STM JOURNALS

Research & Reviews: A Journal of Life Sciences
ISSN: 2249-8656 (Online), ISSN: 2348-9545 (Print)
Volume 8, Issue 3
www.stmjournals.com

Comprehensive In silico Analysis of Single Nucleotide Polymorphisms (SNPs) in Human BDNF Gene

Shovit Ranjan, Praveen Kumar Sharma*
Centre for Life Sciences, Central University of Jharkhand, Ranchi, Jharkhand, India

Abstract
Since a number of single nucleotide polymorphisms (SNPs) are associated with the BDNF gene mutations involved in different diseases. Hence, it is important to find out the probable functional SNPs before planning a larger population study. So, to look out for the functional nsSNPs (non-synonymous single nucleotide polymorphisms) in BDNF gene, existing data present in the dbSNP database and different bioinformatics tools like SIFT, PolyPhen, nsSNPAnalyzer, F-SNP, I-mutant, Pymol were used for the analysis. The results showed that out of the total 2929 SNPs, 58 were nsSNPs (non-synonymous single nucleotide polymorphisms), 44 occurred in the mRNA 3' UTR, 108 occurred in 5' UTR region, 257 occurred in intronic regions and the rest were other types of SNPs. SIFT and PolyPhen programs predicted 3 out of 58 nsSNPs (rs8192466, rs1048218, and rs77787410) as not tolerable, deleterious and damaging. F-SNP revealed that the rs1048221, rs1048220, rs1048218, rs6265 and rs8192466 SNP in the non-synonymous coding region may have protein-coding and splicing regulator functions. PDBSum and UniProtKB predicted the number of protein structures, sharing 100% similarity with the BDNF amino acid sequence. Moreover, I-Mutant and nsSNPAnalyzer showed a decrease in stability for these nsSNPs upon mutation. Protein structural analysis with these amino acid variants was performed by using I-Mutant and PyMOL. This study suggested that three nsSNPs, rs8192466, rs1048218, and rs77787410 were identified as deleterious, thus destabilizing the protein stability, amino acid interactions, and hydrogen bond networks of the protein directly or indirectly. Hence, these SNPs can explain the functional deviation of protein to some extent.

Keywords: *BDNF, F-SNP, I-Mutant, nsSNPAnalyzer, PolyPhen, SIFT, SNP*

****Author for Correspondence*** E-mail: pksharma.cuj@gmail.com

INTRODUCTION

A single nucleotide polymorphism (SNP) is the simplest and most common type of genetic polymorphism in the human genome (since it covers 90% of all human DNA polymorphisms). It refers to a single base mutation in DNA. There are several publically available databases for SNPs, such as dbSNP, GWAS Central and SwissVar, out of which dbSNP is the most extensive one. With the increase in a number of human SNPs by more than 50 million, including both synonymous and non-synonymous SNPs [1], only the non-synonymous SNPs (nsSNPs), also called as missense variants are particularly important as they result in the change of translated amino acid residue sequence. Since it is known that nsSNPs play a major role in the functional diversity of coded proteins in human populations and have been linked with many diseases. Therefore, they may affect the protein function by reducing protein solubility or destabilizing protein structure and may also affect gene function by altering transcription and translation [2-5].

Brain-derived neurotrophic factor (BDNF) a member of the neurotrophic factor and encoded by a gene located on human chromosome 11p13 [6] plays important role in the growth, development, differentiation, and regeneration of various types of neurons in the central nervous system [7]. Earlier studies have already shown reduced levels of BDNF accompanies impaired glucose metabolism in T2D patients [8], and a low level of circulating BDNF in obese patients [9]. Moreover, another study has also shown decreased BDNF mRNA expression and protein in the substantia nigra of Parkinson's Disease (PD) patients [10, 11]; and reduced mRNA expression of BDNF in the hippocampus of Alzheimer's Disease (AD)

patients [12]. Several investigators have also demonstrated that decreased BDNF and its receptor level, tyrosine receptor kinase B level in the frontal cortex and hippocampus of AD patients [13]. Thus, making BDNF, an important candidate gene for Type 2 Diabetes (T2D), Obesity, PD and AD risk. There are many single nucleotide polymorphisms (SNPs) in and around the BDNF gene, found associated with different diseases.

Several polymorphisms in the upstream regulatory region of the BDNF gene is reported to be associated with different diseases like T2D, Obesity, AD, and PD. Among these, few important ones are rs6265, rs925946, rs4074134, rs4923461 in T2D and obesity; rs11030104, rs16917204, rs7103411, rs6265, rs2030324 in AD; and rs6265, rs56164415 in PD. BDNF seems a plausible target for T2D, obesity, AD, PD, and other neurodegenerative disorders studies. Hence, it is important to sort out the possible functional SNPs before planning a larger population study. For accomplishing this goal, data available in the dbSNP database and different bioinformatics tools were used. However, knowledge about the clinical relevance for many of the BDNF SNPs is still limited. Taking into account all these considerations and the central role played by BDNF in many diseases, this study was undertaken to determine the effect of various nsSNPs on the protein structure, stability and function of the BDNF gene, that may have an important role in disease susceptibility.

