Research Journal of Pharmaceutical, Biological and Chemical Sciences

Computational approaches for identifying deleterious nonsynonymous variants from human Glucokinase gene.

Narayana Swamy A, Praveen Kumar Vemuri, and Srinivasulu Kamma*.

Department of Biotechnology, KL University, Vaddeswaram, India.

ABSTRACT

The human Glucokinase gene encodes an enzyme called Glucokinase, which plays key role in Glucose homeostasis. Many single nucleotide polymorphisms found in Glucokinase (GCK) gene have been associated with various disorders including, hyperglycemia, obesity and Neonatal Diabetes Mellitus. In this study, we performed a comprehensive analysis of functional and structural impacts of all known nsSNPs in Glucokinase gene using available computational prediction tools. A total of 107 nonsynonymous SNPs consisting of 98 missense and 9 nonsense variations were found in the Glucokinase gene. 51 of the 98 missense variants were predicted to be damaging or deleterious by three different software programs (PolyPhen-2, SIFT and PROVEAN), and 65 of them were predicted to be less stable using both the I-Mutant2.0 and MUpro software. Additionally, 9 nonsense variants were predicted to be produce a truncated GCK protein, further 30 nsSNPs which common to both deleterious and less protein stability, homology modeled structures were built by swiss PDB viewer using template 3F9M. The total energy and the RMSD values for the homology modeled structure and the mutant-structures were calculated. However, minor variations have been reported in total energies and RMSD's in contrary with the native model as well as mutant models.
Keywords: Muster, PolyPhen-2, SIFT, PROVEAN, MUpro, I-Mutant2.0.

Corresponding author

INTRODUCTION

Glucokinase (GCK) comes under the family of hexokinase gene and plays a key role in glucose homeostasis as a glucose sensor in pancreatic β-cells. It catalyses the initial step in the pathway: the ATP (adenosine triphosphate) dependent phosphorylation to form glucose-6-phosphate (G-6-P) from glucose [1]. A reduction in amount or activity of β-cell GCK increases the glucose threshold for insulin secretion, causing typical fasting hyperglycaemia. GCK is expressed in hepatocytes, pancreatic β-cells, and variety of neural/neuroendocrine cells including pancreatic α-cells, L- and K- gut enterocytes and selected neurons [2]. The GCK expressed from pancreas, liver and brain are coded by the same gene with 12 exons on chromosome 7 (7p15.3-p15.1) with similar in kinetic activity, but their primary structures in the N-terminal are different due to distinct splicing of the RNA transcript. The enzymes contain 465 amino acids and exon 1 varies in the diverse tissues because of the different promoter regions. In view of its crucial role in the regulation of glucose-stimulated insulin secretion, it is possible that mutations in the GCK gene can cause both hyperglycaemia and hypoglycaemia. Genetic studies have shown that GCK mutations are responsible for three different disorders of glucose regulation. Missense mutations of GCK represent the most frequent cause of MODY2; to date more than 200 mutations with distinct enzymatic characteristics have been found. These mutations were also detected in 5-6 % women with gestational diabetes. SNPs are the most common polymorphisms of DNA sequence variation for mapping complex genetic traits. About 500,000 SNPs present in the coding regions of the human genome. Among these, the nonsynonymous SNPs cause changes in the amino acid residues and likely to be an important factor contributing to the functional diversity of the encoded proteins in the human population [3]. A number of SNPs databases are available, such as the National Center for Biotechnology Information database, the human genome variation database (HGVBase) [4], and dbSNP [5]. The effect of intron region SNPs on gene regulation is difficult to understand. So attention is being focused towards nonsynonymous coding SNPs. These types of mutations are believed to be more likely to alter the function of a protein by changing the protein structure. These nsSNPs will affect the gene expression by modifying DNA and transcription factor binding [6, 7] and inactivate active sites of enzymes or change splice sites, thereby produce defective/unknown gene products [8, 9]. Epidemiologic association studies focus a great amount of effort into identifying SNPs in genes that may have an association with disease risk. Often, the SNPs that have an association with disease are those that are known as nonsynonymous SNPs, which result in an amino acid substitution. Many molecular epidemiologic studies focus on studying   SNPs found in coding regions in hopes of finding significant association between SNPs and disease susceptibility but often find little or no association [10]. With the availability of high-throughput SNP detection techniques, the population of nsSNPs is increasing rapidly, providing a platform for studying the genotype-phenotype interaction of human diseases. To select a nsSNP for an association study can be enhanced by first examining the potential impact of an amino acid variant on the function of the encoded protein with the use of different SNP detection programs such as Polymorphism phenotyping v2 (PolyPhen-2) [11], Sorting Intolerant From Tolerant (SIFT) [12], Protein Variation Effect Analyzer (PROVEAN) [13], I-Mutant2.0 [14], and MUpro [15, 16] software. Identifying the deleterious nsSNPs out of a pool of all the SNPs will be very useful for epidemiological population based studies. So the main aim of this study is to identify deleterious nsSNPs associated with GCK gene involved in GDM.

