, and Asn315 as highly conserved with a conservation score of 9 and predicted as buried or exposed as well as functional or structural residues. The Arg607 was also predicted as highly conserved (conservation score 8) and exposed and predicted as functional residue. Arg596 predicted variable residue, and 7 amino acids were predicted averagely conserved (Table 3).
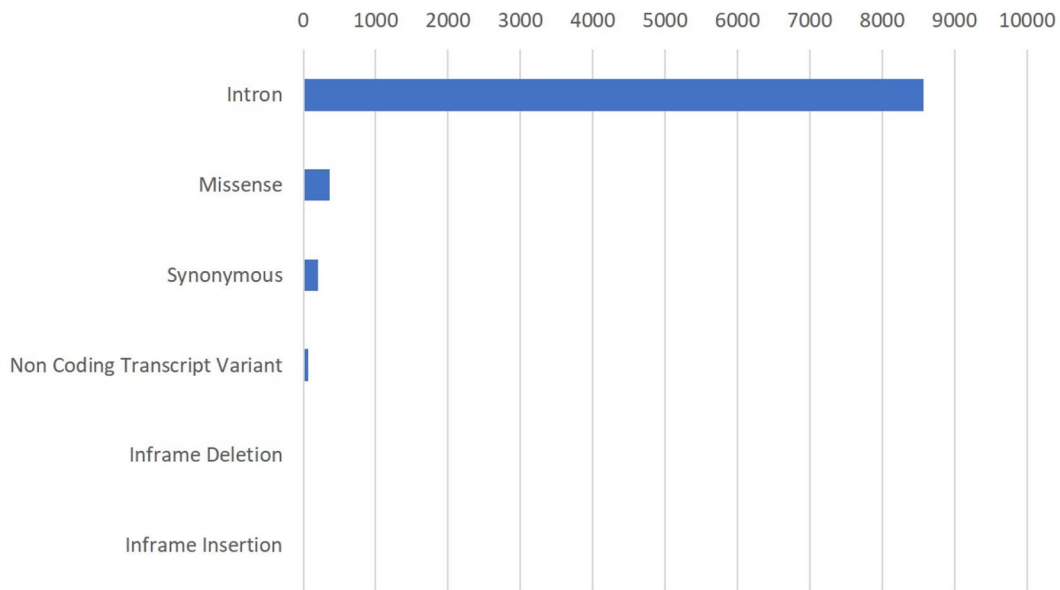
Figure 2. Distribution of SNPs according to the dbSNP database among different *SLC6A4* gene functional classes.
dbSNP indicates database of SNP; SNP, single nucleotide polymorphisms.

Prediction of the posttranslational modification sites

Posttranslational modification sites associated with the selected 15 most potent nsSNPs were predicted using Musite and PROSITE. Ten out of the 15 most-significant nsSNPs were predicted to be involved in PTM, including O-linked glycosylation, N-linked glycosylation, proteolytic cleavage, phosphorylation, methylation, and hydroxylation. Residues R607, W282, and P156 were anticipated to have sites for proteolytic cleavage, whereas W607 and P156 also had methylation and hydroxylation sites, respectively. The results of Musite and PROSITE are shown in Table 3.

Prediction of minor allelic frequency (MAF)

The MAF data for the selected nsSNPs of the *SLC6A4* gene was extracted from the gnomAD database. The highest frequency was found for T192M, P533L, and G530S, while the lowest frequency was found for I270T, G342E, P303H, R607C, and N351S. The result of the MAF is given in Table 3.

Prediction of nsSNPs position in different protein domains

NCBI-conserved domain search tool figured 2 major domains in the *SLC6A4* gene. One was SLC6sbd-SERT domain (Na (+) and Cl (–)-dependent serotonin transporter SERT), which comprises 79-615 amino acids, and another 1 was 5-HT_transport_N domain (Serotonin (5-HT) neurotransmitter transporter, N-terminus) which comprises 24-64 amino acid. In SLC6A4, 208 and 217 amino acids were present in the putative glycosylation site; 94-437 amino acid sequences were present in Na-binding site 2; 96-168 amino acid sequences were present in Na binding site 1; 95-442 amino acid sequences were present in putative substrate-binding site 1 and 103-407 amino acid were present in putative substrate-binding site 2 (Figure 3).

Ligand-binding site prediction

The SLC6A4 protein–ligand binding site was predicted by the COACH server. COACH server predicted IXX and site 4 ligand could bind to the Thr192 site. Again, Y01, CLR could bind to the Trp282 site. The confidence score (C-score) and the predicted binding site residues by COACH were shown (Supplementary File 4).

Prediction and validation of mutant 3D protein structure

The 3D structure of SLC6A4 protein is available in the protein data bank (PDB id: 6VRH). From the ConSurf, PTM, and MAF results, a total of 8 SNPs were the most potential disease-causing nsSNPs. The 3D models of these 8 nsSNPs mutant proteins were built using the SPARKS-X server. It gave the 10 best protein structures. From them, the structure with the highest Z-score was taken for our study. For each of the mutant protein structures, first one was selected. Again, these mutant structures were validated by Verify 3D and analyzed using PROCHECK for Ramachandran plot analysis. The RMSD values and TM score between wild-type and mutant models were analyzed using TM-align (Supplementary File 5). The 3D structure of wild-type and mutant proteins was analyzed through UCSF Chimera. These 8 mutant proteins showed a significant alteration in H-bonding interactions of amino acids compared to native (Supplementary File 1 and 2).