MATERIAL AND METHODS

Datasets

The data of the human BDNF gene was collected from the National Center for Biological Information (NCBI) database. The information about BDNF gene SNPs and their protein sequences in the FASTA format were retrieved from the dbSNP database [14-16] and ‘‘.pdb’’ files for BDNF subunits were retrieved from the RCSB Protein Data Base [17, 18] for performing the computational analysis.

Sequence Homology-Based Prediction of Deleterious nsSNPs by using SIFT

The Sorting Intolerant from Tolerant (SIFT) is an online bioinformatics server that is used to predict the tolerated or deleterious coding non-synonymous SNPs [19]. The SIFT program takes a query sequence and uses multiple sequence alignment to sort out the functionally neutral and deleterious amino acid changes due to SNPs in the coding regions of genes [20]. For this, it utilizes amino acid sequence homology and the physical properties of the proteins in combination with naturally occurring nsSNPs by aligning paralogous and orthologous protein sequences. The algorithm for the SIFT program use the default settings (UniProt-TrEMBL 39.6 database, median conservation of sequences of 3.00, and allowance to remove sequences more than 90% identical to query sequence).

SIFT is a multistep procedure that after giving an input (in the form of dbSNP rs ID or protein RefSeq ID or GI number), searches for similar sequences, then chooses closely related sequences sharing similar function, then obtains the multiple alignments of these chosen sequences, and then calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a tolerance index of 0.05 are predicted to be intolerant or deleterious; those greater than or equal to 0.05 are predicted to be tolerated [21, 22]. In the current study, the dbSNP rsID of all nsSNPs of human BDNF gene filtered from the dbSNP database were analyzed.

Structural Homology-Based Prediction of Functional Consequences of coding nsSNPs using PolyPhen-2

The Polymorphism Phenotyping (PolyPhen-2) is an online bioinformatics server that is used to predict the influence of nsSNPs or amino acid substitution on the structure and function of proteins by using the specific empirical rules [23-25]. The PolyPhen-2 server requires the input as protein sequence in FASTA format, protein/SNP identifier, database ID/accession number, amino acid position and amino acid variant details.PolyPhen-2 categorizes the SNPs into different classes as ‘‘benign,’’ ‘‘possibly damaging’’ or ‘‘probably damaging’’ based on site-specific sequence conservation among mammals, as well as their location in the three-dimensional structure of the protein molecule. The term ‘‘damaging’’ used by PolyPhen-2 reflects the mutations altering the protein structure and not the loss or gain of protein function [26].

In the current study, the protein identifier from the UniProt database for the BDNF protein ‘‘P23560’’ was submitted with the position of variation along with the wild-type and mutant amino acids. Based on this, PolyPhen-2 estimates the position-specific independent count (PSIC) score for every variant and calculates the score difference between variants. The PSIC score difference between the two variants tells the number of functional consequences exerted by that nsSNP. The PSIC score difference thus may be considered as directly proportional to the impact of a particular amino acid substitution [27].

nsSNPAnalyzer

Non-Synonymous SNP Analyzer (nsSNP Analyzer) is an online bioinformatics tool that is used to classify the nsSNPs by using a machine learning method called Random Forest. It uses a curated SNP dataset prepared from the SwissProt database. By using the input, this tool calculates the structural environment of the SNP, including the solvent accessibility, environmental polarity and secondary structure [28]; the normalized probability of the substitution in the multiple sequence alignment [21]; and the similarity-dissimilarity between the original and mutated amino acid.

The protein sequence in FASTA format and a substitution file denoting the SNP identities to be analyzed serves as an input for this program. The file format of Substitution file is denoted as:
X#Y; where X is the original amino acid in one letter, # is the position of the substitution (starting from 1) and Y is the mutated amino acid in one letter. New line characters should separate the number of multiple substitutions.

Prediction of functional SNPs for disease association studies by using F-SNP

SNPs in the coding, intronic and UTR sites regulate the gene expression in many ways like RNA transcript splicing site or transcription factor binding site alteration [29-30]. Hence, all these sites were also analyzed for their functional SNPs. F-SNP examines SNPs for deleterious effects with respect to each functional category like protein coding, splicing regulation, transcriptional regulation, and post-translation. For each category, a series of tests are executed to determine whether the SNP has a functional impact or not [31-33].

F-SNP database requires the input either in the form of candidate gene name using the gene symbol; single `dbSNP rs ID’’; or chromosomal region. The input of the candidate gene symbol (BDNF) was used for the analysis. The SNP prioritization result was a list of SNPs with its Functional Significance (FS) Score and possible functional effects.