EXPERIMENTAL PROCEDURE

Analysis of functional effect on protein: There are many types of web-based softwares available that allow one to predict whether nonsynonymous coding SNPs may have functional effects on proteins. We selected Sorting Intolerant From Tolerant (SIFT) [17] to perform protein conservation analysis and predict the phenotypic effect of amino acid substitutions. Variants that present at conserved alignment positions are expected to be tolerated less than those that present at diverse positions. To construct a multiple sequence alignment of proteins that can be globally aligned to the query sequence and belong to the same clade by using Dirichlet mixture regularization [18] and a modified version of PSIBLAST [19]. Principle of this program is that it generates alignments with a large number of homologous sequences and assigns scores to each residue, ranging from 0.00 to 1.00. Based on the SIFT scores [20] the effects of nsSNPs on protein can be classified as intolerant (0.00–0.05), potentially intolerant (0.051–0.10), borderline (0.101–0.20), or tolerant (0.201–1.00). The higher the tolerance index of a particular amino acid substitution shows lesser is its likely impact.

Damaged GCK nsSNPs analysis: Polymorphism Phenotyping v2 (**PolyPhen-2**) is an automatic tool to predict the possible impact of an amino acid substitution on the structure and function of perticular protein. This prediction is based on straightforward empirical rules that are applied to the sequence, phylogenetic and

structural information characterizing the substitution. We will submit the query in the form of protein sequence with mutational position and two amino acid variants. PolyPhen searches for multiple alignments of homologous sequences, 3D structures of protein and amino acid contact information in various protein structure databases. The PolyPhen-2 server discriminates nonsynonymous SNPs into three main categories: benign, possibly damaging (less confident prediction), or probably damaging.

PROVEAN (Protein Variation Effect Analyzer), which predicts the functional impact for all classes of protein sequence variations of single amino acid substitutions and also deletions, insertions, and multiple substitutions. PROVEAN is a software tool that predicts whether an amino acid substitution has an impact on the biological function of a protein grounded on the alignment-based score [13]. Steps involved in the analysis were collect a set of homologous sequences and followed by compute a Provean score and make a prediction. The score measures the change in sequence similarity of a query sequence to a protein sequence homolog between without and with an amino acid variation of the query sequence. If the PROVEAN score ≤ -2.5, the protein variant is predicted to have a "deleterious" effect, while if the PROVEAN score is > -2.5, the variant is predicted to have a "neutral" effect.

Protein stability analysis:

I-Mutant2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) software was used to predict the protein stability changes caused by nsSNPs in a perticular protein. It is a support vector machine- (SVM-) based tool for the automatic prediction of protein stability change upon single amino acid substitution [14]. The software computed the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus unfolding Gibbs free energy value of the native protein (kcal/mol). A negative DDG value indicates that the mutated protein possesses low stability and vice versa.

MUpro (http://www.ics.uci.edu/~baldig/mutation.html) is also a support vector machine-based tool for the prediction of whether the protein stability is increased or decreased by nsSNPs of protein [15, 16]. The value of the energy change is predicted, and a confidence score between -1 and 1 for measuring the confidence of the prediction is calculated. A score less than 0 means the variant decreases the protein stability; conversely, a score greater than 0 means the variant increases the protein stability.