Table 1. Prediction of the effect of nsSNP by different tools.

SNP	AA CHANGE	SIFT	POLYPHEN 2	SNAP2	PROVEAN	PANTHER	ALIGN GVGD	P-MUT	PHD-SNP	SNPS& GO	MUPRO	I-MUTANT
rs6352	K605N	Tol	B	E	Del	ProD	C65	Dis	N 7	N	Dec	Dec
rs28914832	I425V	Tol	B	N	N	ProD	C25	N	N 8	N	Dec	Dec
rs138004662	V283L	Tol	B	E	N	ProD	C25	Dis	Dis 7	N	Dec	Dec
rs6355	G56A	Tol	B	E	N	ProB	C55	N	N 8	N	Dec	Dec
rs2228673	K201N	Tol	B	N	N	PosD	C65	N	N 6	N	Dec	Dec
rs28914833	F465L	Tol	B	E	Del	ProD	C15	N	N 6	N	Dec	Dec
rs28914834	L550V	Tol	B	N	N	PosD	C25	N	N 3	N	Dec	Dec
rs55848249	D193N	Tol	B	N	N	PosD	C15	N	N 3	N	Dec	Dec
rs55908511	V488M	Tol	B	N	N	PosD	C15	N	N 1	N	Dec	Dec
rs56316081	I108V	Tol	B	N	N	PosD	C25	N	N 3	N	Inc	Inc
rs60067068	G41A	Tol	B	N	N	PosD	C55	N	N 7	N	Dec	Dec
rs74330808	I270T	Del	PD	E	Del	PosD	C65	Dis	Dis 0	N	Dec	Dec
rs75354642	P601S	Tol	PD	N	Del	ProD	C65	Dis	N 2	N	Dec	Dec
rs75808495	E2D	Tol	B	N	N	PosD	C35	N	N 9	N	Dec	Dec
rs112636079	M370V	Tol	B	N	N	PosD	C15	N	Dis 2	N	Dec	Dec
rs117750329	V236I	Tol	PD	N	N	ProD	C25	Dis	N 6	N	Dec	Dec
rs140206260	I179V	Tol	PD	N	N	ProD	C25	N	N 6	N	Dec	Dec
rs140436169	G41R	Del	PD	N	N	PosD	C65	N	N 5	N	Dec	Dec
rs140484986	A65V	Tol	B	N	N	PosD	C55	N	N 5	N	Dec	Inc
rs142071015	I266V	Tol	B	N	N	ProB	C25	N	N 7	N	Dec	Dec
rs142441982	R626H	Tol	B	N	N	PosD	C25	N	N 5	N	Dec	Dec
rs142505940	R241L	Tol	B	E	Del	PosD	C65	N	N 5	N	Dec	Inc
rs143632225	A419V	Tol	PD	N	Del	PosD	C55	N	Dis 7	N	Dec	Inc
rs144427337	T192M	Del	PD	E	Del	PosD	C65	Dis	Dis 5	N	Dec	Inc
rs145643221	I157V	Tol	PD	N	N	PosD	C25	Dis	N 2	N	Dec	Dec

(Continued)

Table 1. (Continued)

SNP	AA CHANGE	SIFT	POLYPHEN 2	SNAP2	PROVEAN	PANTHER	ALIGN GVGD	P-MUT	PHD-SNP	SNPS& GO	MUPRO	I-MUTANT
rs145732192	T433M	Tol	PD	N	Del	ProD	C65	N	Dis 4	Dis	Inc	Dec
rs147306146	R144Q	Tol	PD	E	Del	ProD	C35	N	Dis 2	Dis	Dec	Dec
rs184149069	N211S	Tol	B	N	N	ProD	C45	N	N 4	N	Dec	Dec
rs190758123	V457I	Tol	B	N	N	ProB	C25	N	N 9	N	Dec	Dec
rs191881524	L222P	Tol	-	N	N	ProB	C65	N	Dis 5	N	Dec	Dec
rs199504488	E215K	Tol	B	N	N	ProB	C55	N	Dis 2	N	Dec	Dec
rs199727635	G25R	Del	PD	N	N	PosD	C65	N	N 8	N	Dec	Dec
rs199821523	D57E	Tol	B	N	N	ProB	C35	N	N 7	N	Dec	Dec
rs199832478	I576V	Tol	B	N	N	ProB	C25	N	N 8	N	Dec	Dec
rs199840777	I553V	Tol	B	N	N	PosD	C25	N	N 7	N	Dec	Dec
rs199873504	T421I	Tol	PD	E	Del	ProD	C65	Dis	Dis 3	N	Dec	Dec
rs199876253	G342E	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Inc	Dec
rs199886280	M389I	Tol	B	N	N	ProD	C0	N	Dis 4	N	Dec	Inc
rs199890537	A449G	Tol	B	N	N	ProD	C55	N	N 0	N	Dec	Inc
rs199909202	S293F	Tol	B	N	N	ProB	C65	N	Dis 4	N	Dec	Inc
rs199913287	I266T	Del	B	E	Del	ProB	C65	N	Dis 2	N	Dec	Dec
rs200015551	R607C	Del	PD	E	Del	ProD	C65	Dis	Dis 2	Dis	Inc	Inc
rs200078896	A72V	Tol	B	N	N	ProB	C55	N	N 2	N	Dec	Inc
rs200080084	A505V	Tol	B	N	N	ProB	C55	N	N 0	N	Dec	Inc
rs200126594	P621L	Tol	PD	N	Del	PosD	C65	N	N 4	N	Dec	Dec
rs200180716	R464Q	Tol	B	N	N	ProB	C35	N	N 4	N	Dec	Dec
rs200204643	A181V	Del	PD	E	Del	ProD	C55	Dis	Dis 5	N	Dec	Dec
rs200263321	N45S	Tol	B	N	N	ProD	C45	N	N 3	N	Dec	Dec
rs200339864	S48L	Tol	B	N	N	PosD	C65	N	N 4	N	Inc	Inc
rs200341915	F377S	Del	PD	E	Del	PosD	C65	Dis	Dis 5	Dis	Dec	Dec
rs200405588	I518F	Tol	B	N	N	PosD	C15	Dis	Dis 5	N	Dec	Dec