Modeling of Amino Acid Substitution Effects due to nsSNPs on Protein Structure

The Closest Related Protein Structures

The EMBL-EBI Web-based tool PDBsum and UniProtKB was used to find the proteins similar to the BDNF gene. PDBsum and UniProtKB provide the collection of every information about the macromolecular structure deposited in the Protein Data Bank (PDB) [34, 35]. It performs a FASTA search against all sequences in the protein data bank (PDB) to obtain a list of the closest matches. The PDB code, UniProt id or FASTA sequence of the BDNF protein was provided in the query space.

Models of Substituted Amino Acids

PyMOL (version 1.8.2.0) was used to generate the mutated models of each of the selected PDB entries for the corresponding amino acid substitutions. PyMOL Software allows browsing through a rotamer library to change amino acids. A ‘‘Mutagenesis” option under Wizard Tab was used to replace the native amino acid with a new one. The mutagenesis option facilitates the replacement of the native amino acid by the ‘‘best’’ rotamer of the new amino acid. The native and mutant models were generated and saved in the .png format.

Identification of Cis-Regulatory Elements

PROSCAN (version1.7)

PROSCAN version 1.7 WWW Promoter Scan Service [36] predicts promoter regions based on scoring homologies with putative eukaryotic Pol II promoter sequences [37]. The analysis is done by using the suite of programs available at the site, which is serviced and maintained by Dr. Dan Prestridge at the Advanced Biosciences Computing Center, University of

Minnesota. It requires input in the form of Nucleic Acid Sequences.

Softberry TSSG

Softberry TSSG [38] is the most accurate mammalian promoter prediction program, that helps in the recognition of human Pol II promoter regions and transcription start sites [39]. It requires input in the form of nucleotide sequences.

Promoter 2.0

Promoter 2.0 Prediction Server [40] predicts transcription start sites of vertebrate Pol II promoters in DNA sequences [41]. It has been developed as an evolution of simulated transcription factors that interact with sequences in promoter regions. It builds on principles that are common to neural networks and genetic algorithms. The site is serviced and maintained by Steen Knudsen at the Center for Biological Sequence Analysis, Technical University of Denmark.

Modeling of nsSNP Locations on Protein Structure and Stability

Prediction of Change in Protein Structure Stability due to Mutation

For predicting the change in the stability of the protein due to mutation, a support vector machine (SVM)-based tool server, I-Mutant (version 2.0) was used. I-Mutant is a neural network based tool that helps in the routine analysis of protein stability changes by taking into account the single point mutations. The prediction of the mutational effect on protein stability can be done by providing input in the form of protein structure or sequence. I-Mutant can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related Gibbs-free energy change ($\Delta\Delta G$) [42].

I-Mutant provides the scores for the Gibbs-free energy change, calculated with the FOLD-X energy based web server. FOLD-X is a computer algorithm for quantitative estimation of interactions facilitating the protein stability and thus, it provides the comparison between wild-type and mutant models in the form of van der Waals clashes, which greatly influence the energy decomposition [43, 44].

Scanning of nsSNPs for their position in different protein domains and Modeling of mutant structure

The Prosit-ExPaSy tool [45] was used to find the nsSNPs and the amino acid changes associated with it that results in the change in different domains of the protein structures [46]. The input was provided in the form of UniProtKB ID in the query column, and then, the UniProt database was searched for motifs and domains of BDNF. Then, the normal and mutant structure was modeled and high-quality 3D images were generated by using PyMOL software (version 1.8.2.0). PyMOL is an open-source molecular visualization program that is used for visualizing and comparing the high-quality 3D images of small molecules and biological macromolecules such as proteins [47, 48].

RESULTS AND DISCUSSION

Nowadays, determining deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) for induction of disease-associated phenomena in the target candidate genes is the most important field of computational genomic research. In this investigation, we have applied multiple computational methods and tools to identify the most likely pathogenic point mutations in the BDNF gene along with its effect on the structure and function of the protein.

Collection of SNP Dataset from dbSNP Database

Polymorphism data for the BDNF gene were retrieved in this paper from the dbSNP-NCBI database that contains both validated and non-validated polymorphisms. In spite of this drawback, we preferred this database because it is the most extensive SNP database and allelic frequency of most of nsSNPs of BDNF has been recorded there (except 61 out of 119). It contained a total of 2929 SNPs in human (validated by frequency), out of which 58 (1.97%) were nsSNPs; 28 (0.96%) were sSNPs; 152 (5.19%) occurred in the mRNA 5’