Modeling of GCK: Its Mutant Forms and RMSD Calculations

Structural analyses are performed based on the crystal structure of the protein for evaluating the structural stability of native and mutant proteins. We used dbSNP to identify the protein coded by GCK gene. Homology modeling approach was used for GCK 3D structure prediction. The modeling was performed by automated homology modeling program, SWISS MODEL [21]. The following steps were followed: template structure search using BLAST (http://www.ncbi .nlm.nih.gov/). The FASTA sequence of GCK was submitted to NCBI BLAST. Following BLAST query, the active conformation of human Glucokinase is not altered by allosteric activators. (PDB ID: 3F9M) was selected as template sequence (22). Mutant models were designed by using SWISS MODEL and energy minimisation was performed by DESMOND server. Total energy and RMSD values calculated by swiss

RESULTS AND DISCUSSION

SNP Analysis: By examining SNPs in the GCK gene using the dbSNP and ensembl databases, a total of 107 nonsynonymous SNPs were found. These SNPs consisted of 98 missense variations, 9 nonsense variations.

Deleterious nsSNP found by SIFT program: To determine which missense variants were damaging or deleterious, SIFT software was applied for the 98 missense variants of the GCK gene (Table 1). Each of the missense variants were submittedindependently to the SIFT program. Lower the tolerance index score, higher the functional impact of a perticular aminoacid substitution in a protein and vise-versa. A tolerance index score of ≤0.05 is considered to be deleterious. Out of 98 nsSNPs 70 were found to be deleterious (71.4%) with a tolerance index score of ≤ 0.05. It was also noted that, out of the 70 deleterious nsSNPs, 50 nsSNPs exhibited a highly deleterious tolerance index score of **0.00** as shown in table:1.
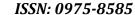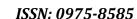
Table 1: PolyPhen SIFT and PROVEAN results of GCK

