Bioresearch Communications
Volume 05, Issue 02, July 2019
Journal Homepage: www.bioresearchcommunications.com

Original Article

A Computational Study of the Functional and Structural Impact of Deleterious SNPs in the BCL11B Gene

Tanima Sharker[1*] , Liza Teresa Rozario[1] , Md. Reaz Morshed[1]

[1]*Department of Biochemistry and Molecular Biology, Noakhali Science and Technology University, Noakhali-3814*

ABSTRACT: Single nucleotide polymorphisms (SNPs) play a significant role in analyzing the genetic basis of complex human diseases. Deleterious SNPs cause changes in the amino acid residues and therefore, are important factors in contributing structural and functional diversity of the encoded protein. The specific functions of BCL11B gene have not been determined yet, but the encoded protein is known to be a key regulator of both differentiation and survival of T-lymphocytes with potential impact on T-ALL (T-cell acute lymphoblastic leukaemia). In this study, a sequence and structure-based computational analysis was used to sort out deleterious SNPs of BCL11B gene to understand its function and possible involvement in carcinomas and other diseases. The analysis identified 3 deleterious mutations (P697L, R620C, T243M) in the coding region of BCL11B gene. However, to establish their role in the pathogenesis of disease, further studies should be done on the deleterious mutations of BCL11B gene.

Keywords: Single nucleotide polymorphisms (SNPs), Deleterious SNPs/ mutations, BCL11B gene, T-ALL, Computational analysis.

Article History
Received: 24 February 2019
Accepted: 21 May 2019

Scan the QR code to see the online version or,visit- www.bioresearchcommunications.com

Corresponding author
Tanima Sharker
Department of Biochemistry and Molecular Biology, Noakhali Science and Technology University, Noakhali-3814
Email: tanimajoti@gmail.com
Mobile: +880 1681022733

Citation: Tanima Sharker , Liza Teresa Rozario , Md. Reaz Morshed; A Computational Study of the Functional and Structural Impact of Deleterious SNPs in the BCL11B Gene. Biores Comm. V5-(2). (758-764)

INTRODUCTION

BCL11 (B-cell lymphoma/leukemia 11) gene family constitutes a superfamily of C2H2 zinc-finger transcription factors which are involved in hematopoietic malignancies and lymphopoiesis [1,2,3]. BCL11A and BCL11B are the members of BCL11 gene family and located on chromosome 14q32.3 and 14q32.1 respectively[1].Chromosomal aberrations of *BCL11A* at 2p16.1 have been reported in a variety of B-cell malignancies and its deficiency in mice leads to a profound block in B-cell development[3]. BCL11A may be involved in lymphoid malignancies through either chromosomal translocation or amplification. BCL11A is highly homologous to BCL11B gene (67% identical at the nucleotide level and 61% identical overall at the protein level). BCL11B, like BCL11A, contain 6 C2H2 zinc fingers and proline rich and acidic regions with 95% identity in the zinc finger domains and is remarkable for having a large 59 CpG island. Similarly, it is also implicated to be involved in human lymphoids and other cancers as BCL11A[1]. It is a known regulator of early thymocyte development and plays an important role in supporting the commitment of thymic progenitors to the T cell lineage. It is expressed in thymus, peripheral lymphoid

organs including spleen and lymph nodes. It is also expressed in peripheral blood lymphocytes, mature naive and activated CD4 T lymphocytes, as well as in the human CD4 T-cell line Jurkat. BCL11B participates in the control of the interleukin-2 (IL2) gene expression following activation through T-cell receptor (TCR). Moreover, it associates with the p300 coactivator in CD4 T cells activated through TCR, which may account for its transcriptional activation function. CCR7 and CCR9 receptors play an important role in directing the movement of progenitors in mammals from the bone marrow to the thymus. By modulating the expression of these receptors, BCL11B crucially regulates the responsiveness of hematopoietic stem cells to chemotactic signals[4]. BCL11 proteins act as transcriptional repressor and the endogenous BCL11B with the nucleosome remodeling and histone deacetylase (NuRD) complex, is one of the key transcriptional corepressor complexes in mammalian cells[5].