© STM Journals 2018. All Rights Reserved

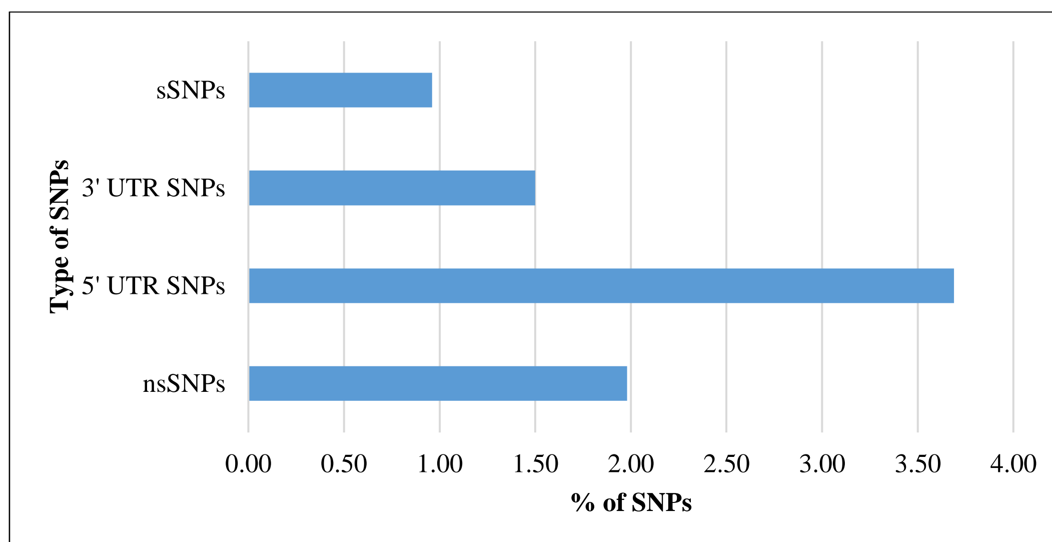


Fig. 1: *A graphical representation of the distribution of nonsynonymous (ns), 5'UTR and 3' UTR SNPs for the BDNF gene (based on the dbSNP database).*

and 3' UTRs and 2857 (9.54%) occurred in intronic regions. There are 108 SNPs (3.69%) in the 5' UTR and 44 SNPs (1.50%) in the 3' UTR separately. A graphical representation of the distribution of SNPs in the coding region and the UTRs are depicted in terms of percentage in Figure 1.

Deleterious nsSNPs predicted by SIFT

The sequence homology-based tool SIFT (Sorting Intolerant from Tolerant) determines the conservation level of a particular amino acid position in a protein. All the 58 nsSNPs were submitted as input to the SIFT server page independently to calculate the tolerance index (TI). The functional impact of the amino acid substitution is inversely proportional to the tolerance index, i.e. the lower the tolerance index, the higher the functional impact a particular amino acid residue substitution is likely to have and vice-versa. A tolerance index score of ≤ 0.05 is considered to be damaging. Table 1 summarizes the results obtained from the SIFT server. It was observed that out of the total 58 nsSNPs analyzed, 3 nsSNPs were identified to be deleterious with a tolerance index score of ≤ 0.05. Two nsSNPs (rs8192466 and rs77787410) showed a highly deleterious tolerance index score of 0.00. The remaining one deleterious nsSNP (rs1048218) showed tolerance index scores of 0.04. Three deleterious nsSNPs showed a nucleotide change from A/C/T, A/G/T and G/T respectively.

SIFT has been tested on many human SNP databases and was found to be suitable for distinguishing the disease-associated SNPs from a neutral one with only a 20% false positive error. The sensitivity of SIFT is confirmed by the subset of nsSNP from dbSNP, that predicts the affected function associated with the disease. Furthermore, the SIFT algorithm works mainly on sequence for prediction, performing similarly to tools that use structure. 74% of nsSNPs identified by the SNP Consortium, were sufficiently similar to homologs in protein sequence databases for SIFT prediction. Hence, using SIFT is advantageous over other tools [20, 49].

Damaging nsSNPs predicted by PolyPhen-2 Server

The nsSNPs involved in structural modification was determined by the PolyPhen-2 (Polymorphism and Phenotype) program. PolyPhen-2 software predicts the possible impact of an amino acid substitution on the structure and function of a protein through specific empirical rules on the sequence. PolyPhen-2 identifies homologues of the input sequences via BLAST, then calculates position-specific independent count (PSIC) scores for every variant and estimates the difference between the variant scores, the difference of >0.339 is detrimental. The program carries out a BLAST query of a sequence against a protein structure database (PDB and PQS) for mapping of the

Table 1: *List of nsSNPs predicted as damaging or non-tolerant by SIFT.*

SNP ID	Nucleotide Change	Amino acid Change	Protein ID	Tolerance Index Score
rs8192466	A/C/T	T2I	NP_001137279	0.00
rs1048218	G/T	Q75H	NP_001137277	0.04
rs77787410	A/G/T	V170I	NP_001137277	0.00

substitution site to known protein 3-dimensional structures. PolyPhen-2 uses the DSSP database to obtain secondary structure and solvent accessible surface area for the mapped amino acid residues. For sequence-based characterization of the substitution site, PolyPhen-2 uses the TMHMM algorithm, Coils2 program and SignalP program to predict transmembrane, coiled-coil and signal peptide regions of the protein sequences. There are certain empirical rules applied to the sequences and the accuracy of that is approximately 82% with a chance of 8% false positive prediction [23].

