DOI : https://dx.doi.org/10.26808/rs.ph.i8v3.04
International Journal of Pharmaceutical Science and Health Issue 8, Vol.3 (May-June 2018)
Available online on http://www.rspublication.com/ijphc/index.html ISSN 2249 – 5738

The mutations identified in human HDAC-10 gene and their prognosis in cancer cases

Sabeena M[1], Kaiser Jamil*[2], A.V.N Swamy[3]

[1]Department of Biotechnology, Jawaharlal Nehru Technological University Anantapur (JNTUA), Anantapuramu, Andhra Pradesh, India

[2]Center for Biotechnology and Bioinformatics, School of Life Sciences, Jawaharlal Nehru Institute of Advanced Studies (JNIAS), Secunderabad, Telangana, India

[3]Department of Chemical Engineering, Jawaharlal Nehru Technological University Anantapur (JNTUA), Anantapuramu, Andhra Pradesh, India

ABSTRACT

Mutations in genes are the hallmark in cancer development. Most of these mutations are Single-nucleotide polymorphisms (SNPs) which play a major role in complexity of human diseases. Identification of deleterious SNPs from tolerant SNPs is important for characterizing the genetic basis of human diseases like cancer. Numerous SNPs have been found in genes coding for human Histone Deacetylase (HDAC) family proteins but there is not much information on the relationship between the genotype and phenotype of these SNPs in HDAC10 gene and its associated mutations in cancers. This study aims to determine the functional SNPs in HDAC10 gene, which we have analyzed *in silico* by selecting various computational approaches. Various computational methods such as SIFT/PROVEAN, PolyPhen 2.0, I-mutant 3.0, SNP&GO and Mutpred tools were used to determine the most probable mutations that might be associated with HDAC10 gene. Further to understand the atomic arrangement in 3D space, between the native and mutant forms of HDAC-10 gene we modeled these two structures using molecular modeling and energy minimization methods to verify our SNP results. A total of 652 SNPs were initially retrieved from the dbSNP database, which could be segregated as damaging, tolerant and deleterious. Among these a set of 4 mutations that were found to be associated with high risk SNPs namely A103V (rs143228101), G133R (rs145098215), V297D (rs113918787) and G304S (rs138792486) as most deleterious and disease associated. Also, our results showed that V297D and G304S were the most sensitive and high-risk SNPs among them. We present a set of computational methods which helped us to identify deleterious and damaging SNPs in HDAC10 which were associated with complete loss of function of HDAC10 gene. This study also opens an arena of identifying deleterious SNPs in hundreds of target genes, and thereby help identify novel therapeutics.

Key words: HDAC10; Biomarkers; Mutations; SNPs; computational analysis; cancers

Corresponding Author: Dr. Kaiser Jamil

INTRODUCTION

Histones are the basic proteins and its acetylation can be considered as an important factor governing gene expression by its effects on chromatin structure and assembly [1]. There are many studies that reveal the mechanism of histone acetylation and deacetylation and its role in transcription. Histone Deacetylase (HDACs) regulates various cellular processes through enzymatic deacetylation of both histone and nonhistone proteins [2]. In eukaryotes, histone deacetylases superfamily consists of 18 genes which are divided into two families and four classes such as I, II, III and IV. All classes consist of 11 family members, which are referred to as "classical" HDACs, whereas 7 of class III members are called "sirtuins"[3]. Several recent studies have implicated that, individual HDAC enzymes as potential therapeutic targets in the development of cancer. HDAC as a therapeutic target for treatment of endometrial cancer has been reported[4]. The study of genetic variation in genes can elucidate critical determinants in environmental exposure and cancer, which could have future implications for preventive and early intervention strategies [5].

Although many Single nucleotide polymorphisms or SNPs are phenotypically neutral, non-synonymous SNPs (nsSNPs) often have deleterious effects on protein structure or function [6]. In other words, SNPs in protein-coding regions that cause amino acid variants are most likely to affect phenotypes. nsSNPs can alter the structure, stability, or function of proteins, and are often associated with human disease [7]. Some researches describes that approximately 50% of the mutations involved in inherited genetic disorders are due to nsSNPs [8][9]. Identification of SNPs affecting human phenotype, especially leading to risks of complex disorders including cancer is one of the key problems of medical genetics. Hence our study was focused to determine the high risk and potential nsSNPs present in HDAC10 gene.