Nucleotide	Protein	dbSNP ID	PolyPhen-2 prediction (score)	SIFT prediction (score)	PROVEAN prediction (score)
c.74T>G	p.Leu25Arg	rs193922325	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-5.353)
c.131G>A	p.Gly44Asp	rs193922279	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-6.610)
c.130G>A	p.Gly44Ser	rs267601516	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-5.619)
c.175C>T	p.Pro59Ser	rs193922287	Probably damaging (1.000)	Damaging (0.04)	Deleterious (-7.170)
c.203G>A	p.Gly68Asp	rs373418736	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-6.565)
c.214G>A	p.Gly72Arg	rs193922289	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.794)
c.253A>T	p.Arg85Trp	rs193922290	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.447)
c.322T>G	p.Tyr108Asp	rs193922292	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-8.857)
c.385T>C	p.Cys129Arg	rs377106269	Probably damaging (1.000)	Damaging (0.00)	Deleterious(11.24)
c.433C>T	p.Pro145Ser	rs150779253	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.491)
c.440G>A	p.Gly147Asp	rs193922296	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-6.993)
c.449T>C	p.Phe150Ser	rs193922297	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.992)
c.457C>T	p.Pro153Ser	rs193922300	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.992)
c.509G>T	p.Gly170Val	rs193922303	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.475)
c.527C>G	p.Ala176Gly	rs193922304	Probably damaging (0.999)	Damaging (0.00)	Deleterious (-3.041)
c.532G>A	p.Gly178Arg	rs193922305	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.344)
c.563C>T	p.Ala188Val	rs193922307	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-3.853)
c.615C>G	p.Asp205Glu	rs193922312	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-3.909)
c.629T>A	p.Met210Lys	rs80356654	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-5.797)
c.630G>T	p.Met210Ile	rs193922313	Probably damaging (0.994)	Damaging (0.03)	Deleterious (-3.776)
c.635C>T	p.Ser212Phe	rs150077934	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-4.226)
c.659G>A	p.Cys220Tyr	rs193922316	Probably damaging (0.998)	Damaging (0.00)	Deleterious (-10.40)
c.683C>T	p.Thr228Met	rs80356655	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-5.808)
c.752T>C	p.Met251Thr	rs193922326	Probably damaging (0.995)	Damaging (0.00)	Deleterious (-5.417)
c.758T>G	p.Val253Gly	rs193921400	Probably damaging (0.984)	Damaging (0.00)	Deleterious (-6.562)
c.760A>C	p.Asn254His	rs193922327	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-4.693)
c.768G>C	p.Glu256Asp	rs193922328	Probably damaging (0.986)	Damaging (0.01)	Deleterious (-2.828)
c.779T>C	p.Phe260Ser	rs193922330	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.148)
c.781G>A	p.Gly261Arg	rs104894008	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.612)
c.895G>C	p.Gly299Arg	rs104894009	Probably damaging (1.000)	Damaging (0.00)	Deleterious(-7.398)
c.907C>T	p.Arg303Trp	rs193922336	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-7.425)
c.917T>C	p.Leu306Pro	rs193922337	Probably damaging (0.999)	Damaging (0.00)	Deleterious (-6.473)
c.944T>A	p.Leu315His	rs193922338	Probably damaging (0.997)	Damaging (0.00)	Deleterious (-6.387)
c.952G>T	p.Gly318Trp	rs193922340	Probably damaging (0.999)	Damaging (0.00)	Deleterious (-7.345)
c.971T>C	p.Leu324Pro	rs193922341	Probably damaging (0.997)	Damaging (0.00)	Deleterious (-6.375)
c.1015G>A	p.Glu339Lys	rs397514580	Probably damaging (0.996)	Damaging (0.00)	Deleterious(-3.630)
c.1124C>T	p.Ser375Phe	rs193922263	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-5.416)
c.1130G>A	p.Arg377His	rs193922264	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-4.575)
c.1132G>T	p.Ala378Thr	rs104894016	Probably damaging (0.992)	Damaging (0.00)	Deleterious (-2.579)
c.1133C>T	p.Ala378Val	rs193929374	Probably damaging (0.990)	Damaging (0.00)	Deleterious (-3.398)
c.1136C>A	p.Ala379Glu	rs193922265	Probably damaging (0.979)	Damaging (0.00)	Deleterious (-4.585)
c.1142T>G	p.Met381Arg	rs193922266	Probably damaging (0.997)	Damaging (0.00)	Deleterious (-3.833)
c.1153G>A	p.Gly385Arg	rs193922267	Probably damaging (1.000)	Damaging (0.01)	Deleterious (-5.127)
c.1157T>C	p.Leu386Pro	rs193922268	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-4.855)
c.1160C>A	p.Ala387Glu	rs193921338	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-4.122)
c.1160C>T	p.Ala387Val	rs193921338	Probably damaging (1.000)	Damaging (0.03)	Deleterious (-3.429)
c.1169T>A	p.Ile390Asn	rs193921340	Probably damaging (0.999)	Damaging (0.01)	Deleterious (-4.979)
c.1190G>T	p.Arg397Leu	rs193929375	Probably damaging (0.998)	Damaging (0.02)	Deleterious (-5.193)
c.1189C>T	p.Arg397Cys	rs370464857	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-6.005)
c.1240A>G	p.Lys414Glu	rs193922272	Probably damaging (0.999)	Damaging (0.02)	Deleterious (-3.571)
c.1289T>C	p.Leu430Pro	rs193922277	Probably damaging (1.000)	Damaging (0.00)	Deleterious (-6.211)
c.1307T>A	p.Ile436Asn	rs193922278	Probably damaging (1.000)	Damaging (0.01)	Deleterious (-5.491)
c.1345G>A	p.Ala449Thr	rs193922282	Probably damaging (1.000)	Damaging (0.04)	Deleterious (-3.380)

Identification of Damaged GCK nsSNPs: To determine the GCK nsSNPs that affected protein structure, polyphen-2 server was used to analyze the possible impact of aminoacids substitution on the protein structure and function. The GCK protein (NP_000153) with every nsSNP position and their both wild type as well as mutant aminoacid was submitted as input for analyzing the protein structural changes because of aminoacids

substitution. Our analysis showed 74 out of 98 nsSNPs to be showed probably damaging (76.5%), among those 49 nsSNPs showed maximum PSIC score 1.000, others were predicted to be possibly damaging or Benign (Table:1).