The protein accession number of BDNF (P23560) and the amino acid substitutions corresponding to each of 58 nsSNPs were submitted separately to PolyPhen-2 server. Table 2 summarizes the results obtained from the PolyPhen-2 server. A position-specific independent count (PSIC) score difference was assigned using the categories 'probably damaging', 'possibly damaging', 'potentially damaging', 'borderline' and 'benign' having PSIC score difference of 2.00 or more, 1.40–1.90, 1.20–1.50, 1.00–1.20 and 0.00–0.90 respectively. So, a total of three nsSNPs (5.17%) were predicted as damaging, and the PSIC scores fell into the range of 0.90 to 1.00. Three nsSNPs predicted to be deleterious by the SIFT program were also predicted to be damaging by the PolyPhen-2 server. rs8192466 and rs77787410 had a SIFT TI of 0.00 and a PolyPhen-2 PSIC of 1.000 and 0.987 respectively. Moreover, rs1048218 had a SIFT TI of 0.04 and a PolyPhen-2 PSIC of 0.956. Therefore, it can be inferred from Table 1 that the results retrieved from the evolutionary-based approach SIFT correlated well with the results obtained from the structural based approach PolyPhen-2, suggesting that these three nsSNPs may disrupt both the protein function and structure. Hence the mutations occurring with these 3 nsSNPs (rs8192466, rs77787410, and rs1048218) would be of prime importance while manifesting itself in the diseases caused by the BDNF gene, according to SIFT and PolyPhen-2 results.

Table 2: *List of nsSNPs predicted to be significantly damaging by PolyPhen2.*

SNP ID	Nucleotide Change	Amino acid Change	PSIC SD
rs8192466	A/C/T	T2I	1.000
rs1048218	G/T	Q75H	0.956
rs77787410	A/G/T	V170I	0.987

nsSNPAnalyzer

nsSNPAnalyzer is a tool that helps in the prediction of the phenotypic effect of nsSNP and also helps to facilitate the interpretation of results about these SNP like the structural environment and multiple sequence alignment. nsSNPAnalyzer makes predictions about the SNP based on the information contained in the multiple sequence alignment and three-dimensional protein structure.

Three amino acid variants (T2I, Q75H, and V170I) were submitted to nsSNPAnalyzer tool along with protein sequence of BDNF in FASTA format. We could not find any significant result for the Q75H and T2I variant. The output of the final result of the nsSNPAnalyzer tool is shown in Table 3.

Functional SNPs for Disease Association Studies Predicted by F-SNP

The F-SNP (Functional Single Nucleotide Polymorphisms) database was used to predict the functionally important SNPs in the synonymous coding, non-synonymous coding, intronic, intergenic, 3' and 5' UTRs. It efficiently identifies the functional SNPs at the splicing, transcriptional, translational and post-translational level. F-SNP is unique in a feature that the prediction of functional effects of SNPs is always based on the most up-to-date information and provides integrated information about the functional effects of SNPs from 16 different bioinformatics tools

Table 3: *Output for nsSNPAnalyzer.*

Amino Acid Variant	Phenotype	Environment	Area buried	Frac Polar	Secondary Structure
T2I	Disease				
Q75H	Disease				
V170I	Neutral	B1S	0.493	0.240	S

and databases. Therefore, the F-SNP database also helps to identify and focus on SNPs with potential pathological effect on human health.

The F-SNP search was performed by querying the gene symbol (BDNF). Table 4 lists the SNPs in the coding, intronic and UTR region, which are predicted to be functionally significant by F-SNP. The SNP rs1048221, rs1048220, rs1048218 and rs6265 in the non-synonymous coding region may have splicing regulator and protein-coding functions. rs8192466 in the non-synonymous coding region may have splicing regulator, protein-coding functions and post-translation functions.

Table 4: *List of SNPs in coding, intronic and UTR (mRNA) predicted to be functionally significant by F-SNP. For this table, the functional effects have been assessed for the SNP, and the SNP was determined to have a potentially deleterious functional effect. FS Score means Functional Significance Score.*

SNP Id	Nucleotide Change	Region (Chromosomal Location)	Possible Functional Effects	FS Score
rs1048221	G/T	Non-synonymous Coding (Chr 11, 27636308)	Protein Coding Splicing Regulation	0.585
rs1048220	G/T	Non-synonymous Coding (Chr 11, 27636314)	Protein Coding Splicing Regulation	0.595
rs1048218	G/T	Non-synonymous Coding (Chr 11, 27636463)	Protein Coding Splicing Regulation	0.917
rs6265	G/A	Non-synonymous Coding (Chr 11, 27636492)	Protein Coding Splicing Regulation	0.650
rs8192466	C/T	Non-synonymous Coding (Chr 11, 27636683)	Protein Coding Splicing Regulation Post Translation	0.533

Modeling of amino acid substitution effects due to nsSNPs on protein structure

The closest related protein structures

By using the EMBL-EBI Web-based tool PDBsum and UniProtKB, the BDNF gene product related protein structures were searched. Fourteen related protein structures, namely Q5IS78, A0A0E3SU01, F7F914, A0A0D9SD87, G3S0L4, C7SSB6, Q9BFJ2, Q71BN9, E2J835, P23560-2, I2CTP4, P23560-3, P23560-4, and P23560-5 were found to share 100% amino acid sequence similarity (Table 5).

