Computational Analysis of Single Nucleotide Polymorphism (SNPs) in Human SLC5A1 Gene

Rashid Abualamah Albasheer Abbas, Afra Mohamed Suliman Albakry, Mona Abdelrahman Mohamed Khaier, Hind Abdelaziz Elnasri*

Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri, Khartoum, Sudan

Email address:
rashidabualma93@gmail.com (R. A. A. Abbas), aframoh2016@gmail.com (A. M. S. Albakry), munakhaier@gmail.com (M. A. M. Khaier), hindnasri2017@gmail.com (H. A. Elnasri)

*Corresponding author

To cite this article:
Rashid Abualamah Albasheer Abbas, Afra Mohamed Suliman Albakry, Mona Abdelrahman Mohamed Khaier, Hind Abdelaziz Elnasri. Computational Analysis of Single Nucleotide Polymorphism (SNPs) in Human SLC5A1 Gene. International Journal of Biomedical Science and Engineering. Vol. 7, No. 4, 2019, pp. 85-91. doi: 10.11648/j.ijbse.20190704.12

Received: August 15, 2019; Accepted: November 8, 2019; Published: December 23, 2019

Abstract: Glucose galactose malabsorption (GGM) is an autosomal recessive disease manifesting within the first weeks of life. It is characterized by a selective failure to absorb dietary glucose and galactose from the intestine leading to severe life threatening diarrhea and dehydration. Mutations in the Na+/glucose co-transporter gene (SLC5A1) gene have been determined to be associated with congenital GGM. In this study different computational tools were used to investigate the nsSNPs (Single nucleotide polymorphisms) in the SLC5A1 gene and to determine their effects on the protein function and structure. The protein structural analysis was done by modeling using Project Hope and Chimera after homology modeling by CPH models 3.2. In addition Gene MANIA software was used to study the association between this gene and related ones. A total of 166 nsSNPs were obtained from the SNPs database in NCBI during 2019. A total of 37 SNP were predicted to be deleterious using SIFT software, while 25 SNPs were predicted to be probably damaging by PolyPhen-2 and 30 SNPs were predicted to be deleterious by PROVEAN. The results of SIFT, PolyPhen-2, PROVEAN, SNPs&GO, PHD-SNP collectively revealed that 16 SNPs were predicted to be highly damaging.

Keywords: Computational Analysis, Glucose–Galactose Malabsorption, SLC5A1 Gene

1. Introduction

Glucose / galactose malabsorption (GGM) is an autosomal recessive disease manifesting within the first weeks of life and is characterized by a selective failure to absorb dietary glucose and galactose from the intestine [1]. Patients with GGM are presented with the neonatal onset of severe life-threatening watery diarrhea and dehydration [2]. It was first described in 1962 [3]. The diarrhea ceases within one hour after removing oral intake of lactose, glucose, and galactose, but promptly returns with the introduction of one or more of the offending sugars into the diet [4].

Secondary active transport of glucose occurs via symport with sodium, using SGLT proteins (sodium-glucose transport protein), in the choroid plexus, proximal tubules of kidneys, and the intestine [4]. Mutations in the Na+/glucose co-transporter gene SLC5A1 (Solute Carrier Family 5 Member 1 (Sodium/Glucose Cotransporter) can cause structural and functional deletion in the SGLT-1 proteins thus glucose and galactose are not absorbed from the intestine leading to clinical manifestations [5]. A total of more than 40 SLC5A1 mutation have been identified in patients with congenital Glucose / galactose malabsorption up to date [6].

The SLC5A1 gene encoding the SGLT1 membrane protein was cloned and sequenced in 1987 [7]. This gene is located within chromosome 22q13.1 and is composed of 15 exons. Expression of SLC5A1 gene is mainly in the intestine and kidney. The translated protein is composed of 664 amino acids with a molecular mass of approximately 73 kDa, consisting of a core of 13 transmembrane domains [8-9]. For various reasons it might not be feasible to perform
laboratory studies for all SNPs in a specific gene or even the whole genome. Thus computational studies are now becoming indispensable for the identification and prioritization of SNPs with functional importance from an enormous number of non-risk alleles. Computational methods are sufficiently fast and flexible and can provide predictions of functionally significant SNPs with a high accuracy of 80–85% [10] if combined with other techniques such as sequencing, structure and phylogenetic relationships.