Using computational approaches, our attempt was to address the problem to predict which of the SNPs of HDAC10 are deleterious to gene function or likely to be disease association. For this study we looked at all the huge datasets in the public repositories and selected the three most commonly used data sets for locating and classifying each SNP. Identification of these deleterious variants could be an important target for structural genomics projects.

MATERIALS AND METHODS

Data Mining for nsSNPs of Human HDAC-10 Gene

The NCBI dbSNP database ((http://www.ncbi.nlm.nih.gov/SNP/ was selected to retrieve the information on all the phenotypes datasets of HDAC-10 SNPs. A total of 652 SNPs were retrieved from the dbSNP database [10]. The information on the HDAC10 protein sequences with id Q969S8 in FASTA format was also retrieved from UniprotKb, which is commonly used as knowledgebase for molecular sequences. Most of the sequences in UniProt KB are derived from the conceptual translation of nucleotide sequences. This program provides a stable, comprehensive, freely accessible central resource on protein sequences and functional annotation [11].

Determination of nsSNPs in Human HDAC-10 Gene by SIFT and PROVEAN

Prediction of the putative effects of nsSNPs on protein function was performed using SIFT and PROVEAN (http://sift.jcvi.org/)[12]. Protein Variation Effect Analyzer (PROVEAN) is a new prediction tool which works for both SNPs and indels. SIFT (Sorting Intolerant From Tolerant) predicts whether an amino acid substitution affects protein function which is based on the degree of conservation of amino acid residues in sequence

alignments derived from closely related sequences. Single submission of all the 652 SNPs data of Human HDAC10 gene from NCBI dbSNP returned its functional predictions. Each line in the input reference SNP data .txt file included rsIds, chromosome location/position and its allele information. The algorithms for the SIFT program uses the updated SWISS-PROT and TrEMBL databases to find related sequences of the query protein [13]. Prediction the functional effects of protein sequence was performed using (PROVEAN) which provided protein sequence variations including single or multiple amino acid substitutions, and in-frame insertions and deletions. [13]

Determination of nsSNPs involved in structural modification by the PolyPhen program

To determine the nsSNPs involved in structural modification of HDAC-10, our query data was analyzed by the PolyPhen-2 (Polymorphism Phenotyping v2) program. Polyphen-2 predicts possible impact of an amino acid substitution on the structure and function of protein using physical and comparative considerations [14]. The query protein and the corresponding amino acid substitutions was submitted to the program SIFT Provean (http://genetics.bwh.harvard.edu/pph2/) to obtained the results.

Prediction of protein stability changes by I-Mutant 2.0

The accuracy of SIFT/PROVEAN and PolyPhen results were validated through the stability changes studies. I-Mutant 2.0 is a Support Vector Machine (SVM) based tool used for the automatic prediction of protein stability change upon amino acid substitution [15]. The protein stability change was determined for HDAC-10 protein sequence Q969S8 using I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html). This software computes the predicted free energy change value/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 positive DDG value indicates that the mutated protein possesses high stability. Likewise, scores <0 are predicted by the algorithm to indicate decreased stability, whereas scores >0 are considered to indicate increased stability. I-Mutant 2.0 clearly stabilizes or destabilizes the protein in 80% of the cases when the 3-D structure is known and 77% of the cases when only the protein sequence is available (http://folding.biofold.org/i-mutant/i-mutant2.0.html)

Determination of disease related mutations by SNPs and Gene Ontology (GO)

HDAC-10 protein sequence Q969S8 was submitted as input to the SNPs and GO algorithms (http://snps.biofold.org/snps-and-go/) server. We used this server to predict the impact of protein variations using functional information encoded by Gene Ontology terms of the three main roots: Molecular function, Biological process, and cellular component *E*. [15]**.**

Validation of harmful mutations by MutPred

MutPred (http://mutpred.mutdb.org/) is the web application algorithm used to validate the harmful mutations present in HDAC10 nsSNPs. MutPred builds on the established SIFT method but offers improved classification accurately based upon protein sequence, and model changes of structural features and functional sites between wild-type and mutant sequences with output of probabilities of gain or loss of structure and function[16]. MutPred uses SIFT, PSI-BLAST, and Pfam profiles along with some structural disorder prediction algorithms such as TMHMM, MARCOIL, I-Mutant 2.0, B-factor prediction, and DisProt. Functional analysis included the prediction of DNA-binding site, catalytic domains, calmodulin-binding targets, and posttranslational modification sites. In addition, it predicts molecular cause of disease. The tool requires a protein sequence, a list of amino acid substitutions, and an email address.

