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

Md Arzo Mia1, Md Nasir Uddin1*, Yasmin Akter1, Jesmin2 and Lolo Wal Marzan1

1Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong, Bangladesh. 2Department 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, Mjpro, and MufPred2; 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, 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

RECEIVED: January 24, 2022. ACCEPTE: May 6, 2022.
TYPE: Original Research
FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lolo Wal Marzan, Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong 4331, Bangladesh. Email: marzan.geb@cu.ac.bd

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.9,9 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.14,15

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-Var 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.

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/hom/open-access-at-sage).
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.\textsuperscript{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.\textsuperscript{20} Conservation of the amino acid residues was predicted using ConSurf.\textsuperscript{19} We also investigated the posttranslational modification (PTM) sites in the human SLC6A4 protein using Musite and PROSITE.\textsuperscript{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.\textsuperscript{16,23-25} Together, this study conducted a thorough, in-depth computational analysis on all the nsSNPs of the \textit{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).

\textbf{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.\textsuperscript{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.\textsuperscript{28} The rs IDs of \textit{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.\textsuperscript{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.\textsuperscript{30} The input query was the FASTA format of the SLC6A4 protein sequence.

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.\textsuperscript{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. 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. 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. The input key was the SLC6A4 protein sequence and amino acid substitution.

Predicator 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. 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 (SNP&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. The input query was the UniProt accession number of the SLC6A4 protein (P31645) and its amino acids substitute variants.

PMut (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. The input query was the FASTA format of the SLC6A4 protein sequence and variations.

MUplo (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. 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. 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. 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. 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. 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. 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, pyridoline carboxylic acid, palmitoylation, and hydroxylation is identified by the MusiteDeep server. 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. 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.66,67 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://services.mbi.ucla.edu/Verify3D) and PROCHECK (https://services.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.ccmh.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). 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).
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 SLC6sb-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        | Dec   | Dec      |
| rs28914832   | I425V     | Tol  | B          | N     | N       | ProD    | C25        | N     | 8       | N        | Dec   | Dec      |
| rs138004662  | V283L     | Tol  | B          | E     | N       | ProD    | C25        | Dis   | N       | 7        | Dec   | Dec      |
| rs6355       | G56A      | Tol  | B          | E     | N       | ProB    | C55        | N     | 8       | N        | Dec   | Dec      |
| rs2228673    | K201N     | Tol  | B          | N     | N       | PosD    | C65        | N     | 6       | N        | Dec   | Dec      |
| rs28914833   | F465L     | Tol  | B          | E     | Del     | ProD    | C15        | N     | 6       | N        | Dec   | Dec      |
| rs28914834   | L550V     | Tol  | B          | N     | N       | PosD    | C25        | N     | 3       | N        | Dec   | Dec      |
| rs55848249   | D193N     | Tol  | B          | N     | N       | PosD    | C15        | N     | 3       | N        | Dec   | Dec      |
| rs55908511   | V488M     | Tol  | B          | N     | N       | PosD    | C15        | N     | 1       | N        | Inc   | Inc      |
| rs56316081   | I108V     | Tol  | B          | N     | N       | PosD    | C25        | N     | 3       | N        | Inc   | Inc      |
| rs60067068   | G41A      | Tol  | B          | N     | N       | PosD    | C55        | N     | 7       | N        | Dec   | Dec      |
| rs74330808   | I270T     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | 0       | N        | Dec   | Dec      |
| rs75354642   | P601S     | Tol  | PD         | N     | Del     | ProD    | C65        | Dis   | 2       | N        | Dec   | Dec      |
| rs11263079   | M370V     | Tol  | B          | N     | N       | PosD    | C15        | N     | 2       | N        | Dec   | Dec      |
| rs117750329  | V236I     | Tol  | PD         | N     | N       | PosD    | C25        | Dis   | N       | 6        | Dec   | Dec      |
| rs140206260  | I179V     | Tol  | PD         | N     | N       | PosD    | C25        | N     | 6       | N        | Dec   | Dec      |
| rs14036169   | G41R      | Del  | PD         | N     | N       | PosD    | C65        | N     | 5       | N        | Dec   | Dec      |
| rs14048986   | A65V      | Tol  | B          | N     | N       | PosD    | C55        | N     | 5       | N        | Inc   | Dec      |
| rs142701015  | I266V     | Tol  | B          | N     | N       | PosB    | C25        | N     | 7       | N        | Dec   | Dec      |
| rs14244982   | R626H     | Tol  | B          | N     | N       | PosD    | C25        | N     | 5       | N        | Dec   | Dec      |
| rs142505940  | R241L     | Tol  | B          | E     | Del     | PosD    | C65        | N     | 5       | N        | Dec   | Inc      |
| rs143632225  | A419V     | Tol  | PD         | N     | Del     | PosD    | C55        | N     | 7       | N        | Dec   | Inc      |
| rs144427337  | T192M     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | 5       | N        | Dec   | Inc      |
| rs145643221  | I157V     | Tol  | PD         | N     | N       | PosD    | C25        | Dis   | 2       | N        | Dec   | Dec      |