In this study different computational methods were used to identify the SNPs (Single nucleotide polymorphisms) in SLC5A1 gene and the effects of the predicted mutation on the protein function and structure.

2. Methodology

SLC5A1 gene was investigated in dbSNP/NCBI database using computational analysis. The SNPs and the related ensembles protein (ESNP) were obtained from the SNPs database (dbSNPs) http://www.ncbi.nlm.nih.gov/snp and Uniprot database during the year 2019. Several software were used for analysis.

2.1. GeneMANIA

(http://www.genemania.org). GeneMANIA finds related genes to the input genes, using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. Gene MANIA can be used to find new members of a pathway, additional genes which where missed in screening or find new genes with a specific function [11]. The input was SLC5A1 gene name and the results are usually shown as a diagram and tables showing the relation between the different genes.

2.2. SIFT: “Sorting Intolerant from Tolerant”

(http://sift.mit.edu/www/SIFT_dbSNP.html). It is a sequence homology-based tool that presumes important amino acids will be conserved in the protein family. Hence, changes at well-conserved positions tend to be predicted as deleterious or tolerated. A list of nonsynonymous ID (rsID) that were obtained from the dbSNP database were the input for SIFT and then only the deleterious SNPs were chosen for further analysis. The cutoff value in the SIFT program is a tolerance index of ≥0.05. The higher the tolerance index, the less functional impact a particular amino acid substitution is likely to have [12].

2.3. PolyPhen-2 (Polymorphism Phenotyping v2)

(http://genetics.bwh.harvard.edu/pph2/). It is an online bioinformatics program that predicts the possible impact of amino acid substitution on the stability and function of human proteins using structural and comparative evolutionary considerations. This program basically searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, then calculates position specific independent count scores (PSIC) for each of the two variants, and then computes the PSIC scores difference between two variants. The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have [13]. Prediction outcomes could be classified as benign, possibly damaging or probably damaging. For structural and functional predictions, SNPs that were predicted to be deleterious by SIFT were submitted to PolyPhen-2 as protein sequence in FASTA format (obtained from Expasy), along with the position of the mutation, native and the new substituent amino acids.

2.4. PROVEAN (Protein Variation Effect Analyzer)

(http://provean.jcvi.org/index.php). It is a software tool which predicts the effect of all classes of protein sequence variations such as single amino acid substitutions, insertions, deletions, and multiple substitution on the function of protein. Prediction outcomes could be classified as deleterious or neutral [14]. The protein sequence in FASTA was again the input for this software.

2.5. SNPs&GO (Predicting Disease Associated Variations Using GO (Gene Ontology Terms))

SNPs&GO (http://snps.biofold.org/snpands-and-go/snpands-and-go.html). It is an accurate method that, starting from a protein sequence, can predict whether a mutation is disease related or not by exploiting the protein functional annotation. SNPs&GO collects in unique framework information derived from protein sequence, evolutionary information, and function as encoded in the Gene Ontology terms, and outperforms other available predictive methods [15]. The protein sequence and mutation sites were the input for this software.

2.6. PHD-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphisms)

PHD-SNP is a web-based tool available at (http://snps.biofold.org/phd-snp/phd-snp.html). PhD-SNP is a Support Vector Machines (SVMs) based method that predicts disease associated nsSNPs using sequence information. The protein sequence and mutation positions were the input. For each mutation, PhD-SNP returns an output score (ranging from 0-1) that represents the probability of this nsSNP being associated with disease. The method considers 0.5 to be the threshold above it the nsSNPs are predicted to be disease-associated [16].