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in human genome[6]. They can act as biological markers helping scientists locate genes that are associated with disease. Single-nucleotide polymorphisms may fall within coding sequences of genes, non-coding regions of genes or in the intergenic regions[7,8]. Synonymous SNPs do not affect the protein sequence, while nonsynonymous SNPs change the amino acid sequence of protein and therefore, are important factors in contributing structural and functional diversity of the encoded protein[9,10].

In terms of their disease causing potential, most of the SNPs of BCL11B gene are still uncharacterized. Recently, *in silico* approaches have been widely used to identify the effect of deleterious SNPs in candidate genes by utilizing information such as conservation of residues, structural attributes and physiochemical properties of peptides [11,12,13,14]. The advantages of *in silico* approaches over the lab based characterization are their reliability, convenience, speed and lower cost to find such variants that have the potential impact on the structure and function of bcl11b protein[15]. So, the aim of the present study is to explore the effect of deleterious SNPs on the stability and function of the BCL11B gene. For this study, a set of computational algorithms has been employed to identify the deleterious SNPs in the BCL11B gene.

MATERIALS AND METHODS

Retrieval of SNPs

Data on the human BCL11B gene and its protein sequence (FASTA format) were collected from NCBI (http://www.ncbi.nlm.nih.gov/) and UniProtKB (http://www.uniprot.org/uniprot/). SNPs located in BCL11B gene were retrieved from dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/)[16].

Prediction of the effect of deleterious SNPs

To analyze the functional and structural consequences of deleterious SNPs of BCL11B gene, SIFT, SNAP2, PolyPhen-2, SNPs&GO, PhD-SNP and PANTHER were used sequentially.

Deleterious SNPs carried out amino acid substitution was first screened by SIFT (Sorting Intolerant from Tolerant) server. SIFT (https://sift.bii.a-star.edu.sg/) uses sequence homology based method to speculate the deleterious and tolerated SNPs so that the phenotypic and functional changes upon protein molecule due to amino acid substutions can be analyzed. In our study, we submitted rsIDs retrieved from dbSNP as a query to make prediction. Deleterious SNPs prediction by SIFT server was further used to find their effect on the structure and function of BCL11B gene[17].

PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) was used to search the possible effect of an amino acid substitution on the structure and function of bcl11b protein. Protein sequence, database ID/ accession number and details of amino acids substitution are used as input options for PolyPhen-2[18].

SNAP2 (Screening for Nonacceptable Polymorphisms) (https://www.rostlab.org/services/SNAP/) predicts functional effects of mutations. Taking a variety of sequence and variant features into account, it differentiates between effect and neutral variants. FASTA format of protein sequences is used as input option for SNAP2[19].

In addition, we used SNPs&GO ((http://snps.biofold.org/snps-and-go/snps-and-go.html), PhD-SNP (Predictor of human deleterious single nucleotide polymorphisms) and PANTHER (Protein

Analysis Through Evolutionary Relationships) tools to screen the disease-related SNPs. The protein sequence in FASTA format is the input option for SNPs&GO server[20]. The server also provides the outcome for PhD-SNP[21] and PANTHER[22] algorithms.

Conserved residues prediction by ConSurf

ConSurf (http://consurf.tau.ac.il/2016/) uses an empirical Bayesian inference to analyze evolutionary conservation of amino acid substitutions in protein based on the phylogenetic relations between homologous sequences. The FASTA sequence of BCL11B protein is provided to ConSurf tool. Along with the color scheme, it gives conservation score (Score 9 means most conserved amino acid whereas 1 means variable amino acid)[23].

Prediction of Protein stability change by I-Mutant2.0 and NetSurfP

I-Mutant2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) is a neural-network-based web server for the automatic prediction of protein stability

changes upon deleterious SNPs. I-Mutant2.0 can classify the signs of the protein stability changes due to a variation and it can predict the relative change in Gibbs-free energy (ΔG) at a given temperature[24].

NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP/) server predicts the surface accessibility, secondary structure, disorder, and phi/psi dihedral angles of amino acids in an amino acid sequence. To speculate the secondary structure, surface, and solvent accessibility of amino acids, the FASTA sequence of BCL11B protein was submitted to NetSurfP. The prediction of this server states buried (low accessibility), partially buried (moderate accessibility) and exposed (high accessibility) region in protein structure[25].