(Continued)

Table 1. (Continued)

SNP	AA CHANGE	SIFT	POLYPHEN 2	SNAP2	PROVEAN	PANTHER	ALIGN GVGD	P-MUT	PHD-SNP	SNPS& GO	MUPRO	I-MUTANT
rs200435184	W282S	Del	PD	E	Del	ProD	C65	Dis	Dis 2	Dis	Dec	Dec
rs200486204	T225M	Tol	PD	N	N	ProB	C65	N	N 1	N	Dec	Dec
rs200510224	E494D	Tol	B	N	N	ProD	C35	N	N 2	N	Dec	Dec
rs200544663	R596Q	Tol	PD	N	N	PosD	C35	N	N 7	N	Dec	Dec
rs200548683	H235L	Tol	B	N	Del	PosD	C65	Dis	Dis 5	N	Inc	Inc
rs200670218	Q22H	Tol	B	N	N	PosD	C15	N	N 9	N	Dec	Dec
rs200740988	R564Q	Tol	B	N	N	PosD	C35	N	N 7	N	Dec	Dec
rs200850098	P339L	Tol	PD	E	Del	ProD	C65	N	Dis 2	Dis	Inc	Dec
rs200924626	P303H	Del	PD	E	Del	ProD	C65	Dis	Dis 4	Dis	Dec	Dec
rs200953188	R104C	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Dec	Dec
rs200977199	N316S	Del	B	N	N	PosD	C45	Dis	N 2	N	Dec	Dec
rs200983126	M389V	Tol	B	N	N	ProD	C15	N	N 0	N	Dec	Dec
rs201041934	S62C	Tol	B	N	N	PosD	C65	N	N 5	N	Dec	Inc
rs201114547	P533L	Del	PD	E	Del	ProD	C65	Dis	Dis 7	Dis	Dec	Dec
rs201228840	R464W	Tol	B	E	N	ProB	C65	N	Dis 4	N	Dec	Dec
rs201369668	R523G	Tol	B	E	N	ProB	C65	N	Dis 2	N	Dec	Dec
rs201387005	A183V	Del	PD	E	Del	ProD	C55	N	Dis 1	N	Dec	Dec
rs201425535	A331S	Tol	B	N	N	ProD	C65	N	Dis 4	N	Dec	Dec
rs201480140	P131L	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Dec	Dec
rs201481838	R596W	Del	PD	E	Del	PosD	C65	Dis	N 5	Dis	Dec	Dec
rs201506679	N217S	Tol	B	N	N	PosD	C45	N	N 5	N	Dec	Dec
rs201518786	N208S	Tol	B	N	N	ProD	C45	Dis	N 0	N	Dec	Dec
rs201520306	I471S	Tol	B	N	N	ProB	C65	Dis	Dis 1	N	Dec	Dec
rs201520429	K610E	Tol	B	N	N	PosD	C55	Dis	N 8	N	Dec	Dec
rs201688096	N351S	Del	PD	E	Del	ProD	C45	Dis	Dis 7	Dis	Dec	Dec
rs201688297	T4M	Del	PD	N	N	PosD	C65	N	N 9	N	Dec	Inc

(Continued)

Table 1. (Continued)

SNP	AA CHANGE	SIFT	POLYPHEN 2	SNAP2	PROVEAN	PANTHER	ALIGN GVGD	P-MUT	PHD-SNP	SNPS& GO	MUPRO	I-MUTANT
rs201802369	A138T	Del	PD	N	Del	ProD	C55	Dis	Dis 1	Dis	Dec	Dec
rs201833332	I161T	Del	PD	E	Del	ProD	C65	Dis	Dis 6	N	Dec	Dec
rs201940331	P156L	Del	PD	E	Del	ProD	C65	Dis	Dis 7	Dis	Inc	Dec
rs202152288	V50L	Tol	B	N	N	PosD	C25	N	N 6	N	Dec	Dec
rs202166264	T497A	Tol	B	N	N	PosD	C55	N	N 8	N	Dec	Dec
rs202181933	I609T	Tol	B	N	N	ProB	C65	N	N 6	N	Dec	Dec
rs371274847	V363A	Tol	B	N	Del	PosD	C55	N	N 3	N	Dec	Dec
rs372056901	G35E	Tol	B	N	N	PosD	C65	N	N 4	N	Dec	Dec
rs374144565	G530S	Del	PD	E	Del	ProD	C55	Dis	Dis 7	Dis	Dec	Dec
rs374583307	A401T	Tol	B	N	N	ProB	C55	N	N 0	N	Dec	Dec
rs375503605	V26A	Tol	B	N	N	ProB	C55	N	N 8	N	Dec	Dec
rs375913512	P617L	Del	B	N	N	ProB	C65	N	N 3	N	Inc	Dec

Abbreviations: B, benign; Dec, decrease; Del, deleterious; Dis, disease; E, effect; Inc, increase; N, neutral; nsSNP, nonsynonymous SNPs; PD, probably damaging; PosD, possibly damaging; ProB, probably benign; ProD, probably damaging; SIFT, Sorting Intolerance from Tolerance; SNAP2, screening of nonacceptable polymorphism 2; SNP, single nucleotide polymorphism; SNPs&GO, single nucleotide polymorphism and gene ontology; Tol, tolerated.