Functional impact of aminoacid substitution in protein:

To findout the functional effect of the aminoacid substitution in protein PROVEAN software tool was used. Total protein in fasta format and mutation position with wild type and mutant aminoacid was submitted to the server as input. Results shown 78 nsSNPs out of 98 to be deleterious with threshold value ≥ -2.5, remain were neutral. Total 53 GCK variants were predicted to be deleterious which common to SIFT, PolyPhen-2 and PROVEAN.

Identification of functional nsSNPs:

Next, the changes in the protein stability of the missense variants were examined using I-Mutant2.0 and MUpro software. In the I-Mutant2.0 (DDG) prediction, 84 (85.7%) of the 98 variants were predicted to be decreasing the protein stability and the others were predicted to be increasing the protein stability, in case of MUpro prediction, 76 (77.5%) of the 98 missense variants were predicted to be reducing the protein stability and others were predicted to be increasing the stability of protein (Table 2). A total of 65 variants (66.3%) out of the 98 missense variants, including 30 common damaging or deleterious variants namely p.Leu25Arg, p.Gly44Asp, p.Pro59Ser, p.Gly72Arg, p.Arg85Trp, p.Cys129Arg, p.Gly147Asp, p.Phe150Ser, p.Pro153Ser, p.Gly170Val, p.Ala176Gly, p.Asp205Glu, p.Met210Lys, p.Met251Thr, p.Val253Gly, p.Asn254His, p.Glu256Asp, p.Phe260Ser, p.Gly261Arg, p.Arg303Trp, p.Leu306Pro, p.Leu315His, p.Leu324Pro, p.Arg377His, p.Ala378Thr, p.Leu386Pro, p.Ile390Asn, p.Lys414Glu, p.Leu430Pro and p.Ala449Thr as determined using the PolyPhen-2, SIFT and PROVEAN software, were predicted to be less stable using both the I-Mutant2.0 and the MUpro software.

Table 2: I-Mutant2.0 (DDG) and MUpro prediction results

Protein	I-Mutant 2.0 prediction (DDG)	MUpro prediction (score)
p.Ala11Thr	Decrease (-1.22)	Decrease (-1.00)
p.Glu17Gly	Decrease (-1.97)	Decrease (-1.00)
p.Ile19Met	Decrease (-0.18)	Decrease (-0.47919321)
p.Leu25Arg	Decrease (-1.44)	Decrease (-1.00)
p.Arg36Pro	Decrease (-1.84)	Decrease (-1.00)
p.Gly44Asp	Decrease (-1.09)	Decrease (-0.11226136)
p.Thr49Asn	Decrease(-0.36)	Decrease (-0.53046485)
p.Pro59Ser	Decrease(-1.72)	Decrease (-0.30912217)
p.Gly72Arg	Decrease (-1.53)	Decrease (-0.85076235)
p.Arg85Trp	Decrease (-0.63)	Decrease (-0.22957849)
p.Cys129Arg	Decrease (-0.31)	Decrease (-0.045416174)
p.Ser131Pro	Decrease (-2.99)	Decrease (-0.55305962)
p.Met139Leu	Decrease (-1.33)	Decrease (-0.26218606)
p.Gly147Asp	Decrease (-0.53)	Decrease (-0.24484083)
p.Phe150Ser	Decrease (-2.31)	Decrease (-0.2238206)
p.Pro153Ser	Decrease (-1.17)	Decrease (-1.00)
p.Arg155Gly	Decrease (-2.31)	Decrease (-1.00)
p.Gly170Val	Decrease (-2.06)	Decrease (-0.45306308)
p.Ala176Gly	Decrease (-0.93)	Decrease (-1.00)
p.Val181Ala	Decrease (-1.53)	Decrease (-0.85089369)