RESULT AND DISCUSSION

To identify SNPs responsible for specific phenotypes with molecular approaches seems to be expensive and time-consuming[26]. Therefore, *in silico* approaches can help in narrowing down the number of missense mutations to be screened in genetic association studies and for a better understanding of the functional and structural aspects of the parent protein. Previous studies using computational analysis helped in predicting the functional nonsynonymous SNPs associated with BCL11A gene[27]. So, in our study we used different *in silico* tools to identify the deleterious SNPs and their impact on BCL11B gene.

Retrieval of SNPs

All the SNPs of BCL11B gene (Gene ID: 64919) were retrieved from dbSNP database. A total of 24625 SNPs was found in human BCL11B gene out of which only 940 were reported in the exonic region. Among them 417 were coding synonymous. We selected rest of the 523 SNPs which are coding nonsynonymous, 3' splice site, 5' splice site and frameshift variants for our enquiry. Nonsynonymous SNPs change the amino acid sequence of a protein which can contribute the structural and functional diversity of the encoded protin[9,10]. For gene functioning appropriate splicing is crucial and its disruption may be strongly deleterious. So, SNPs in splice site may distort the actual functions of the protein[28]. And the frameshift variants are recognized to play an important role in some genetic diseases[29].

Prediction of deleterious SNPs

Prediction of Phenotypic Impacts by SIFT

SIFT predicts whether an amino acid substitution affects protein function so that the substitutions can be prioritized for further study[30]. It speculates whether the substitution is deleterious or tolerated. The SIFT values less than 0.05 are deleterious, those greater than or equal to 0.05 are tolerated[17]. 523 rsIDs were submitted for SIFT input and it predicted 21 substitutions as tolerated and 6 as deleterious as shown in (Table 1). These 27 substitutions were nonsynonymous SNP (nsSNP) and more specifically missense mutation (obtained data from dbSNP database). Remaining rsIDs were not found by SIFT server.

Simulation of Functional Consequences by PolyPhen-2

PolyPhen-2 predicts the possible effect of amino acid substitution on function and structure of a protein based on several criteria like phylogenetic, structural information and sequence of protein. The PolyPhen-2 score ranges from 0.0 (tolerated) to 1.0 (deleterious) so values nearer 1.0 are more confidently predicted to be deleterious[18]. Out of the 27 nsSNPs (obtained from SIFT), 7 (26%) deleterious nsSNPs were speculated as probably damaging (more confident prediction), 9 (33.3%) as possibly damaging (less confident prediction) and the remaining 40.7% as benign by PolyPhen-2. Therefore, this classification of SNPs based on PolyPhen-2 scores allow us to assess the potential quantitative effect of SNPs on wild type protein.

By comparing different *in silico* tools, SIFT and PolyPhen-2 were reported to have better performance in identifying deleterious nsSNPs[29]. The accuracy of SIFT and PolyPhen-2 was further validated, which makes these tools more applicable for the prediction[30].

Functional Significance of Substitution by SNAP2

SNAP2 anticipates whether the impact of amino acid substitutions is neutral or not. The output of this server predicts a score (ranges from −100 strong neutral prediction to +100 strong effect prediction) that reflects the likelihood of specific mutation to alter the native protein function with expected accuracy[19].

The findings from SNAP2 server indicated 14 (52%) of the variants were significant and rest of them 48% were insignificant (Table 1). So, these 14 variants can alter the function of the native protein.

From the outcomes of these 3 servers SIFT, PolyPhen-2 and SNAP2, it was concluded that 7 deleterious nsSNPs with rsIDs of rs145770156, rs200835381, rs367645713, rs368885255, rs372020796, rs372022855 and rs376769895 were significant (Table 1) and thereby the result was refined. For further validation and authentication of this result additive *in silico* tools were used in the present study.