Table 2. Concurrence of all the analyzing tools.

SNP	AA CHANGE	SIFT	POLYPHEN 2	SNAP2	PROVEAN	PANTHER	ALIGN GVGD	P-MUT	PHD-SNP	SNPS& GO	MUPRO	I-MUTANT
rs74330808	I270T	Del	PD	E	Del	PosD	C65	Dis	Dis 0	N	Dec	Dec
rs144427337	T192M	Del	PD	E	Del	PosD	C65	Dis	Dis 5	N	Dec	Inc
rs199876253	G342E	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Inc	Dec
rs200015551	R607C	Del	PD	E	Del	ProD	C65	Dis	Dis 2	Dis	Dec	Inc
rs200341915	F377S	Del	PD	E	Del	PosD	C65	Dis	Dis 5	Dis	Dec	Dec
rs200435184	W282S	Del	PD	E	Del	ProD	C65	Dis	Dis 2	Dis	Dec	Dec
rs200924626	P303H	Del	PD	E	Del	ProD	C65	Dis	Dis 4	Dis	Dec	Dec
rs200953188	R104C	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Dec	Dec
rs201114547	P533L	Del	PD	E	Del	ProD	C65	Dis	Dis 7	Dis	Dec	Dec
rs201480140	P131L	Del	PD	E	Del	ProD	C65	Dis	Dis 9	Dis	Dec	Dec
rs201833332	I161T	Del	PD	E	Del	ProD	C65	Dis	Dis 6	N	Dec	Dec
rs201940331	P156L	Del	PD	E	Del	ProD	C65	Dis	Dis 7	Dis	Inc	Dec
rs201481838	R596W	Del	PD	E	Del	PosD	C65	Dis	N 5	Dis	Dec	Dec
rs201688096	N351S	Del	PD	E	Del	ProD	C45	Dis	Dis 7	Dis	Dec	Dec
rs374144565	G530S	Del	PD	E	Del	ProD	C55	Dis	Dis 7	Dis	Dec	Dec

Abbreviations: Dec, decrease; Del, deleterious; Dis, disease; E, effect; Inc, increase; N, neutral; PD, probably damaging; PosD, possibly damaging; ProD, probably damaging; SIFT, Sorting Intolerance from Tolerance; SNAP2, screening of nonacceptable polymorphism 2; SNP, single nucleotide polymorphism; SNPs&GO, single nucleotide polymorphisms; SNPs&GO, single nucleotide polymorphism and gene ontology.

Table 3. Most deleterious nsSNP showing conservation predicted from ConSurf and their posttranslation sites prediction by Musite and PROSPER with their minor allelic frequency (MAF).

SNP	AA VARIANT	CONSERVATION SCORE	B/E	F/S	PTM	MAF
rs74330808	I270T	7	B	-		0.00000398
rs144427337	T192M	4	E	-	O-linked Glycosylation, Phosphorylation	0.0000119
rs199876253	**G342E**	**9**	**B**	**S**		**0.00000398**
rs200015551	**R607C**	**8**	**E**	**F**	**Methylation, Proteolytic Cleavage**	**0.00000812**
rs200341915	F377S	6	B	-	-	
rs200435184	**W282S**	**9**	**B**	**S**	**Proteolytic Cleavage**	
rs200924626	P303H	3	E		Hydroxylation	0.00000398
rs200953188	**R104C**	**9**	**E**	**F**	**Methylation**	
rs201114547	P533L	4	E		Hydroxylation	0.0000119
rs201480140	**P131L**	**9**	**B**	**S**	**Hydroxylation**	
rs201833332	I161T	7	B			
rs201940331	**P156L**	**9**	**E**	**F**	**Hydroxylation, Proteolytic Cleavage**	
rs201481838	R596W	1	B		Methylation	
rs201688096	**N351S**	**9**	**E**	**F**	**N-linked glycosylation**	**0.00000398**
rs374144565	G530S	8	B			0.0000142