2.7. Protein Stability

In order to predict the effect of single point mutation on the protein stability, two software were used:

2.7.1. I-Mutant Suite

(http://gpcr2.biocomp.unibo.it/cgi/predictors/IMutant3.0.cgi). It is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single
point mutations. The input was the protein sequence, the position of the SNP in the protein and the new residue. The method allows to predict if a mutation can largely destabilize the protein (Gibbs-free energy change DDG <-0.5 Kcal/mol) or largely stabilize (DDG >0.5 Kcal/mol) or have a weak effect (-0.5≤G≤0.5 Kcal/mol)[17].

2.7.2. MUpro

http://mupro.proteomics.ics.uci.edu/). It is another web server for prediction of protein stability changes upon mutations. It use support vector machines to predict protein stability changes for single-site mutations by using sequence information. The protein sequence and point of mutation was the input and the output is either increased or decreased stability [18].

2.8. Project hope

(http://www.cmbi.ru.nl/hope/). It is a fully automatic program that analyzes the structural and functional effects of point mutations. It builds a report with text, figures, and animations [19]. The protein sequence in FASTA format, wild and new amino acid and point of substitution were the input for Project hope.

2.9. CPHmodels3.2

(http://www.cbs.dtu.dk/services/CPHmodels/) is a web server predicting protein 3D structure by using asingle template homology modeling. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions [20].

2.10. Chimera

(http://www.cgl.ucsf.edu/chimera). It is used to generate the mutated 3D model models of each protein [21].

3. Results and Discussion

The goal of this study was to analyze the nsSNPs in SLC5A1 gene and the effect of predicted mutations at the proteomic level. SLC5A1 gene plays a vital role in human body and it was found to be co-expressed and shared domains with 11 genes as predicted by GeneMANIA (Figure 1 and Table 1).

![Figure 1. Genetic interactions, pathways, co-expression, co-localization and protein domain similarity of SLC5A1 gene.](image)

| Function                                      | FDR       | Genes in network | Genes in genome |
|-----------------------------------------------|-----------|------------------|-----------------|
| Solute:Sodium Symporter Activity              | 9.42E-06  | 4                | 12              |
| Sodium Ion Transmembrane Transporter Activity | 6.59E-05  | 5                | 65              |
| Symporter Activity                            | 0.00027572| 4                | 36              |
| Solute:Cation Symporter Activity              | 0.00027572| 4                | 35              |
| Active Transmembrane Transporter Activity     | 0.00039644| 5                | 111             |
| Secondary Active Transmembrane Transporter Activity | 0.00194446 | 4              | 64              |
| Monovalent Inorganic Cation Transmembrane Transporter Activity | 0.0031379 | 5              | 180             |
| Metal Ion Transmembrane Transporter Activity  | 0.01081159| 5                | 238             |
| Inorganic Cation Transmembrane Transporter Activity | 0.01651465 | 5              | 266             |
| Alcohol Transmembrane Transporter Activity    | 0.12788567| 2                | 12              |
A total of 166 ns SNPs were obtained from the SNPs database (dbSNPs) in NCBI. Following analysis using SIFT software, 37 SNPs were predicted to be deleterious. A total of 25 SNPs were predicted to be probably damaging by PolyPhen-2 and 30 SNPs were predicted to be deleterious by PROVEAN as shown in (Tables 2, Appendix A1).

Table 2. The results SIFT, Polyphen-2, and PROVEAN.

| SOFTWARE | RESULTS |
|----------|---------|
| SIFT     | Total: 166 ns SNPs. 37 deleterious. |
| PolyPhen-2 | i 25 probably damaging. ii 9 possibly damaging. iii 3 Benign. |
| PROVEAN  | i 30 Deleterious. ii 7 Neutral |

Analysis with SNPs &GO and PHD-SNP showed different results, 23 SNPs were predicted to be disease related with SNPs &GO compared to 32 with PHD-SNP (Figure 2, Appendix A2).

![Image](image)

**Figure 2.** The result of SNPs & GO compared to PHD-SNP

From the results of all five software (SIFT, PolyPhen-2, PROVEAN, SNPs&GO and PHD-SNP) 16 SNPs were predicted to be highly damaging (Table 3).