Functional Characterization by SNPs&GO, PhD-SNP, PANTHER

All the 3 algorithms envisage whether the amino acid substitution is disease associated or not. SNPs&GO can predict whether a variation is disease associated or not by exploiting the corresponding protein functional annotation. The probability scores higher than 0.5 indicates the disease related effect of mutation on the parent protein function[20]. The findings from SNPs&GO indicated that among the 7 deleterious nsSNPs, 6 nsSNPs are disease associated whereas the

prediction from PhD-SNP and PANTHER specified 3 of the nsSNPS as disease related (Table 2).
In this study, the results of the SIFT, PolyPhen-2, SNAP2, SNPs&GO, PhD-SNP and PANTHER algorithms were combined to prioritize the damaging nsSNPs and increase the accuracy of the analysis. By combining the speculation of these bioinformatical algorithm, 3 (rs200835381, rs372020796, rs376769895) out of 7 deleterious nsSNPs were found to be more damaging and disease instigating.

Table 1. Prediction of the effect of SNPs by using SIFT, SNAP-2 and PolyPhen-2 server.

SNP	AMINO ACID CHANGE	SIFT PREDICTION	PolyPhen-2	SNAP-2 PREDICTION
rs145770156	P98L	Tolerated	Probably Damaging	Effect
rs200835381	R620C	Deleterious	Probably Damaging	Effect
rs367645713	P267L	Deleterious	Probably Damaging	Effect
rs368885255	P184L	Deleterious	Probably Damaging	Neutral
rs372020796	P697L	Tolerated	Probably Damaging	Effect
rs372022855	M47T	Tolerated	Benign	Neutral
rs376769895	T243M	Tolerated	Probably Damaging	Effect

Table 2. Prediction of the effect of SNPs by using SNPs&GO, PhD-SNP and PANTHER server.

SNP	AMINO ACID CHANGE	SNPs&GO	PhD-SNP	PANTHER
rs145770156	P98L	Disease	Neutral	Neutral
rs200835381	R620C	Disease	Disease	Disease
rs367645713	P267L	Disease	Neutral	Neutral
rs368885255	P184L	Disease	Neutral	Neutral
rs372020796	P697L	Disease	Disease	Disease
rs372022855	M47T	Neutral	Neutral	Unknown
rs376769895	T243M	Disease	Disease	Disease

Prediction of conserved amino acid residues by Consurf

In ConSurf, the evolutionary rate is estimated based on the evolutionary connection between the protein and its homologues and considering the similarity between amino acids as reflected in the substitutions matrix[23]. In this study, bcl11b protein was found in conserved region. The sequence alignment showed that R620C mutant was present in highly conserved region whereas P697Land T243M were present in conserved and average region respectively (Figure 1).

Prediction of protein stability change and solvent accessibility by I-Mutant2.0 and NetSurf P

To envisage the change of protein stability due to mutation, the selected 3 deleterious nsSNPs were tested by I-Mutant2.0 server. The output is either a free energy change value (ΔΔG) of protein after and before mutation or the sign of DDG. Positive DDG value (DDG>0) leads to increased stability whereas negative DDG value (DDG<0) indicates decreased stability[22]. The variants R620C and T243M showed negative DDG values and were considered to be less stable whereas P697L showed a positive DDG value and was referred to have more stability (Table 3).
From primary sequence information NetSurfP predicts protein surface accessibility and secondary structure. The variants R620C, P697L and T243M were evaluated for solvent accessibility and stability by using the server NetSurfP. All these 3 deleterious mutants were found to be exposed (Table 3).
Solvent accessibility and hydrophobicity play important roles in the structure and functions of proteins[33]. In proteins, polar side chains tend to be exposed to the solvent whereas hydrophobic residues tend to be buried in the interior of the protein away from the solvent [34,35]. The stability of proteins increases when the area of water-accessible hydrophobic surface reduces[36]. All these 3 deleterious

nsSNPs and their respective wild variants are exposed to the surface area. R620C entails a substitution of arginine (basic charged amino acid) to cysteine (polar uncharged amino acids). The loss of charge in the mutant variant may have impact on the protein function. In P697L, proline is substituted by leucine where both amino acids are non-polar in nature. However, T243M entails the substitution of threonine (polar amino acid) by methionine (non-polar amino acid). The exposure of the non-polar residue in place of the polar one on the surface area may decrease protein stability. Furthermore, this residue involves in helix formation. Because of the polar hydroxyl (-OH) group, threonine engages in H-bond formation. H-bond is crucial for stabilizing the secondary structure of protein. The substitution of threonine by methionine thus may disrupt the secondary structure in the mutated protein.