Abbreviations: B, buried; E, exposed; F, functional; MAF, minor allelic frequency; nsSNP, nonsynonymous SNPs; PTM, posttranslational modification; S, structural; SNP, single nucleotide polymorphisms.
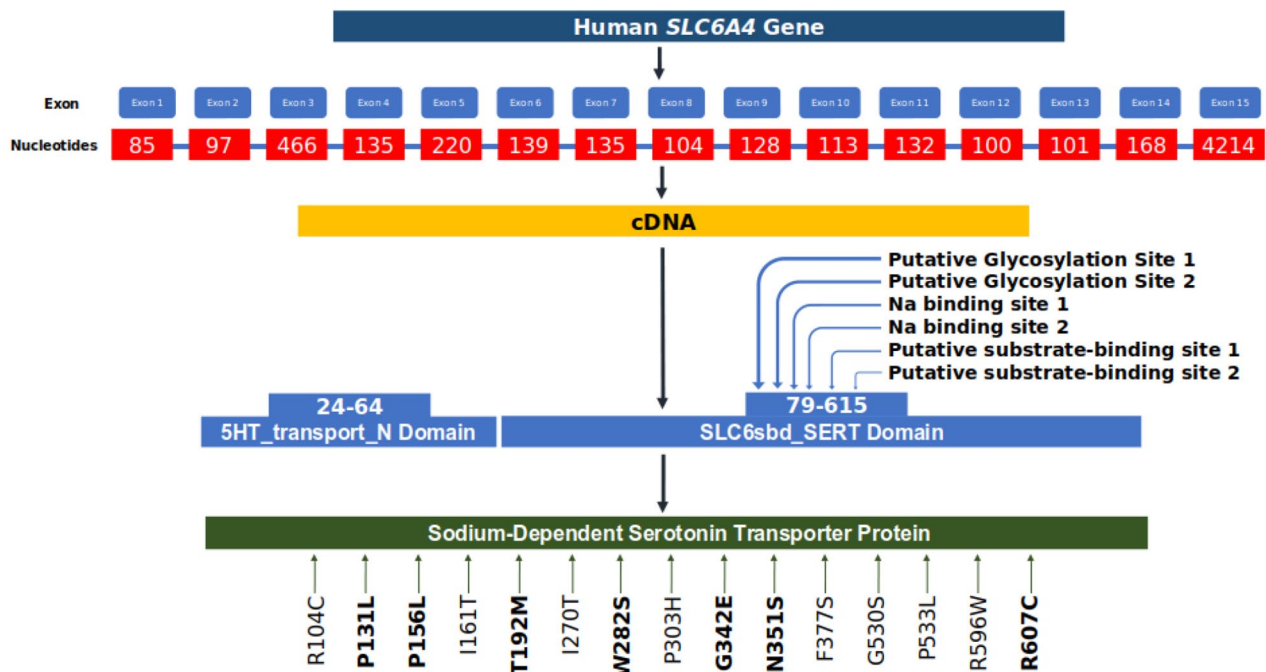The significant results are shown in bold in the table.

Figure 3. Graphical representation of the domain and position of nsSNP in *SLC6A4* gene and protein.
nsSNP indicates nonsynonymous SNPs.

Discussion

SLC6A4 gene encodes a serotonin transporter protein that carries the serotonin neurotransmitter from the synaptic cleft into presynaptic neurons. This protein ceases the action of serotonin and reuses it in a sodium-dependent manner.[52] Also, it is a target of taking many antidepressant drugs.[53,54] Polymorphisms in the *SLC6A4* gene has been shown to influence the rate of serotonin reuptake and play a significant role in numerous disease
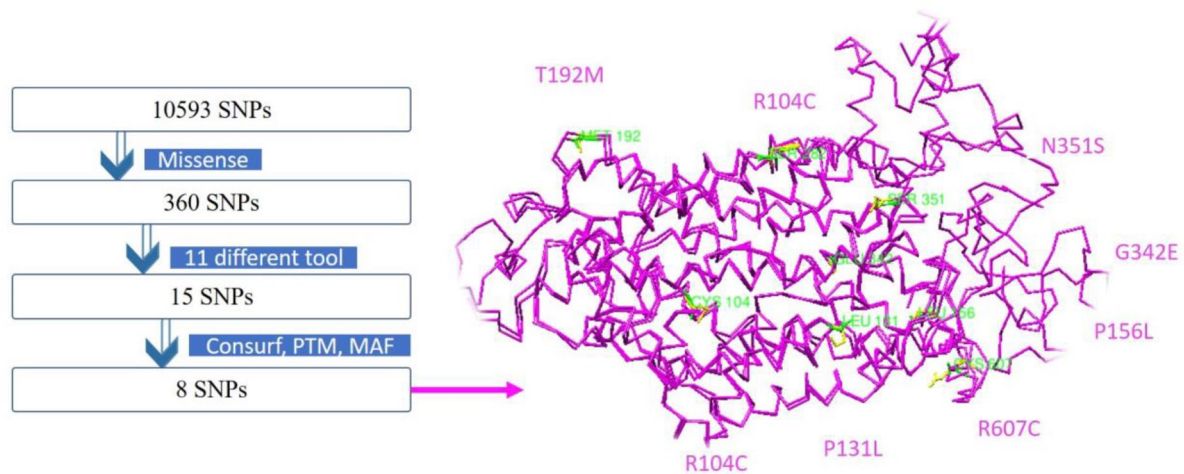
Figure 4. Concurrence of all the deleterious SNPs.
SNP indicates single nucleotide polymorphisms.

like autism, OCD, and major depressive disorder (MDD).[55] The G56A substitution in exon 2 of the *SLC6A4* gene has a prominent association with autism.[55] An I425V variation in exon 9, A1438G, and T102C SNPs of the *SLC6A4* gene were reported to be related to OCD.[56,57] Another study reported that the I550V polymorphism in exon 12 and the K605N in exon 13 of the *SLC6A4* gene were associated with MDD and nonfatal suicidal behavior in cases of autism and OCD in Chinese patients.[58]