p.Gly193Arg	Decrease (-0.89)	Decrease (-0.47861572)
p.Glu196Gly	Decrease (-0.99)	Decrease (-1.00)
p.Met202Val	Decrease (-1.14)	Decrease (-1.00)
p.Met202Thr	Decrease (-0.66)	Decrease (-0.91830769)
p.Asp205Glu	Decrease (-0.17)	Decrease (-0.00173817)
p.Met210Lys	Decrease (-1.59)	Decrease (-0.029189533)
p.Glu221Lys	Decrease (-0.51)	Decrease (-0.63878742)
p.Val226Ala	Decrease (-0.43)	Decrease (-1.00)
p.Ala232Thr	Decrease (-1.01)	Decrease (-0.048656291)
p.Met235Thr	Decrease (-1.91)	Decrease (-0.18332828)
p.Arg250His	Decrease (-0.97)	Decrease (-1.00)
p.Met251Thr	Decrease (-1.53)	Decrease (-1.00)
p.Val253Ala	Decrease (-0.71)	Decrease (-1.00)
p.Val253Gly	Decrease (-2.36)	Decrease (-1.00)
p.Asn254His	Decrease (-1.24)	Decrease (-0.01808416)
p.Glu256Asp	Decrease (-0.69)	Decrease (-0.48284055)
p.Phe260Ser	Decrease (-2.37)	Decrease (-0.55792462)
p.Gly261Arg	Decrease (-1.76)	Decrease (-0.18765988)
p.Gly264Ser	Decrease (-1.08)	Decrease (-0.52492178)
p.Leu271Pro	Decrease (-1.00)	Decrease(-0.48548809)
p.Leu276Val	Decrease (-1.69)	Decrease (-1.00)
p.Glu279Gln	Decrease (-0.31)	Decrease (-0.35940335)
p.Glu279Gly	Decrease (-0.96)	Decrease (-0.77508288)
p.Arg303Trp	Decrease (-0.35)	Decrease (-0.67722617)
p.Leu306Pro	Decrease (-0.74)	Decrease (-1.00)
p.Leu315His	Decrease (-1.86)	Decrease (-1.00)
p.Phe316Tyr	Decrease (-0.72)	Decrease (-0.97745194)
p.Leu324Pro	Decrease (-1.05)	Decrease (-1.00)
p.Ser340Gly	Decrease (-2.18)	Decrease (-1.00)
p.Ile348Phe	Decrease (-1.62)	Decrease (-1.00)
p.Arg377His	Decrease (-0.90)	Decrease(-0.0096188303)
p.Ala378Thr	Decrease (-0.81)	Decrease (-0.16361444)
p.Leu386Pro	Decrease (-1.71)	Decrease (-0.86992943)
p.Ile390Asn	Decrease (-1.55)	Decrease (-0.69497455)
p.Ile390Thr	Decrease (-3.45)	Decrease (-0.74331655)
p.Arg392Leu	Decrease (-0.73)	Decrease(-0.080982809)
p.Arg403Gly	Decrease (-2.53)	Decrease (-0.97494644)
p.Lys414Glu	Decrease (-0.33)	Decrease (-0.32475836)
p.Phe423Tyr	Decrease (-0.4)	Decrease (-0.41807022)
p.Arg429Lys	Decrease (-2.37)	Decrease (-0.95622978)
p.Leu430Pro	Decrease (-0.83)	Decrease (-1.00)
p.Thr431Ala	Decrease (-0.83)	Decrease (-1.00)
p.Arg447Gly	Decrease (-2.13)	Decrease (-1.00)
p.Ala449Thr	Decrease (-0.89)	Decrease (-0.69146143)
p.Val455Met	Decrease (-1.30)	Decrease (-0.35492159)

Regarding the 9 nonsense variations, c.304A>T (p.Lys102Ter), c.556C>T (p.Arg186Ter), c.645C>A (p.Tyr215Ter), c.793G>T (p.Glu265Ter) variants were predicted to loss the Nucleotide-Binding Domain of the sugar region where Protein which binds a nucleotide, a phosphate ester of a nucleoside consisting of a purine or pyrimidine base linked to ribose or deoxyribose phosphates, and c.793G>T (p.Glu265Ter), c.835G>T (p.Glu279Ter), c.871A>T (p.Lys291Ter), c.1114G>T (p.Glu372Ter) variations (Table:3) were predicted to loss the Hexokinase-2 domain activity in the GCK protein (22).