Constructing HDAC10 wild and mutated models for our studies

(i) Modeling of the Native form of HDAC10 protein, and structure evaluation and eenergy minimization

As the 3D structure of HDAC10 protein is not available, we analyzed the HDAC10 sequence (Q969S8) for homology to other proteins. 3D-Jigsaw was used to generate 3D structural model for wild type HDAC10 (Q969S8). The 3D-Jigsaw searches multiple sequence databases (e.g. PFAM and PDB) and builds the structure based on homologues of known structure [17]. We could visualize the predicted model using the SWISS-PDB viewer [18].. The energy minimization process was performed by GROMACS module in SWISS-PDB viewer. The energy-based calculations were conducted at the atomic levels in protein structures and the output energy profile value for the 9 amino acid residues (P28L, A103V, G133R, R196C, A240V, V297D, G304S, R496W and P483L) were recorded. SWISS-PDB viewer uses GROMOS as the default energy minimization method [19].

(ii) Modeling of protein HDAC10 Mutant form, structure evaluation studies and eenergy minimization studies

The mutation of the modeled structure with 9 high-risk nsSNPs was performed by Swiss PDB viewer. This mutant structure also underwent energy minimization process using the GROMOS in SWISS-PDB viewer[20]. The energy values for each 9 mutant positions (P28L, A103V, G133R, R196C, A240V, V297D, G304S, R496W and P483L) are reported based on the fact that the deviation between the two structures was evaluated by their energy values which could affect stability and functional activity. These two models of wild and mutant forms- are our target proteins of HDAC10 which will be used to verify our predicted SNP results, to evaluate the structural and functional aspects of HDAC10.

RESULTS

SNPs retrieval and analysis

The HDAC10 gene investigated in this study had 30 nsSNPs (nonsynonymous), 109 were synonymous SNPs, and 1 nonsense SNP. SIFT, PROVEAN and PolyPhen-2 software's were applied to categorize the nsSNPs into three groups such as (i) damaging, (ii) tolerant (iii) deleterious. SIFT predicted whether an amino acid substitution affected protein function. SIFT software was applied to naturally occurring nonsynonymous polymorphisms or laboratory-induced missense mutations. In the SIFT software analysis, 15 variants were predicted as damaging and the other 15 were predicted as tolerant. In the PROVEAN analysis, 15 variants were predicted as deleterious, but the other 15 were neutral. In the PolyPhen-2 analysis, 16 variants were predicted as probably damaging and 14 were neutral- Table-1. SIFT and PolyPhen were shown to have better performance in identifying functional nsSNPs among other *in-silico* methods [22]. We observed a set of 12 nsSNPs namely rs146669691, rs142006720, rs143228101, rs145098215, rs139503758, rs144393433, rs143033632, rs113918787, rs138792486, rs148078988, rs146379292 and rs61748567 were common in all three of the above categories. Therefore, these nsSNPs are presented as most likely damaging or deleterious (Table-1). Among these, the three nsSNPs which were found to be highly deleterious included - rs146669691, rs145098215 and rs138792486 with SIFT score of 0.00.

DOI : https://dx.doi.org/10.26808/rs.ph.i8v3.04
International Journal of Pharmaceutical Science and Health Issue 8, Vol.3 (May-June 2018)
Available online on http://www.rspublication.com/ijphc/index.html ISSN 2249 – 5738

Table1: SNP analysis using SIFT, PROVEAN and PolyPhen-2 software's: The 12 nsSNPs of HDAC10 which were found to be significant are highlighted as bold.