In this study, *in silico* approaches were applied to screen and foresee the impacts of different SNPs on the structure-function of the *SLC6A4* gene. To date, more than 10 000 SNPs of the *SLC6A4* gene are reported in the dbSNP of NCBI, of which 360 polymorphisms are nonsynonymous (nsSNPs). The nsSNPs could either have a neutral effect or a major deleterious effect on protein 3D structure and function. For most of the nsSNPs of the *SLC6A4* gene, still, the potential to cause disease is not characterized. So, in this research work, we screened retrieved all the nsSNPs of the *SLC6A4* gene and then analyzed to identify the potential nsSNPs in the human *SLC6A4* gene that was deleterious, damaging, and disease-causing. We further studied the impacts of this nsSNP on the 3D protein structure, stability, and biological function using different Bioinformatics tools and algorithms. To evaluate the pathogenicity of the identified nsSNPs of the human *SLC6A4* gene, diverse structure-based algorithms along with machine learning tools were employed to infer and validate the predictions. We used 6 different bioinformatics tools (SIFT, PROVEAN, PolyPhen2, SNAP2, PANTHER, and Align GVGD) to evaluate the functional implications of nsSNPs of the human *SLC6A4* gene. In addition, 3 other tools (P-Mut, SNPs & GO, and PhD-SNP) were applied to determine the disease-causing nsSNPs of the human *SLC6A4* gene. Alterations of protein stability due to nsSNPs were predicted using MUpro and I-Mutant.

A total of 360 nonsynonymous SNPs of the human *SLC6A4* gene were retrieved and analyzed. SIFT identified 89 amino acid variants, of which 22 were predicted to have deleterious effects on the structure and the rest predicted as tolerable. These 89 SNPs were further explored to validate their effects on protein structure-function using MUpro and I-Mutant. MUpro-analyzed data indicated 81 out of 89 nsSNPs with decreased protein stability, whereas I-Mutant analysis predicted 76 nsSNPs associated with decreased protein stability. In literature, SNP rs25531 has been extensively studied in *SLC6A4*. This SNP is extensively studied in the population to check its relation with autism, depression, and anxiety, insomnia, irritable bowel syndrome, and ADHD.[59-63] Interestingly, the computational analysis of this study predicted this SNP as not deleterious. From all these analyses, we identified 15 substitutions that were found common using all the tools in this study. These 15 SNPs were predicted as deleterious or disease-causing and decreasing protein stability in the human *SLC6A4* gene product (Table 2).

These 15 screened nsSNPs of the human *SLC6A4* gene were further analyzed using bioinformatics tools: MutPred, ConSurf web server, Musite, PROSPER, gnomAD, NCBI conserved domain search tool, SPARKS-X, TM-Align, Varify3D, PROCHECK, COACH, GeneMenia, and STRING to evaluate the structural and functional properties *in silico*. MutPred results predicted 7 variants: W282S, P303H, R104C, P131L, and N351S as the highest damaging SNPs. These substitutions might alter the structures in ways that might alter cell membrane; a gain of a helix, a gain of relative solvent accessibility, loss of catalytic site or loss of metal-binding sites in the protein. The ConSurf-analyzed data predicted that G342, R607, W282, R104, P131, P156, N351, and G530 variants of the human *SLC6A4* gene were highly conserved (The conservation score is 8 to 9); of which 3 variants: G342, W282, and P131 were buried as important structural residues and R607, R104, P156, and N351 were exposed as functional residues (Table 3).

In the MAF results we have found that R607C, G342E, and N351S showed more frequency to occur than others.

Although the most of values of MAF indicates rare value to occur frequently, these values will help future in the various community.

In this study, Musite and PROSPER tools were applied for the post-PTM sites. The analyzed data indicated proteolytic cleavage sites at R607, W282, and P156 residues; hydroxylation sites were present at P303, P533, P131, and P156 residues; methylation sites were situated at R607, R104, and R596 residues; O-linked and N-linked Glycosylation sites were at T192 and N351 residues, respectively. From these analyzed data, it is evident that among the 15 functionally significant nsSNPs, both methylation and proteolytic cleavage sites were predicted to be at R607 residue; and both hydroxylation and proteolytic sites were found at P156 residue (Table 3). Therefore, these 2 mutations might significantly affect PTM of the human *SLC6A4* gene product.

COACH-analyzed data indicated that the T192 and W282 residues of the human *SLC6A4* gene product were involved in the interactions of the ligand-binding site. The ligand that can bind to the T192 and W282 sites can affect the structural conformation or functional consequences. The outputs of the NCBI-conserved domain search tool showed that the R104C, P131L, P156L, and I161T variants were present in sodium ion (Na+) binding site 1 and the R104C, P131L, P156L, I161T, T192M, I270T, G342E, F377S, W282S, P303, and N3351S variants were located in Na+ binding site 2. The human *SLC6A4* gene product is a sodium-dependent serotonin transporter, so a mutation in the Na+ binding site might interfere with serotonin transporter activity. Taken together with the result of MutPred, ConSurf, Musite, PROSPER, COACH, and NCBI conserved domain search tools, we have selected 8 nsSNPs out of 15 for further structural analysis.

For checking the effects of the mutant variants of the human *SLC6A4* gene on the protein structure and binding interactions, the 3D protein models of the 8 variants: T192M, G342E, R607C, W282S, R104C, P131L, P156L, and N351S were generated and validated using SPARKS-X, and Verify3D and then further analyzed using PROCHECK. All of these mutants had almost the same TM score, which means that their topological similarity is high with the wild-type protein. But in the case of RMSD value, R607C had the highest deviation, and P131L had the lowest. Verify 3D analyzed data suggested that all the structures had around 80%; that is, almost all amino acids had scored ≥0.2 in the 3D and 1D Profile. PROCHECK results also suggested that all the mutant protein structures had 90% or above amino acid residues in the favorable region, and hence they were used for further analysis. Native 3D model of the SLC6A4 protein was retrieved from the Protein Data Bank (PDB id: 6VRH) compared with the mutant protein structures. Further structural study of the wild-type and mutant proteins predicted alterations in H-bonding interactions in T192M, G342E, R607C, W282S, R104C, P131L, P156L, and N351S. The alteration in hydrogen bonds

might cause structural instability, which in turn might cause defects in protein function.