Owing to the inaccessibility of the PDB ID (Protein Data Bank ID) and 3D structure of bcl11b protein, the actual impact of these amino acid substitutions on protein activity and stability is still unfathomable.

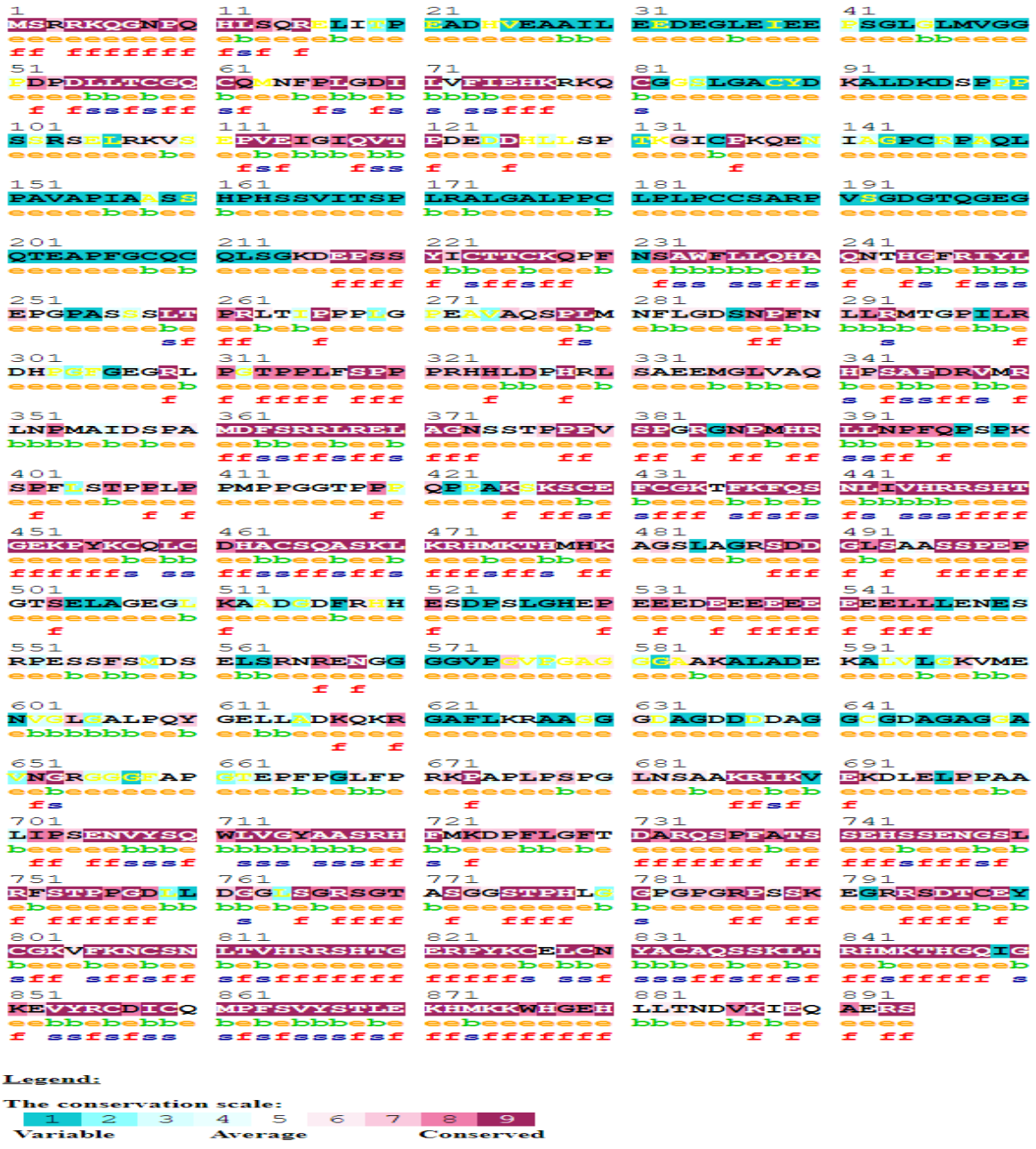

Figure 1. Unique and conserved amino acids in BCL11B protein were predicted by ConSurf. Amino acids were ordered based on a conservation scale of 1–9 and highlighted as follows: blue residues (1–4) are variable, white residues (5) are average, and purple residues (6–9) are conserved. (e) Exposed residues are colored via an orange letter. (b) Buried residues are marked via a green letter. (f) Putative functional highly conserved and exposed residues are revealed with a red letter. (s) Predicted structural residues which are highly conserved and buried are indicated via blue letter.

CONCLUSION

Although the specific functions of BCL11B gene are not established exactly, from some recent studies it has been reported that it plays an important role in the differentiation as well as survival of T cells. Almost all the SNPs of this gene especially those with disease causing potency is uncharacterized yet. So, this study is of great value as it is the first attempt to signify the functional and structural impact of deleterious nsSNPs on BCL11B gene using different computational algorithms. It differentiates disease causing mutations from neutral ones as listed in SNP database. 3 deleterious nsSNPs (P697L, R620C, and T243M) in the coding region of BCL11B gene were identified in the present study where the P697L and R620C were found to be more potential due to their propensity to cause diseases. Furthermore, the predicted disease associated nsSNPs can be studied to establish their role in different disease development as well as potent drug discovery.

REFERENCE

1. Satterwhite, E., Sonoki, T., Willis, T.G., Harder, L., Nowak, R., Arriola, E.L., Liu, H., Price, H.P., Gesk, S., Steinemann, D., Schlegelberger, B., Oscier, D.G., Siebert, R., Tucker, P.W., Dyer, M.J. 2001 December. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood*. 98 (12): 3413–20.
2. Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., et al. 2008 February. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A.*. 105 (5): 1620–5.
3. Liu, H., Ippolito, G.C., Wall, J.K., Niu, T., Probst, L., Lee, B.S., Pulford, K., Banham, A.H., Stockwin ,L., Shaffer, A.L., Staudt, L.M., Das, C., Dyer, M.J. and Tucker, P.W. 2006 May 16. Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Mol Cancer.* ;5:18.