S.No	AA Substitution	dbSNP Id	SIFT Prediction (Score)	Provean Prediction (Score)	PolyPhen-2 Prediction (Score)
1.	M1T	rs142968114	Damaging	Neutral	Probably Damaging
2.	T12M	rs147929426	Tolerated	Deleterious	Probably Damaging
3.	P28L	**rs146669691**	Damaging	Deleterious	Probably Damaging
4.	E67K	**rs142006720**	Damaging	Deleterious	Probably Damaging
5.	L71V	rs12169695	Tolerated	Neutral	Benign
6.	K80E	rs148308549	Tolerated	Neutral	Benign
7.	A85V	rs186782930	Tolerated	Neutral	Benign
8.	A103V	**rs143228101**	Damaging	Deleterious	Probably Damaging
9.	A107T	rs149121852	Tolerated	Neutral	Benign
10.	G133R	**rs145098215**	Damaging	Deleterious	Probably Damaging
11.	A139V	rs115099639	Tolerated	Neutral	Benign
12.	D185N	rs34437225	Tolerated	Deleterious	Benign
13.	R196C	**rs139503758**	Damaging	Deleterious	Probably Damaging
14.	R216Q	rs117636118	Tolerated	Neutral	Benign
15.	A240V	**rs144393433**	Damaging	Deleterious	Probably Damaging
16.	A250V	**rs143033632**	Damaging	Deleterious	Probably Damaging
17.	V297D	**rs113918787**	Damaging	Deleterious	Probably Damaging
18.	G304S	**rs138792486**	Damaging	Deleterious	Probably Damaging
19.	M333V	rs141840429	Tolerated	Neutral	Benign
20.	C336R	**rs148078988**	Damaging	Deleterious	Probably Damaging
21.	T418M	rs112311672	Damaging	Neutral	Probably Damaging
22.	R446W	**rs61748567**	Damaging	Deleterious	Probably Damaging
23.	P483L	**rs146379292**	Damaging	Deleterious	Probably Damaging
24.	R496W	rs61748567	Damaging	Deleterious	Probably Damaging
25.	R532S	rs144296501	Tolerated	Neutral	Benign
26.	E535D	rs149579149	Tolerated	Neutral	Benign
27.	A536T	rs138168321	Tolerated	Neutral	Benign
28.	V550M	rs61748566	Tolerated	Neutral	Benign
29.	E611Q	rs13054930	Tolerated	Neutral	Benign
30.	N612D	rs75596977	Tolerated	Neutral	Benign

Protein Stability prediction studies

To study the changes in the protein stability of the missense variants, we used I-Mutant 2.0 software. This tool was developed to test 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 [21]. This software can efficiently predict whether a protein mutation affects the stability of the protein structure or not, based on the principle - the more negative the DDG value, the less stable the mutation. We found that the I-Mutant 2.0 server predicted 21 nsSNPs that were less stable, further we observed that 2 nsSNps namely rs146669691 and rs144393433 from the 12 shortlisted SNPs were found to be more stable 1.23 and 1.17 respectively

To examine how the sequence level affects the functions upon mutation, we analyzed the native form of HDAC10 using MutPred. Our output result for these 21 nsSNP's from MutPred contained probability values of deleterious mutation and top five property scores (*p*), where *p* is the P-value on which certain structural and functional properties are impacted and actionable hypotheses which are referred to as confident hypotheses. To classify the shortlisted nsSNPs, we used SNPs & GO program to predict the disease related mutations from protein sequences with a scoring accuracy of 82% and Matthews correlation coefficient of 0.63. The SNPs & Gene Ontology program is based on Support Vector Machine (SVM). For SNPs&GO, FASTA sequence of whole protein is an input option and output is the predicted results, which are based on the discrimination among disease related and neutral variations of protein sequence. SNPs&GO program collects, in a unique framework, information derived from protein sequence, protein sequence profile and protein functions. The probability scores higher than 0.5 shows the disease related mutations on the parent protein function [16].

Combining the scores from of all five computational programs, the accuracy of prediction was filtered to most disease-associated 9 mutations from the initial list of 12 namely P28L, A103V, G133R, R196C, A240V, V297D, G304S, R496W and P483L. To understand the molecular and structural/functional behavior of the disease associated 9 mutations, we performed energy minimization process using wild and mutant HDAC10; hence the modeling and energy minimization of protein structural information was necessary for absolute understanding of its functionality. We also observed notable increase in the energy minimization values for the four residues namely (A103V, G133R, V297D and G304S). This finding was further supported by the careful investigation of all our predicted results Table 2 (Figure 1 and 2)

Table 2: Summary of four significant nsSNPs determined by various software tools

S.No	Mutation	SIFT Prediction (Score)	Provean Prediction (Score)	PolyPhen-2 Prediction (Score)	I-Mutant	MutPred
1.	A103V	Damaging	Deleterious	Probably Damaging	Decrease	Gain of helix Loss of catalytic residue at A103 Gain of Loop
2.	G133R	Damaging	Deleterious	Probably Damaging	Decrease	Gain of MoRF binding Loss of methylation Loss of sheet
3.	V297D	Damaging	Deleterious	Probably Damaging	Decrease	Loss of methylation Loss of catalytic residue