Summarizing all the results of this study, we identified that T192M, G342E, R607C, W282S, R104C, P131L, P156L, and N351S variants of the human *SLC6A4* gene were the most deleterious, pathogenic, and functionally significant nsSNPs in the humans (Figure 4). This in-depth *in silico* structure-function study suggested that these damaging nsSNPs of the *SLC6A4* gene have the potential to be explored as important biomarkers for serotonin-related mental disorders in the future. However, more studies and further experimental validation are needed to confirm the role of *SLC6A4* SNPs in disease susceptibility.

Conclusion

The results of this study identified the most deleterious and risky polymorphisms of the *SLC6A4* gene and analyzed their encoded protein 3D structural alterations in association with the biological functions. We have figured out T192M, G342E, R607C, W282S, R104C, P131L, P156L, and N351S SNPs are the most deleterious SNPs and can reduce the protein stability of *SLC6A4*. The screened nsSNPs will provide deep insight for further exploring the *SLC6A4* gene as an effective biomarker for serotonin-related various mental disorders. Finally, this research can be a strong direction for understanding the molecular basis of serotonin-related disorders and promote more accessible wet-laboratory studies.

Author Contributions

This work is a product of the intellectual effort of the whole team and that all members have contributed in various degrees to the analytical methods used, PROVEAN the research concept, to the experiment design, and the manuscript preparation.

ORCID iDs

Md Nasir Uddin https://orcid.org/0000-0002-2172-8872

Lolo Wal Marzan https://orcid.org/0000-0002-7426-9841

REFERENCES

1. Murphy DL, Fox MA, Timpano KR, et al. How the serotonin story is being rewritten by new gene-based discoveries principally related to SLC6A4, the serotonin transporter gene, which functions to influence all cellular serotonin systems. *Neuropharmacology*. 2008;55:932-960. doi:10.1016/j.neuropharm.2008.08.034.

2. Wendland JR, Moya PR, Kruse MR, et al. A novel, putative gain-of-function haplotype at SLC6A4 associates with obsessive-compulsive disorder. *Hum Mol Genet*. 2008;17:717-723. doi:10.1093/hmg/ddm343.

3. Alaerts M, Ceulemans S, Forero D, et al. Detailed analysis of the serotonin transporter gene (SLC6A4) shows no association with bipolar disorder in the Northern Swedish population. *Am J Med Genet B Neuropsychiatr Genet*. 2009;150:585-592. doi:10.1002/ajmg.b.30853.

4. Gelernter J, Pakstis AJ, Kidd KK. Linkage mapping of serotonin transporter protein gene SLC6A4 on chromosome 17. *Hum Genet*. 1995;95:677-680. doi:10.1007/BF00209486.

5. Serretti A, Calati R, Mandelli L, De Ronchi D. Serotonin transporter gene variants and behavior: a comprehensive review. *Curr Drug Targets*. 2006;7:1659-1669. doi:10.2174/138945006779025419.

6. Dabhi B, Mistry KN. In silico analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. *Meta Gene*. 2014;2:586-595.

7. Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309:1559-1563.

8. Datta A, Mazumder MH, Chowdhury AS, Hasan MA. Functional and structural consequences of damaging single nucleotide polymorphisms in human prostate cancer predisposition gene RNASEL. *BioMed Res Int*. 2015;2015:271458.

9. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894-3900. doi:10.1093/nar/gkf493.

10. Barroso I, Gurnell M, Crowley VEF, et al. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999;402:880-883. doi:10.1038/47254.

11. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol*. 2001;307:683-706. doi:10.1006/jmbi.2001.4510.

12. Lander ES. The new genomics: global views of biology. *Science*. 1996;274:536-539. doi:10.1126/science.274.5287.536.

13. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *Obstet Gynecol Surv*. 1995;50:201-204. doi:10.1097/00006254-199503000-00021.

14. Desai M, Chauhan J. In silico analysis of nsSNPs in human methyl CpG binding protein 2. *Meta Gene*. 2016;10:1-7.

15. Desai M, Chauhan JB. Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. *Comput Biol Chem*. 2017;70:7-14. doi:10.1016/j.compbiolchem.2017.07.001.

16. Desai M, Chauhan JB. Computational analysis for the determination of deleterious nsSNPs in human MTHFR gene. *Comput Biol Chem*. 2018;74:20-30.

17. Desai M, Chauhan JB. Predicting the functional and structural consequences of nsSNPs in human methionine synthase gene using computational tools. *Syst Biol Reprod Med*. 2019;65:288-300.

18. Patel SM, Koringa PG, Reddy BB, Nathani NM, Joshi CG. In silico analysis of consequences of non-synonymous SNPs of Slc11a2 gene in Indian bovines. *Genom Data*. 2015;5:72-79.

19. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010;38:W529-W533. doi:10.1093/nar/gkq399.

20. Bhatnager R, Dang AS. Comprehensive in-silico prediction of damage associated SNPs in human prolidase gene. *Sci Rep*. 2018;8:9430. doi:10.1038/s41598-018-27789-0.

21. Gao J, Thelen JJ, Dunker AK, Xu D. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol Cell Proteomics*. 2010;9:2586-2600. doi:10.1074/mcp.M110.001388.