Table 3. Prediction results of nsSNPs using five different softwares.

| No | SNP ID       | Amino acid change | SIFT prediction | Polyphen-2Predication | Provean Prediction | SNP and GO | PHD SNP |
|----|-------------|------------------|----------------|-----------------------|--------------------|-----------|---------|
| 1  | rs33939896  | T76M             | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 2  | rs121912669 | D28G             | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 3  | rs199573966 | R287H            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 4  | rs199872285 | T81K             | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 5  | rs200004849 | V298A            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 6  | rs200304934 | S393F            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 7  | rs200406921 | A341E            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 8  | rs201079555 | R173C            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 9  | rs201216997 | L211V            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 10 | rs201271081 | R443C            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 11 | rs201598524 | R287C            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 12 | rs202070786 | K321N            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 13 | rs370932142 | T144M            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 14 | rs371505974 | G191R            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 15 | rs372081140 | T421I            | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |
| 16 | rs373203939 | R8Q              | deleterious     | probably damaging     | Deleterious        | Disease   | Disease |

Regarding protein stability, the stability was found to be decreased in all SNPs except in two SNPs: rs200304934 and rs202070786 which showed increased stability when I-mutant software has been used, and only one mutation: rs199872285 showed increased protein stability. The prediction accuracy based on sequence information alone is close to the accuracy of methods that depend on tertiary structure information. MUpro software overcomes one important shortcoming of approaches that require tertiary structures to make accurate predictions. Thus, this method can be used on a genomic scale to predict the stability changes for large numbers of proteins with unknown tertiary structure [18].

The SNPs were further submitted to the Project Hope software to see the effect of amino acid substitution on protein structure. Each amino acid has its own specific size, charge, and hydrophobicity value and the wild type residue and newly introduced mutant residue often differ in these properties. Differences in size in all predicted SNPs can affect the contact with the lipid-membrane. In addition, differences in hydrophobicity can affect the hydrophobic interactions with the membrane lipids and can result in loss of hydrogen bonds and/or disturb correct folding. This was predicted for SNPs: rs121912669, rs33939896, rs199872285, rs200304934, rs201079555, rs201598524, rs371505974, rs201271081, rs370932142 (Table 4).

Difference in charge between wild-type and mutant residue can also affect protein function can cause loss of interactions with other molecules or residues. This was predicted for SNPs: (rs121912669, rs201079555, rs199573966, rs201598524, rs371505974, rs201271081) (Table 4).

Three SNPs namely (rs121912669), (rs371505974) and (rs200406921) have been reported in previous studies [8, 22] to be associated with SGLT and in this study they were predicted to be highly damaging by all software. Also, a recent study revealed two novel SNPs among Saudi population suffering from congenital Glucose galactose malabsorption (G89R and G435D) [8].

Another SNP (rs121912668) has also been reported to be disease related in a previous study [1] while in the current
study it was predicted to be highly damaging by all software programs except in PolyPhen-2 it predicted to be possibly damaging with high score 0.936.

It is thus important to differentiate between disease associated and neutral SNPs since this will help in understanding the relationship between the genotype and phenotype and provide a better diagnosis strategies.

### Table 4. 3D model by Chimera and project hope for SGLT1 protein.

| SNP ID     | Amino acid change | Wild type | mutant type | 3D structure | Project hope |
|------------|-------------------|-----------|-------------|--------------|--------------|
| rs33939896 | T76M              | ![Wild type](image1.png) | ![Mutant type](image2.png) | ![3D structure](image3.png) | ![Wild type](image4.png) ![Mutant type](image5.png) |
| rs200304934| S393F             | ![Wild type](image6.png) | ![Mutant type](image7.png) | ![3D structure](image8.png) | ![Wild type](image9.png) ![Mutant type](image10.png) |
| rs201079555| R173C             | ![Wild type](image11.png) | ![Mutant type](image12.png) | ![3D structure](image13.png) | ![Wild type](image14.png) ![Mutant type](image15.png) |
| rs371505974| G191R             | ![Wild type](image16.png) | ![Mutant type](image17.png) | ![3D structure](image18.png) | ![Wild type](image19.png) ![Mutant type](image20.png) |
| rs200406921| A341V             | ![Wild type](image21.png) | ![Mutant type](image22.png) | ![3D structure](image23.png) | ![Wild type](image24.png) ![Mutant type](image25.png) |