(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     | Inc      | Dec   |
| rs147306146| R144Q     | Tol  | PD         | E     | Del     | ProD    | C35        | N     | Dis     | Dec      | Dec   |
| rs184149069| N211S     | Tol  | B          | N     | N       | ProD    | C45        | N     | N       | 4        | N     | Dec      |
| rs190758123| V457I     | Tol  | B          | N     | N       | ProB    | C25        | N     | N       | 9        | N     | Dec      |
| rs191881524| L222P     | Tol  | -          | N     | N       | ProB    | C65        | N     | Dis     | N        | Dec   |
| rs199504888| E215K     | Tol  | B          | N     | N       | ProB    | C55        | N     | Dis     | 2        | N     | Dec      |
| rs199727635| G25R      | Del  | PD         | N     | N       | PosD    | C65        | N     | N       | 8        | N     | Dec      |
| rs199821523| D57E      | Tol  | B          | N     | N       | ProB    | C35        | N     | N       | 7        | N     | Dec      |
| rs199832478| I576V     | Tol  | B          | N     | N       | ProB    | C25        | N     | N       | 8        | N     | Dec      |
| rs199840777| I553V     | Tol  | B          | N     | N       | PosD    | C25        | N     | N       | 7        | N     | Dec      |
| rs199873504| T421I     | Tol  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 3        | N     | Dec      |
| rs199876253| G342E     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 9        | Dis   | Inc      |
| rs199886280| M389I     | Tol  | B          | N     | N       | ProD    | C0         | N     | Dis     | 4        | N     | Dec      |
| rs199890537| A449G     | Tol  | B          | N     | N       | ProD    | C55        | N     | N       | 0        | N     | Dec      |
| rs199909202| S293F     | Tol  | B          | N     | N       | ProB    | C65        | N     | Dis     | 4        | N     | Dec      |
| rs199913287| I266T     | Del  | B          | N     | N       | ProB    | C65        | N     | Dis     | 2        | N     | Dec      |
| rs200015551| R607C     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 2        | Dis   | Inc      |
| rs200078896| A72V      | Tol  | B          | N     | N       | ProB    | C55        | N     | N       | 2        | N     | Dec      |
| rs200080084| A505V     | Tol  | B          | N     | N       | ProB    | C55        | N     | N       | 0        | N     | Dec      |
| rs200126594| P621L     | Tol  | PD         | N     | Del     | PosD    | C65        | N     | N       | 4        | N     | Dec      |
| rs200180716| R464Q     | Tol  | B          | N     | N       | ProB    | C35        | N     | N       | 4        | N     | Dec      |
| rs200204643| A181V     | Del  | PD         | E     | Del     | ProD    | C55        | Dis   | Dis     | 5        | N     | Dec      |
| rs200263321| N45S      | Tol  | B          | N     | N       | ProD    | C45        | N     | N       | 3        | N     | Dec      |
| rs200339864| S48L      | Tol  | B          | N     | N       | PosD    | C65        | N     | N       | 4        | N     | Inc      |
| rs200341915| F377S     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | Dis     | 5        | Dis   | Dec      |
| rs200405588| I518F     | Tol  | B          | N     | N       | PosD    | C15        | Dis   | Dis     | 5        | N     | Dec      |