Table 5: *The available PDB structure for the BDNF gene with a similarity (100%) with BDNF FASTA sequence at UniProtKB.*

Protein Accession No	Length(aa)	Similarity
Q5IS78	247	100%
A0A0E3SU01	247	100%
F7F914	247	100%
A0A0D9SD87	247	100%
G3S0L4	247	100%
C7SSB6	247	100%
Q9BFJ2	247	100%
Q71BN9	247	100%
E2J835	247	100%
P23560-2	255	100%
I2CTP4	255	100%
P23560-3	262	100%
P23560-4	329	100%
P23560-5	276	100%

Models of Substituted Amino Acids

BDNF gene (PDB Id: 1B8M) were scanned manually to identify amino acid polymorphisms. 1B8M accounted for three nsSNPs: rs8192466 (Thr2Ile), rs1048218 (Gln75His) and rs77787410 (Val170Ile) (Table 6). All the functional nsSNPs predicted using the SIFT and PolyPhen tools and present in the three structures mentioned above were subjected to the PyMOL Mutagenesis tool. The models for each functional nsSNP were made

and visualized as a comparison using PyMOL (version 1.8.2.0) (Figure 2).

Analysis of cis-regulatory elements
PROSCAN: version 1.7
No promoter regions were predicted.

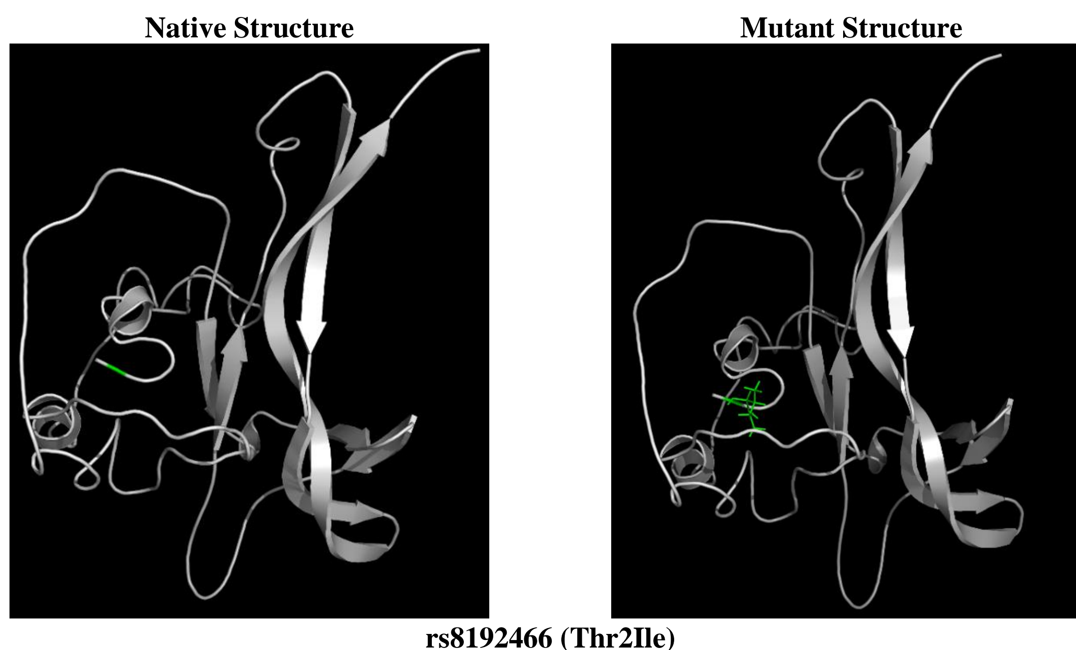
Softberry TSSG
1 promoter is predicted. The promoter position is 177 LDF (Threshold for LDF- 4.00). TATA box is predicted at 146 LDF.

Promoter 2.0 Prediction Server
No promoter predicted.

Modeling of nsSNP locations on protein structure and stability
Prediction of change in protein structure stability due to mutation
The I-Mutant 2.0 server is a neural network based routine tool used to predict the change in protein structure stability by considering the single-site mutation. The input option for this tool is the protein sequence or 3D structure. This tool was tested with the data extracted from ProTherm, which is the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability due to mutation and it results in the efficient prediction of the stability of the protein structure due to mutation. Depending on the usage of protein structural or sequence information, the prediction of the stability is 80% or 70% respectively. I- mutant also provides the scores of free energy change predictions, calculated with the FOLD-X energy based web server. By incorporating the FOLD-X estimations with those of I-Mutant, a precision of 93% can be achieved, thus making an I-Mutant tool helpful in protein design and mutation [42].