DOI : https://dx.doi.org/10.26808/rs.ph.i8v3.04
International Journal of Pharmaceutical Science and Health Issue 8, Vol.3 (May-June 2018)
Available online on http://www.rspublication.com/ijphc/index.html ISSN 2249 – 5738

						Loss of sheet Loss of helix Loss of stability
4.	G304S	Damaging	Deleterious	Probably Damaging	Decrease	Loss of catalytic residue Gain of phosphorylation Gain of disorder Loss of helix Loss of sheet

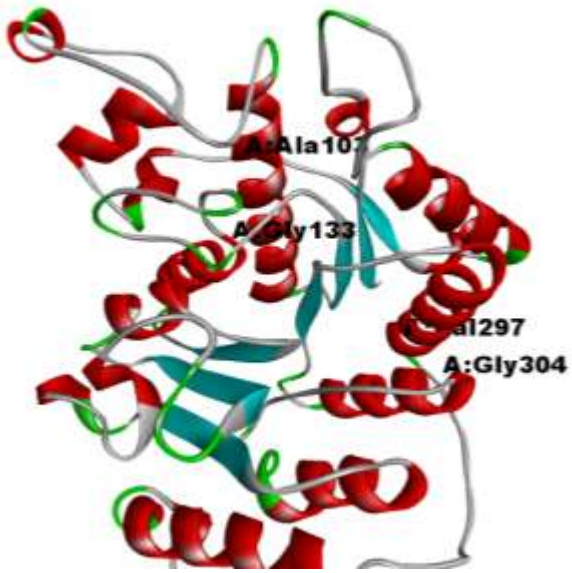

Fig 1: Shows the wild type with four high risk SNPs at positions 103, 133, 279 and 304

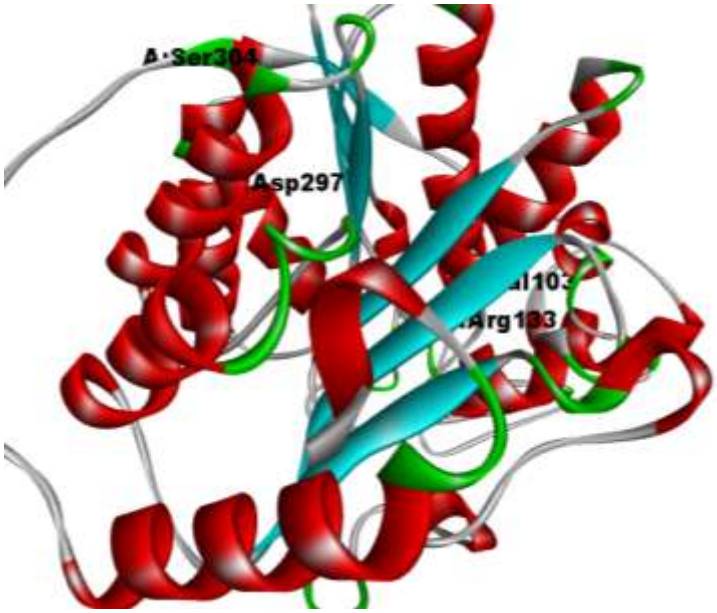

Fig 2: Shows the mutant type with four high risk SNPs at positions 103, 133, 279 and 304

DISCUSSION

Understanding functional roles of mutation has impacted as biomarkers in identifying gene susceptibility and disorder studies [21]. This study first identified the effects of HDAC-10 residues and its expressing mutations by categorizing them as most likely damaging or deleterious. These results showed the pathological nature of the residues. It is important to understand the disease progression mechanism in cancer, thus driver mutations studies are significant [22]. Next, the protein stability prediction approach allowed us to draw several conclusions regarding the nature of mutations. From these results, we observed that there was a significant relation between all the results obtained by the analytical methods using I-Mutant, Mutpred and molecular modeling-based energy minimization studies. In addition to these findings, we could also identify several HDAC10 residues which took part in many of diseases or underwent post-translational modifications. It was clearly observed that there was a loss of catalytic residue at A103V, loss of sheet and loss of methylation at G133R. Loss of methylation, loss of catalytic residue, loss of sheet, loss of helix and loss of stability were identified at V297D. Loss of catalytic residue, gain of phosphorylation, gain of disorder, loss of helix, and loss of sheet were observed at G304S. The above results indicated that V297D and G304S were the most sensitive and high risk nsSNP among the SNPs studied. It was further indicated that native structure had more flexibility than mutant structures and the predicted (A103V, G133R, V297D and G304S) mutations. Analysis of nonsense SNPs and its effects showed that a coding nonsense SNP with dbSNP id rs11553698 (E256*) due to a nucleotide change from G to T was detected as a truncated protein arising mainly due to nonsense or frameshift variations and may exhibit reduced functional activity. Thus, our results demonstrated that some of the predicted nsSNPs in HDAC10 gene may be deleterious to its structure and function. Most of these high-risk nsSNPs were located at highly conserved amino acid sites in a protein-protein interaction module. This study is the first systematic and comprehensive *in silico* analysis of functional SNPs in the HDAC-10 gene.