(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      |
| rs200548688| 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       | PosD    | C15        | N     | N 0     | N       | Dec   | Dec      |
| rs201041934| S62C      | Del  | 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       | PosD    | C45        | Dis   | N 0     | N       | Dec   | Dec      |
| rs201520306| I471S     | Tol  | B          | N     | N       | ProB    | C65        | 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      |
| 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     | Dec     | Dec   | Dec      |
| rs201833332  | I161T     | Del        | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | Dec     | Dec   | Dec      |
| rs201940331  | P156L     | Del        | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | Dec     | Dec   | Dec      |
| rs202152288  | V50L      | Tol        | B          | N     | N       | PosD    | C25        | N     | N 6     | Dec     | Dec   | Dec      |
| rs202166264  | T497A     | Tol        | B          | N     | N       | PosD    | C55        | N     | N 8     | Dec     | Dec   | Dec      |
| rs202181933  | I609T     | Tol        | B          | N     | N       | ProB    | C65        | N     | N 6     | Dec     | Dec   | Dec      |
| rs371274847  | V363A     | Tol        | B          | N     | Del     | PosD    | C55        | N     | N 3     | Dec     | Dec   | Dec      |
| rs372056901  | G35E      | Tol        | B          | N     | N       | PosD    | C65        | N     | N 4     | Dec     | Dec   | Dec      |
| rs374144565  | G530S     | Del        | PD         | E     | Del     | ProD    | C55        | Dis   | Dis     | Dec     | Dec   | Dec      |
| rs374583307  | A401T     | Tol        | B          | N     | N       | ProB    | C55        | N     | N 0     | Dec     | Dec   | Dec      |
| rs375503605  | V26A      | Tol        | B          | N     | N       | ProB    | C55        | N     | N 8     | Dec     | Dec   | Dec      |
| rs375913512  | P617L     | Del        | B          | N     | N       | ProB    | C65        | N     | N 3     | Dec     | Dec   | Inc      |

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 polymorphisms; SNPs&GO, single nucleotide polymorphism and gene ontology; Tol, tolerated.
### Table 2. Concurrency 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      |
| rs144427337     | T192M     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | Dis     | 5       | N     | Dec      |
| rs199876253     | G342E     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 9       | Dis   | Inc      |
| rs200015551     | R607C     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 2       | Dis   | Dec      |
| rs200341915     | F377S     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | Dis     | 5       | Dis   | Dec      |
| rs200435184     | W282S     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 2       | Dis   | Dec      |
| rs200924626     | P303H     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 4       | Dis   | Dec      |
| rs200953188     | R104C     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 9       | Dis   | Dec      |
| rs201114547     | P533L     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 7       | Dis   | Dec      |
| rs201480140     | P131L     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 9       | Dis   | Dec      |
| rs201833332     | I161T     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 6       | N     | Dec      |
| rs201940331     | P156L     | Del  | PD         | E     | Del     | ProD    | C65        | Dis   | Dis     | 7       | Dis   | Inc      |
| rs201481838     | R596W     | Del  | PD         | E     | Del     | PosD    | C65        | Dis   | N       | 5       | Dis   | Dec      |
| rs201688096     | N351S     | Del  | PD         | E     | Del     | ProD    | C45        | Dis   | Dis     | 7       | Dis   | Dec      |
| rs374144586     | G530S     | Del  | PD         | E     | Del     | ProD    | C55        | Dis   | Dis     | 7       | Dis   | 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 polymorphisms; SNPs&GO, single nucleotide polymorphism and gene ontology.
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...
Bioinformatics and Biology Insights

The G56A substitution in exon 2 of the SLC6A4 gene has a prominent association with autism. An I425V variation in exon 9, A1438G, and T102C SNPs of the SLC6A4 gene were reported to be related to OCD.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.

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 N351S as the highest conserved 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 N351S as the highest conserved 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 N351S as the highest conserved 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.
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. 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.

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

Supplemental Material
Supplemental material for this article is available online.

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. Neuropsychopharmacology. 2008;33:932-960. doi:10.1016/j.neurnpharm.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/ddn343.
3. Alalrert J, 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. Gelentener 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.
58. Rao S, Leung CST, Lam MH, Wing YK, Wai MY, 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.