The three mutations (2, T → I; 75 Q → H; and 170 V → I) of BDNF gene have been selected based on the prediction scores by PolyPhen-2 server. These three mutations were given to I-Mutant server to predict the protein structural stability upon mutation. The results are summarized in Table 6.

Modeling of mutant structure
For mapping the deleterious nsSNPs into protein structure, the information was obtained from the dbSNP database. The available structure for the BDNF gene has a PDB id 1B8M. Three nsSNPs were found to be highly deleterious in nature among all the nsSNPs. Hence, we selected these nsSNPs for structural analysis and modeled the mutant structure. The mutational position and amino acid variant associated with this nsSNP as T → I at the residue position 2; Q → H at the residue position 75; and V → I at the residue position 170 were mapped using PyMOL software (version 1.8.2.0) to get modeled structure as shown in Figure 2.



rs8192466 (Thr2Ile)

Fig. 2: *A comparison of amino acid substitutions due to nsSNPs, rs8192466 (Thr2Ile), rs1048218 (Gln75His) and rs77787410 (Val170Ile). Figure shows the differences of structure and electron cloud density between native and mutant models BDNF Protein (PDB ID: 1B8M). Models were generated by using Pymol (v1.8.2.0).*

Table 6: *Prediction of Protein Structural Stability based on standard free energy change. For all the predictions, pH and Temperature were selected as 7.0 and 25 ºC respectively. WT: Amino acid in wild-type protein; NEW: New Amino acid in mutant protein; ΔΔG: ΔG(New Protein)-ΔG(Wild Type) in Kcal/mol ($\Delta\Delta G < 0$: Decrease Stability; $\Delta\Delta G > 0$: Increase Stability); T: Temperature in Celsius degrees; pH: -log[H+].*

Mutation	Position	WT	New	pH	Temperature	Stability	DDG (Kcal/mol)
2: T-I	2	T	I	7.0	25 ºC	Decrease	-0.21
75: Q-H	75	Q	H	7.0	25 ºC	Decrease	-1.10
170: V-I	170	V	I	7.0	25 ºC	Decrease	-0.36

CONCLUSION

The study suggested that three nsSNPs, rs8192466, rs1048218, and rs77787410 were identified as deleterious, thus destabilizing the protein stability, amino acid interactions, and hydrogen bond networks of the protein directly or indirectly. Hence, these SNPs can explain the functional deviation of protein to some extent.

ACKNOWLEDGMENTS

This work was partially supported by Department of Biotechnology, Government of India BUILDER project (BT/PR-9028/INF/22/193/2013) granted to Centre for Life Sciences, and UGC Central University Fellowship to S.R.

REFERENCES

1. Luu T-D, Rusu A-M, Walter V, et al. MSV3d: database of human MisSense Variants mapped to 3D protein structure. *Database*. 2012; 2012: bas018.
2. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *New England Journal of Medicine*. 1994; 331 (16): 1056-1061p.
3. Lander ES. The new genomics: global views of biology. *Science*. 1996; 274 (5287): 536p.
4. Barroso I, Gurnell M, Crowley V, et al. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999; 402 (6764): 880-883p.
5. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *Journal of molecular biology*. 2001; 307 (2): 683-706p.
6. Maisonpierre PC, Le Beau MM, Espinosa R, et al. Human and rat brain-derived neurotrophic factor and neurotrophin-3: gene structures, distributions, and chromosomal localizations. *Genomics*. 1991; 10 (3): 558-568.
7. Fahnestock M, Garzon D, Holsinger R, Michalski B. Neurotrophic factors and Alzheimer's disease: are we focusing on the wrong molecule? *Ageing and Dementia Current and Future Concepts*. 2002: 241-252p.
8. Krabbe K, Nielsen A, Krogh-Madsen R, et al. Brain-derived neurotrophic factor (BDNF) and type 2 diabetes. *Diabetologia*. 2007; 50 (2): 431-438p.
9. Oh K-J, Lee DS, Kim WK, Han BS, Lee SC, Bae K-H. Metabolic Adaptation in Obesity and Type II Diabetes: Myokines, Adipokines and Hepatokines. *International Journal of Molecular Sciences*. 2016; 18 (1): 8p.
10. Mogi M, Togari A, Kondo T, et al. Brain-derived growth factor and nerve growth factor concentrations are decreased in the substantia nigra in Parkinson's disease. *Neuroscience letters*. 1999; 270 (1): 45-48p.
11. Parain K, Murer MG, Yan Q, et al. Reduced expression of brain-derived neurotrophic factor protein in Parkinson's disease substantia nigra. *Neuroreport*. 1999; 10 (3): 557-561p.
12. Connor B, Young D, Yan Q, Faull R, Synek B, Dragunow M. Brain-derived neurotrophic factor is reduced in Alzheimer's disease. *Molecular Brain Research*. 1997; 49 (1): 71-81p.
13. Ferrer I, Marín C, Rey MJ, et al. BDNF and full-length and truncated TrkB expression in Alzheimer disease. Implications in therapeutic strategies. *Journal of Neuropathology & Experimental Neurology*. 1999; 58 (7): 729-739p.
14. Bhagwat M. Searching NCBI's dbSNP database. *Current Protocols in Bioinformatics*. 2010: 32:1.19. 1-19. 18.
15. Sherry ST, Ward M-H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*. 2001; 29 (1): 308-311p.
16. Upload NCBI. (2017) [online]. Available at: http://www.ncbi.nlm.nih.gov/snp/. [Accessed on January, 2017]
17. Dutta S, Burkhardt K, Bluhm WF, Berman HM. Using the tools and resources of the RCSB Protein Data Bank. *Current Protocols in Bioinformatics*. 2005.
18. Upload RCSB. (2017). Available from [online] http://www.rcsb.org/pdb/home/home.do [Accessed on January, 2017]