### 4. Conclusions

In this study found 16 ns SNPs were identified in mutations SLC5A1 gene. Three of the predicted SNPs were also reported in clinical trials, while the others need further confirmative studies. Predicting the phenotypic effect of nsSNPs using computational algorithms will also help in better understanding of the genetic variations in response to diseases, albeit that computation prediction need further conformation using clinical studies.

### Appendix

#### Table A1. Analysis of SLC5A1 nsSNPs predicted with SIFT, P-2olyphen, and PROVEAN programs.

| SNP ID     | amino acid change | SIFT score | PROTEIN ID  | SIFT prediction | Polyphen –2 prediction | Polyphen score | PROVEAN prediction | Score |
|------------|-------------------|------------|-------------|-----------------|------------------------|---------------|-------------------|-------|
| rs33915717 | R15W              | 0.03       | ENSP00000266088 | deleterious | probably damaging | 0.958         | neutral           | -1.21 |
| rs33939337 | L527F             | 0.022      | ENSP00000266088 | deleterious | probably damaging | 0.998         | deleterious        | -3.56 |
| rs33939896 | T76M              | 0.025      | ENSP00000444898 | deleterious | probably damaging | 1             | deleterious        | -5.88 |
| rs33978633 | N354S             | 0.037      | ENSP00000444898 | deleterious | Benign               | 0.06          | deleterious        | -3.69 |
| rs111735032| M201V             | 0.044      | ENSP00000444898 | deleterious | Benign               | 0.058         | neutral           | -1.65 |
| rs121912668| D28N              | 0          | ENSP00000266088 | deleterious | possibly damaging   | 0.936         | deleterious        | -4.38 |
| SNP ID     | Amino acid change | Protein ID         | SNP and GO | RI | Prop- | PHI SN | RI | Prop- |
|-----------|-------------------|--------------------|------------|----|-------|--------|----|-------|
| rs33915717 | R15W              | ENSP0000026608     | Neutral    | 7  | 0.145 | Neutral | 1  | 0.43  |
| rs33939337 | L527F             | ENSP0000026608     | Neutral    | 1  | 0.431 | Disease | 3  | 0.651 |
| rs33939896 | T76M              | ENSP0000044489     | Disease    | 6  | 0.817 | Disease | 8  | 0.902 |
| rs33978633 | N354S             | ENSP0000044489     | Neutral    | 5  | 0.251 | Disease | 3  | 0.66  |
| rs11735032 | M201V             | ENSP0000044489     | Disease    | 0  | 0.501 | Disease | 2  | 0.599 |
| rs12191266 | D28N              | ENSP0000044489     | Disease    | 5  | 0.759 | Disease | 6  | 0.853 |
| rs37093214 | T144M             | ENSP0000044489     | Disease    | 7  | 0.844 | Disease | 8  | 0.909 |
| rs77150974 | G191R             | ENSP0000044489     | disease    | 1  | 0.474 | disease | 2  | 0.61  |
| rs73720393 | R8Q               | ENSP0000044489     | disease    | 1  | 0.792 | disease | 5  | 0.727 |