Table 3: Summary of nonsense variations of the GCK gene

dbSNP ID	Nucleotide	Protein
rs193922329	c.76C>T	Q26Ter
rs193922259	c.103A>T	R35Ter
rs193922291	c.304A>T	K102Ter
rs104894006	c.556C>T	R186Ter
rs144723656	c.645C>A	Y215Ter
rs104894011	c.793G>T	E265Ter
rs104894005	c.835G>T	E279Ter
rs193922335	c.871A>T	K291Ter
rs193922262	c.1114G>T	E372Ter

Modeling of mutant protein: SWISS MODEL Expasy showed one 3D structure which was using the template PDB ID 3F9M. The modeled 3D structure included residues ranging from 4 to 458 of 465 aminoacids of GCK. Later, all mutant 3D structures were generated through SWISS MODEL by using template 3f9m. Following structure modelling, the native and mutants were subjected to energy minimization through DESMOND server with 2000 iterations. Further total energy calculated by Swiss PDB viewer. The energy minimized structures of mutants were only used for further analysis like RMSD calculation by structural alignment in SCHRODINGER server. The total energy and the RMSD values for the native structure and the mutant-structures are given in Table 4. The total energy and RMSD values for all mutants were ranges between from -4342.067639 Kcal mol^{-1} to -4880.8116635 Kcal mol^{-1} and 0.114A° to 0.218A° respectively.

Table 4: RMSD and total energy of modeled structure and its mutant forms

RESIDUE	TOTAL ENERGY kcal/mol	RMSD (A°)
Homology Modeled structure	-4744.565966	-
Mutant (p.Leu25Arg)	-4880.811664	0.222
Mutant (p.Gly44Asp)	-4660.594168	0.208
Mutant (p.Pro59Ser)	-4679.379302	0.202
Mutant (p.Gly72Arg)	-4744.77653	0.201
Mutant(p.Arg85Trp)	-4714.673279	0.212
Mutant (p.Cys129Arg)	-4441.387189	0.208
Mutant (p.Gly147Asp)	-4821.130258	0.225
Mutant (p.Phe150Ser)	-4762.507648	0.143
Mutant (p.Pro153Ser)	-4859.530354	0.224
Mutant (p.Gly170Val)	-4787.048279	0.215
Mutant (p.Ala176Gly)	-4798.789914	0.222
Mutant (p.Asp205Glu)	-4723.619742	0.208
Mutant (p.Met210Lys)	-4829.304732	0.222
Mutant (p.Met251Thr)	-4655.890535	0.201
Mutant (p.Val253Gly)	-4626. 0564054	0.202
Mutant (p.Asn254His)	-4765.168499	0.222
Mutant (p.Glu256Asp)	-4829.979685	0.227
Mutant (p.Phe260Ser)	-4805.453633	0.222
Mutant (p.Gly261Arg)	-4739.74044	0.21
Mutant (p.Arg303Trp)	-4657.101099	0.226
Mutant (p.Leu306Pro)	-4605.012667	0.207
Mutant (p.Leu315His)	-4791.025813	0.225
Mutant (p.Leu324Pro)	-4342.067639	0.222
Mutant (p.Arg377His)	-4804.387428	0.203
Mutant (p.Ala378Thr)	-4776.763862	0.219
Mutant (p.Leu386Pro)	-4754.779637	0.216
Mutant (p.Ile390Asn)	-4707.367352	0.202
Mutant (p.Lys414Glu)	-4655.787763	0.203
Mutant (p.Leu430Pro)	-4655.507887	0.203
Mutant (p.Ala449Thr)	-4824.293977	0.222

Amino acid substitutions presently account for approximately half of the known gene lesions responsible for inherited disease in humans. A major interest in human genetics is to find out the mutations

that are functionally neutral from those that contribute to disease. Identification of nsSNPs that are affects the protein functions and structures related to disease is an important task.