4. Cismasiu VB et al. 2006 Oct 15. BCL11B participates in the activation of *IL2* gene expression in CD4_ T lymphocytes. *Blood.* ;108(8):2695-702.
5. Cismasiu VB[1], Adamo K, Gecewicz J, Duque J, Lin Q, Avram D. 2005 Oct 13. BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene.* ;24(45):6753-64.
6. Ramensky, V., Bork, P., Sunyaev, S. 2002 Sep 1. Human non-synonymous SNPs*:* server and survey. *Nucleic Acids Res.* 30(17):3894-900.
7. Liu, J., Gough, J., Rost, B. 2006. Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genetics*. 2(4): e29.
8. Carninci P., Kasukawa T., Katayama S., et al. 2005. The transcriptional landscape of the mammalian genome. *Science*. ;309(5740):1559–1563.
9. Chakravarti,A. 2001. To a future of genetic medicine. *Nature*. 409(6822):822-3.
10. Ng, P. C., Henikoff, S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annual Review of Genomics and Human Genetics*. 7:61–80.
11. Sunyaev, S., Ramensky, V., Koch, I., Lathe III, W., Kondrashov, A. S. and Bork, P. Prediction of deleterious human alleles. 2001. *Human Molecular Genetics*, vol. 10, no. 6, pp. 591–597.
12. Wang, Z. and Moult, J. 2001. SNPs, protein structure, and disease. *Human Mutation*, vol. 17, no. 4, pp. 263–270.
13. Grantham, R. 1974. Amino acid difference formula to help explain protein evolution.*Science*, vol. 185, no. 4154, pp. 862–864.
14. Chasman, D. and Adams, R. M. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *Journal of Molecular Biology*, vol. 307, no. 2, pp. 683–706.
15. Goswami, A. M. 2015. Structural modeling and *in silico* analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2. *Meta Gene*. 5, 162–172.
16. Sherry, S.T., Ward, M.-H., Kholodov, M.et al. 2001 DbSNP: theNCBI database of genetic variation. *Nucleic Acids Research*, vol. 29, no. 1, pp. 308–311.
17. Kumar, P., Henikoff, S. and Ng, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, vol. 4, no. 7, pp. 1073–1082.
18. Adzhubei, I. A., Schmidt, S., Peshkin, L. et al. 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, vol. 7, no. 4, pp. 248–249.
19. Bromberg, Y., Yachdav, G., and Rost, B. 2008. SNAP predicts effect of mutations on protein function. *Bioinformatics*, vol. 24, no. 20, pp. 2397–2398.
20. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. and Casadio, R. 2009. Functional annotations improve the predictive score of human disease-relatedmutations in proteins. *Human Mutation*, vol. 30, no. 8, pp. 1237–1244.
21. Capriotti, E., Calabrese, R. and Casadio, R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary

Information. *Bioinformatics*, vol. 22,no. 22, pp. 2729–2734.
22. Moore, D. J., Zhang, L., Dawson, T. M. and Dawson, V. L. 2003. A missense mutation (L166P) in DJ-1, linked to familial Parkinson's disease, confers reduced protein stability and impairs homooligomerization. *Journal of Neurochemistry*, vol. 87, no. 6, pp. 1558–1567.
23. Celniker, G., Nimrod, G., Ashkenazy, H. et al. 2013. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, vol. 53, no. 3-4, pp. 199–206.
24. Capriotti, E., Fariselli, P. and Casadio, R. 2005. I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, vol. 33, no. 2, pp.W306–W310.
25. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. and Lundegaard, C. 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology*, vol. 9, no. 1, article 51.
26. Chen, X. and P. Sullivan. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *The pharmacogenomics journal*. 3(2): p. 77-96.
27. Das, S.S., Chakravorty, N. 2019 Mar 7. Identification of deleterious SNPs and their effects on BCL11A, the master regulator of fetal hemoglobin expression. *Genomics.*
28. Yerbol, Z., Kurmangaliyev, Roman, A., Sutormin., Sergey, A., Naumenko, Georgii, A., Bazykin, Mikhail, S., Gelfand. 2013 Sep 1. Functional implications of splicing polymorphisms in the human genome . *Human Molecular Genetics*. Volume 22, Issue 17, Pages 3449–3459.
29. Zimmerman, P.,A., Buckler, White, A., Alkhatib, G., Spalding, T., Kubofcik, J., Combadiere, C., Weissman, D., Cohen, O., Rubbert, A., Lam, G., Vaccarezza, M, Kennedy, P.,E., Kumaraswami, V., Giorgi, J.,V., Detels, R., Hunter, J., Chopek, M., Berger, E.,A., Fauci, A.,S., Nutman, T.,B., Murphy, P.,M. January 1997. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Molecular Medicine* (Cambridge, Mass.). 3 (1): 23–36.
30. Pauline, C., Ng., and Steven, Henikoff. 2003.SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research. Vol. 31, No. 13.
31. Thusberg, J. and M., Vihinen. 2009. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Human mutation*. 30(5): p. 703-714.
32. Hicks, S., et al. 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Human mutation. 32(6): p. 661-668.
33. Shambhu, Malleshappa, Gowder, Jhinuk, Chatterjee, Tanusree, Chaudhuri, and Kusum Paul. 2014. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *The Scientific World Journal.*
34. M., F., Perutz, J., C., Kendrew, and H., C., Watson. Structure and function of haemoglobin: II. some relations between polypeptide chain configuration and amino acid sequence. *Journal of Molecular Biology*, vol.13, no.3,pp.669–678,1965.
35. M., M., Gromiha. 2010. Protein Bioinformatics: From Sequence to Function, *Academic Press*, New York, NY, USA, 1st edition.
36. C,. Strub, C., Alies, A., Lougarre, C., Ladurantie, J., Czaplicki, and D., Fournier. 2004. Mutation of expose hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry*, vol. 5,article 9, pp. 1–6.