Table A2: ns SNPs in SL5C1A gene predicted with SNPs and GO and PHD SNPs programs.
| SNP ID      | Amino Acid Change | Protein ID          | SNP and GO | RI | Prop-ability | PHD SNP | RI | Prop-ability |
|-------------|-------------------|---------------------|------------|----|--------------|---------|----|--------------|
| rs202033427 | K122N             | ENSP00000266088     | Neutral    | 5  | 0.266        | Disease | 3  | 0.65         |
| rs202070786 | K321N             | ENSP00000266088     | Disease    | 1  | 0.547        | Disease | 1  | 0.528        |
| rs377741549 | M198L             | ENSP00000444898     | Disease    | 0  | 0.514        | Disease | 4  | 0.68         |
| rs370932142 | T144M             | ENSP00000444898     | Disease    | 2  | 0.584        | Disease | 0  | 0.5          |
| rs371505974 | G191R             | ENSP00000444898     | Disease    | 7  | 0.847        | Disease | 8  | 0.925        |
| rs372081140 | T421I             | ENSP00000444898     | Disease    | 5  | 0.762        | Disease | 8  | 0.889        |
| rs373203939 | R8Q               | ENSP00000444898     | Disease    | 6  | 0.799        | Disease | 8  | 0.897        |

References

[1] Turk E, ZabelB, MundlosS, Dyer Jand Wright E M. (1991). Glucose–galactose malabsorption caused by defect in Na+/glucose co-transporter. *Nature* 350: 354–356.

[2] Xin B, Wang H. (2011). Multiple sequence variations in SLC5A1 gene are associated with glucose–galactose malabsorption in a large cohort of Old Order Anish. *Clin Genet*. 79: 86–91.

[3] Lindquist B, Meeuwisse GW. (1962) Chronic diarrhea caused by monosaccharide malabsorption. *Acta Paediatr*. 51: 674–685.

[4] Wright EM, Turk E, Martin MG. (2002). Molecular basis for glucose galactose malabsorption. *Cell Biochem Biophys*. 36: 115–121.

[5] Assiri A, Saeed A, Alnimri A, et al. (2013). Five Arab children with glucose-galactose malabsorption. *Paediatr Intern Child Health*. 33: 108–110.

[6] Al-Suyufi Y, Al Saleem K, Al-Mehaidib A, et al. (2018). SLC5A1 mutations in Saudi Arabian patients with congenital glucose- galactose malabsorption. *J Pediatr Gastroenterol Nutr*. 66: 250–2.

[7] Hediger MA, Coady MJ, Ikeda TS, et al. (1987). Expression cloning and cDNA sequencing of the Na+/glucose co-transporter. *Nature*. 330: 379–81.

[8] Turk E, Martin MG, Wright EM. (1994). Structure of the human Na+/glucose co-transporter gene SGLT1. *The Journal of biological chemistry*. 269: 15204-9.

[9] Turk E, Wright EM. (1997). Membrane topology motifs in the SGLT cotransporter family. *J Membr Biol*.159:1-20.

[10] Chasman D, Adams RM (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 307: 683–706.

[11] Wärde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. (2010)The Gene MANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 1; 38 Suppl: W214-20.

[12] Ng PC, HenikoffS. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 31: 3812-3814.

[13] Adzhubei I, Jordan D M, Sunyaev SR, (2013) Predicting Functional Effect of Human Missense Mutations Using Poly Phen-2. *Human Genetics*. 76: 7.20-1-7.20.41.

[14] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PloS ONE* 7 (10): e46688.

[15] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*. 30: 1237-1244.

[16] Capriotti E, Altman RB, Bromberg Y. (2013) Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics*. 14 Suppl 3:S2.

[17] Capriotti E, Fariselli P, Casadio R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33 (Web Server issue): W306-W310.

[18] Cheng J., Randall A., and Baldi P. (2005). Prediction of Protein Stability Changes for Single Site Mutations Using Support Vector Machines. *Proteins: Structure, Function, vol*. 62, no. 4, pp. 1125-1132.

[19] Venselaar H, TeBeek TA, Kuipers RK, Hekkelman ML, Vriend G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 8; 11: 548.

[20] Nielsen M., Lundegaard C., Lund O., Petersen TN. (2010) CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Research*, Vol. 38.

[21] Petersen E F, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 25: 1605-12.

[22] Lam, J. T., Martin, M. G., Turk, E., Hirayama, B. A., Bossard, N. U., Steinmann, B., & Wright, E. M. (1999). Missense mutations in SGLT1 cause glucose-galactose malabsorption by trafficking defects. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1453 (2), 297–303.