CONCLUSION

In our analysis, we found 30 nsSNPs which are less stable, deleterious and probably damaging by both Functional and structural analysis prediction tools. 9 nonsense variants showed truncated protein synthesis. The RMSD values of mutant models were not significant although total energies shown difference in contrary to native model. Molecular docking and simulation studies are required to assess the highly deleterious nsSNPs. Use of these prediction tools to find out the potentially polymorphic nsSNPs for epidemiology studies can be an efficient and systemic way to explore the role of genetic variation in disease risk and to curtail cost. Furthermore, predicted impact of these nsSNPs can be tested with the use of animal models or cell lines to determine if functionality of the protein has indeed been altered.

ACKNOWLEDGMENT

REFERENCES

[1]   Matschinsky F, Liang Y, Kesavan P, Wang L, et al. Glucokinase as pancreatic beta cell glucose sensor and diabetes gene. J. Clin. Invest 1993; 92: 2092-2098.
[2]   Schuit FC, Huypens P, Heimberg H and Pipeleers DG. Glucose sensing in pancreatic beta- cells: a model for the study of other glucose-regulated cells in gut, pancreas, and hypothalamus. Diabetes 2001; 50: 1-11.
[3]   Rajasekaran R, Sudandiradoss C, Doss C.G, and Sethumadhavan R, "Identification and in silico analysis of functional SNPs of the BRCA1 gene," Genomics 2007; vol. 90, no. 4, pp. 447– 452.
[4]   Fredman D, Siegfried M, Yuan Y.P, Bork P, Lehv¨aslaiho H, and Brookes A.J, "HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources," Nucleic Acids Research 2002; vol. 30, no. 1, pp. 387– 391.
[5]   Smigielski E.M, Sirotkin K, Ward M, and Sherry S.T, "dbSNP: a database of single nucleotide polymorphisms," Nucleic Acids Research 2000; vol. 28, no. 1, pp. 352–355.
[6]   Barroso I, Gurnell M, Crowley V.E et al., "Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension," Nature 1999; vol. 402, no. 6764, pp. 880–883.
[7]   Thomas R. McConnell R, Whittacker J, Kirkpatrick P, Bradley J, and Sand ford R, "Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR," American Journal of Human Genetics 1999; vol. 65, no. 1, pp. 39–49.
[8]   Yoshida A, Huang I.Y, and Ikawa M, "Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals," Proceedings of the National Academy of sciences of the United States of America 1984; vol. 81, no. 1, pp. 258– 261.
[9]   Jaruzelska J, Abadie V, d'Aubenton-Carafa Y, Brody E, Munnich A, and Marie J, "In vitro splicing deficiency induced by a C to T mutation at position -3 in the intron 10 acceptor site of   the phenylalanine hydroxylase gene in a patient with phenylketonuria," Journal of Biological Chemistry 1995; vol. 270, no. 35, pp. 20370–20375.
[10]   Johnson M.M, Houck J, and Chen C, "Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response," Cancer Epidemiology Biomarkers and Prevention 2005; vol. 14, no. 5, pp. 1326–1329, 2005.
[11]   Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7: 248-249.
[12]   [12] 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.
[13]   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.
[14]   Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res 2004; 32: D120-121.

[15]   Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 2005; 33: W72-76.

[16]   Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 2006; 62: 1125-1132.

[17]   P. C. Ng and S. Henikoff, "Predicting deleterious amino acid substitutions," Genome Research 2001; vol. 11, no. 5, pp. 863–874.

[18]   S. F. Altschul, T. L. Madden, A. A. Sch¨affer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Research 1997; vol. 25, no. 17, pp.3389–3402.

[19]   Sjolander K, Karplus K, Brown M et al., "Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology," Computer Applications in the Biosciences1996; vol. 12, no. 4, pp. 327–345.

[20]   Arnold K, Bordoli L, Kopp J, and Schwede T, "The SWISSMODEL workspace: a web-based environment for protein structure homology modelling," Bioinformatics 2006; vol. 22, no. 2,pp. 195–201.

[21]   Xi T, Jones I.M, and Mohrenweiser H.W, "Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein   function," Genomics 2004; vol. 83, no. 6, pp. 970–979.

[22]   Petit P, Antoine M, Ferry G, Boutin JA, Lagarde A, Gluais L et al., The active conformation of human Glucokinase is not altered by allosteric activators. Acta Crystallogr D Biol Crystallogr 2011 Nov; 67(Pt 11):929-35.