19. Upload PROVEAN. (2017). Available from [online] http://sift.jcvi.org [Accessed on January, 2017]
20. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*. 2003; 31 (13): 3812-3814p.
21. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research*. 2001; 11(5): 863-874p.
22. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet*. 2006; 7: 61-80p.
23. Upload PolyPhen-2 prediction of functional effects of human nsSNPs.(2017).Available from [online] http://genetics.bwh.harvard.edu/pph2/ (Accessed on January, 2017)
24. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic acids research*. 2002; 30 (17): 3894-3900p.
25. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010; 7 (4): 248-249p.
26. Boyko AR, Williamson SH, Indap AR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*. 2008; 4 (5): e1000083.
27. Rajasekaran R, Sudandiradoss C, Doss CGP, Sethumadhavan R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. *Genomics*. 2007; 90 (4): 447-452p.
28. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991; 164-170p.
29. Prokunina L, Castillejo-López C, Öberg F, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nature genetics*. 2002; 32 (4): 666-669p.
30. Prokunina L, Alarcón-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert reviews in molecular medicine*. 2004; 6 (10): 1-15p.
31. Available from [online] http://compbio.cs.queensu.ca/F-SNP/ (Accessed on January, 2017)
32. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic acids research*. 2008; 36(suppl 1): D820-D4p.
33. Lee PH, Shatkay H. An integrative scoring system for ranking SNPs by their potential deleterious effects. *Bioinformatics*. 2009; 25 (8): 1048-1055p.
34. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic acids research*. 2001; 29 (1): 221-222p.
35. Consortium U. UniProt: the universal protein knowledgebase. *Nucleic acids research*. 2017; 45(D1): D158-D169p.
36. Upload BioInformatics and Molecular Analysis Section. (2017). Available from [online] https://www-bimas.cit.nih.gov/molbio/proscan/ (Accessed on January, 2017)
37. Prestridge DS. Predicting Pol II promoter sequences using transcription factor binding sites. *Journal of molecular biology*. 1995; 249 (5): 923-932p.
38. Available from [online] http://www.softberry.com/ (Accessed on January, 2017)
39. Solovyev VV, Shahmuradov IA, Salamov AA. Identification of promoter regions and regulatory sites. *Computational Biology of Transcription Factor Binding*. 2010; 57-83p.
40. Upload DTU Bioinfomatics. (2017). Available from [online] http://www.cbs.dtu.dk/services/Promoter/ (Accessed on January, 2017)
41. Knudsen S. Promoter2. 0: for the recognition of PolII promoter sequences. *Bioinformatics*. 1999; 15 (5): 356-361p.
42. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*. 2005; 33(suppl 2): W306-W10p.
43. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *Journal of molecular biology*. 1994; 235 (3): 983-1002p.
44. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic acids research*. 2005; 33(suppl 2): W382-W8.

45. Upload PROSITE. (2017). Available from [online] http://prosite.expasy.org/ (Accessed on January, 2017)
46. Sigrist CJ, De Castro E, Cerutti L, et al. New and continuing developments at PROSITE. *Nucleic acids research*. 2012: gks1067.
47. DeLano WL. The PyMOL user's manual. *DeLano Scientific, San Carlos, CA*. 2002; 452.
48. DeLano WL. *The PyMOL molecular graphics system*. 2002.
49. Xi T, Jones IM, Mohrenweiser HW. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics*. 2004; 83 (6): 970-979p.

Cite this Article

Shovit Ranjan, Praveen Kumar Sharma. Comprehensive In silico Analysis of Single Nucleotide Polymorphisms (SNPs) in Human BDNF Gene. *Research & Reviews: A Journal of Life Sciences*. 2018; 8(3): 128–139p.

 © STM Journals 2018. All Rights Reserved