22. Song J, Tan H, Perry AJ, et al. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS ONE*. 2012;7:e50300. doi:10.1371/journal.pone.0050300.

23. Badgujar NV, Tarapara BV, Shah FD. Computational analysis of high-risk SNPs in human CHK2 gene responsible for hereditary breast cancer: a functional and structural impact. *PLoS ONE*. 2019;14:e0220711.

24. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein-structure coordinates. *Int Tab for Crystallography*. 2012;21:684-687. doi:10.1107/97809553602060000882.

25. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-1612. doi:10.1002/jcc.20084.

26. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-1081. doi:10.1038/nprot.2009.86.

27. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40:W452-W457. doi:10.1093/nar/gks539.

28. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812-3814. doi:10.1093/nar/gkg509.

29. Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian SV. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res*. 2006;34:1317-1325. doi:10.1093/nar/gkj518.

30. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat*. 2008;29:1342-1354. doi:10.1002/humu.20896.

31. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*. 2012;7:e46688. doi:10.1371/journal.pone.0046688.

32. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;76:7-20.

33. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*. 2016;32:2230-2232.

34. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006;22:2729-2734. doi:10.1093/bioinformatics/btl423.

35. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat*. 2009;30:1237-1244. doi:10.1002/humu.21047.

36. Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, De La Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 2005;21:3176-3178.

37. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*. 2006;62:1125-1132. doi:10.1002/prot.20810.

38. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005;33:W306-W310. doi:10.1093/nar/gki375.

39. Akhtar M, Jamal T, Jamal H, et al. Identification of most damaging nsSNPs in human CCR6 gene: in silico analyses. *Int J Immunogenet*. 2019;46:459-471.

40. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009;25:2744-2750. doi:10.1093/bioinformatics/btp528.

41. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*. 2017:134981. https://www.biorxiv.org/content/10.1101/134981v1.

42. Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*. 2002;18:S71-S77. doi:10.1093/bioinformatics/18.suppl_1.S71.

43. Williamson K, Schneider V, Jordan RA, Mueller JE, Henderson Pozzi M, Bryk M. Catalytic and functional roles of conserved amino acids in the SET domain of the S. Cerevisiae lysine methyltransferase Set1. *PLoS ONE*. 2013;8:e57974.

44. Sidore C, Busonero F, Maschio A, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat Genet*. 2015;47:1272-1281.

45. Wang D, Liu D, Yuchi J, et al. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res*. 2020;48:W140-W146.

46. Hunter S, Apweiler R, Attwood TK, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009;37:D211-D215. doi:10.1093/nar/gkn785.

47. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015;43:D222-D226.

48. Yang Y, Faraggi E, Zhao H, Zhou Y. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*. 2011;27:2076-2082. doi:10.1093/bioinformatics/btr350.

49. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*. 2001;10:1470-1473.

50. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33:2302-2309. doi:10.1093/nar/gki524.

51. Yang J, Roy A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 2013;29:2588-2595.

52. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet*. 1997;13:163. doi:10.1016/s0168-9525(97)01103-7.

53. Ramamoorthy S, Bauman AL, Moore KR, et al. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *Proc Natl Acad Sci USA*. 1993;90:2542-2546. doi:10.1073/pnas.90.6.2542.

54. Ramoz N, Reichert JG, Corwin TE, et al. Lack of evidence for association of the serotonin transporter gene SLC6A4 with autism. *Biol Psychiatry*. 2006;60:186-191. doi:10.1016/j.biopsych.2006.01.009.

55. Sutcliffe JS, Delahanty RJ, Prasad HC, et al. Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. *Am J Hum Genet*. 2005;77:265-279. doi:10.1086/432648.

56. Kilic F, Murphy DL, Rudnick G. A human serotonin transporter mutation causes constitutive activation of transport activity. *Mol Pharmacol*. 2003;64:440-446. doi:10.1124/mol.64.2.440.

57. Taylor S. Molecular genetics of obsessive–compulsive disorder: a comprehensive meta-analysis of genetic association studies. *Mol Psychiatry*. 2013;18:799-805.

58. Rao S, Leung CST, Lam MH, Wing YK, Waye MMY, Tsui SKW. Resequencing three candidate genes discovers seven potentially deleterious variants susceptibility to major depressive disorder and suicide attempts in Chinese. *Gene*. 2017;603:34-41.

59. Gadow KD, DeVincent CJ, Siegal VI, et al. Allele-specific associations of 5-HTTLPR/rs25531 with ADHD and autism spectrum disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;40:292-297.

60. Kohen R, Jarrett ME, Cain KC, et al. The serotonin transporter polymorphism rs25531 is associated with irritable bowel syndrome. *Dig Dis Sci*. 2009;54:2663-2670. doi:10.1007/s10620-008-0666-3.

61. Pallesen S, Jacobsen DP, Nielsen MB, Gjerstad J. The 5-HTTLPR rs25531 LALA-genotype increases the risk of insomnia symptoms among shift workers. *Sleep Med*. 2019;60:224-229.

62. Schneider I, Kugel H, Redlich R, et al. Association of serotonin transporter gene AluJb methylation with major depression, amygdala responsiveness, 5-HTTLPR/rs25531 polymorphism, and stress. *Neuropsychopharmacology*. 2018;43:1308-1316.

63. Wang Z, Baker DG, Harrer J, Hamner M, Price M, Amstadter A. The relationship between combat-related posttraumatic stress disorder and the 5-HTTLPR/rs25531 polymorphism. *Depress Anxiety*. 2011;28:1067-1073. doi:10.1002/da.20872.