CONCLUSION

It is concluded that Meta-analysis of all available Single Nucleotide Polymorphisms of HDAC10 determined the disease associated nsSNPs. The computational platforms utilized included sequence-based conservation profile, homology-based structure profile information, and support vector algorithm. The highly deleterious SNPs were identified using multiple *in silico* computational methods like SIFT/PROVEAN, PolyPhen 2.0, I-mutant 3.0, SNP&GO and MutPred,. These methods showed greater accuracy for the prediction of most disease-associated mutations in HDAC gene. This study is the first extensive *in silico* analysis of the HDAC10 gene. A total of 12 high-risk nsSNPs were shortlisted, including four that were not only high risk nsSNPs but most sensitive which could predict their role in disease as well. Further, we laid the platform of *in silico* methods which can predict nsSNPs of not only HDAC10 genes, but the same could be used for any gene, and would be useful for genotype-phenotype studies on new lead molecules development and its clinical response in cancers.

REFERENCES

[1] Kuo MH, Allis CD. Roles of histone acetyltransferases and deacetylases in gene regulation. Bioessays 1998;20:615–26.
[2] Delcuve GP, Khan DH, Davie JR, Groth A, Rocha W, Verreault A, et al. Roles of histone

deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. Clin Epigenetics 2012;4:5.

[3] Witt O, Deubzer HE, Milde T OI. HDAC family: What are the cancer relevant targets? Cancer Lett 2009;277:8–21.

[4] Hagelkruys A, Sawicka A, Rennmayr M, Seiser C. The Biology of HDAC in Cancer: The Nuclear and Epigenetic Components, 2011, p. 13–37.

[5] Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. Br J Cancer 2004;90:747–51.

[6] Ramensky V. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894–900.

[7] Kelly JN, Barr SD, Ramensky V, Radivojac P, Vacic V, Haynes C, et al. In Silico Analysis of Functional Single Nucleotide Polymorphisms in the Human TRIM22 Gene. PLoS One 2014;9:e101436.

[8] Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang S-P, et al. A Catalog of Neutral and Deleterious Polymorphism in Yeast. PLoS Genet 2008;4:e1000183.

[9] Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, et al. Identification, analysis, and prediction of protein ubiquitination sites. Proteins 2010;78:365–80. doi:10.1002/prot.22555.

[10] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.

[11] Magrane M, UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011;2011:bar009. doi:10.1093/database/bar009.

[12] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4. doi:10.1093/nar/gkg509.

[13] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP, Altshuler D, et al. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One 2012;7:e46688. doi:10.1371/journal.pone.0046688.

[14] 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;7:248–9.

[15] Capriotti E, Fariselli P, Rossi I, Casadio R, Prevost M, Wodak S, et al. A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics 2008;9:S6.

[16] Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009;25:2744–50. doi:10.1093/bioinformatics/btp528.

[17] Bates PA, Kelley LA, MacCallum RM, Sternberg MJ. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 2001;Suppl 5:39–46.

[18] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014;42:W252-8.

[19] Petrov D, Margreitter C, Grandits M, Oostenbrink C, Zagrovic B. A systematic framework for molecular dynamics simulations of protein post-translational modifications. PLoS Comput Biol 2013;9:e1003154.

[20] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein

homology-modeling server. Nucleic Acids Res 2003;31:3381–5.

[21] Ben-Nissan G, Chotiner A, Tarnavsky M, Sharon M. Structural Characterization of Missense Mutations Using High Resolution Mass Spectrometry: A Case Study of the Parkinson's-Related Protein, DJ-1. J Am Soc Mass Spectrom 2016;27:1062–70.

[22] McFarland CD, Korolev KS, Kryukov G V., Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci U S A 2013;110:2910.