

RESEARCH ARTICLE

Insights on Impact of Missense SNPs of Human USH2A associated with Usher Syndrome II

K. Anbarasu*, A. Mahjabeen, Radha Mahendran

Department of Bioinformatics, School of Life Sciences, VELS University, Chennai 600 091, Tamil Nadu, India.

*Corresponding Author E-mail: anbarasuk.sls@velsuniv.ac.in

ABSTRACT:

Usher syndrome, also known as Hallgren syndrome, is an extremely rare genetic disorder caused by a mutation in any one of at least 11 genes resulting in a combination of hearing loss and visual impairment. It is a leading cause of deaf blindness and is at present incurable. Usher syndrome is classed into three subtypes according to onset and severity of symptoms. All three subtypes are caused by mutations in genes involved in the function of the inner ear and retina. Usher II are generally hard-of-hearing rather than deaf, and their hearing does not degrade over time; moreover, they generally have a normal vestibular system. Usher syndrome type II occurs at least as frequently as type I, but because type II may be under diagnosed or more difficult to detect, it could be up to three times as common as type I. Usher syndrome type II may be caused by mutations in any of three different genes: USH2A, GPR98, and DFNB31. The protein encoded by the USH2A gene, usherin, is located in the supportive tissue in the inner ear and retina. Usherin is critical for the proper development and maintenance of these structures, which may help explain its role in hearing and vision loss. In this study we investigate the most deleterious and disease associated SNPs in USH2A by using PolyPhen, SIFT, PANTHER, I-mutant 2.0, PhD-SNP, SNP&GO, Pmut, and Mutpred tools. The results showed mutation C759F and C536R in USH2A gene found to be most deleterious SNPs among the dataset. Our study will facilitate wet-lab researches to develop a potent drug therapies against these USH2A mutant patients for Usher syndrome II treatment.

KEYWORDS: Usher syndrome III, Usherin, SNP, Mutation, Deleterious.

INTRODUCTION:

Usher syndrome (USH) is an autosomal recessive disorder characterized by deafness and blindness owing to sensor neural hearing loss and progressive pigmentary retinopathy. Usher syndrome is clinically and genetically heterogeneous. Clinically, two major forms of Usher syndrome have been established (USH1 and USH2), one characterized by congenital, severe to profound hearing loss and absence of vestibular function, and a second in which the congenital hearing loss is less severe and vestibular function is normal¹. Usher syndrome is heterogeneous, which means that different mutated genes can cause the same phenotype².

Linkage studies have shown genetic heterogeneity in Usher syndrome and five distinct loci have been mapped for USH1 and USH2³. The clinical picture of Usher syndrome type II is complicated because subtle variations within the Usher II hearing phenotype have been observed in several Dutch studies. In the first, three of 13 type II patients had a mild but definite progression of hearing loss unrelated to pres by cusic; these three families showed linkage to USH2A and not to USH3⁴. Mutations in the USH2A gene on chromosome 1q41 appear to be responsible for most cases of Usher type II.7 The gene has 21 exons and codes for usherin, a novel protein whose structure is partially homologous with the lam-inin protein group. The function of usherin is not yet understood but it has been postulated that it is a cell adhesion molecule or forms part of the basement membrane. The genomic structure of USH2A has been determined and several new USH2A mutations have recently been found⁵.

There is extensive genetic heterogeneity of Usher syndrome. Usher II was originally found to be linked to markers on chromosome 1⁶. Subtle phenotypic variations within the classic Usher II phenotype were observed within a group of 23 affected Dutch patients from 10 different families (all included in this study). Serial audiograms were available for 13 of the 23 patients, and subsequent analysis showed three of these 13 patients with a mild but definite progression of hearing loss⁷. The majority of patients with Usher syndrome usually fall into one of these clinically distinct categories. Usher syndrome type I (USH1) patients have profound hearing loss and vestibular dysfunction from birth. In addition, night blindness appears earlier in life than USH2 patients, who tend to have less severe hearing loss and normal vestibular function⁸.

MATERIALS AND METHODS:

Dataset:

Human USH2A genes were collected from OMIM and Entrez gene on National Center for Biotechnology Information (NCBI) Website and used as dataset. The corresponding SNPs of USH2A was obtained from dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>). The protein information of usherin was retrieved from the Uniprot database with SwissProtId: O75445).

SIFT:

SIFT, Sorting Intolerant From Tolerant method predicted the deleterious and tolerated SNPs for analyzing the impact of single amino acid substitution (<http://sift.jcvi.org/>). The algorithm of SIFT depends on sequence homology that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. The amino acid position was important in protein function based on conserved residues in sequence alignment and unimportant amino acid positions with divergence. SIFT consists of four step prediction procedure (i) searches for similar sequences (ii) chooses closely related sequences share similar function to the query sequence (iii) obtains the alignment of these chosen sequences (iv) calculates normalized probabilities for all possible substitutions from the alignment. The prediction of SNPs with normalized probability value less than 0.05 are deleterious and greater than or equal to 0.05 are tolerated^{9, 10}. The rs ID of human USH2A SNPs was given as query in batch mode.

POLYPHEN:

PolyPhen-2 (Polymorphism Phenotyping v2), a tool that predicted possible impact of an amino acid substitution on the structure and function of a query protein using physical and comparative approach (<http://genetics.bwh.harvard.edu/pph2/>). PolyPhen was

based on combination of sequence and structure based attributes and uses naive Bayesian classifier for the identification of an amino acid substitution and the impact of mutation. The current version of the tool was more advantages (i) High quality multiple sequence alignment pipeline (ii) Probabilistic classifier based on machine-learning method (iii) Optimized for high-throughput analysis of the next-generation sequencing data. The output of tool predicted the probably damaging and possibly damaging were classified as functionally significant (≤ 0.5) and the benign level being classified as tolerated (≥ 0.51)¹¹.

I-MUTANT 2.0:

I-Mutant 2.0, a support vector machine server for the automatic prediction of changes in protein stability due to single site mutations (<http://folding.biofold.org/cgi-bin/i-mutant2.0/>). The tool was based on the trained data set ProTherm, the most comprehensive database of experimental data on protein mutations. The tool prediction was evaluate the stability change upon single site mutation starting from the protein structure or from the protein sequence. The predicted free energy change (DDG) value was based on the differential unfolding Gibbs free energy changes between mutant and native proteins (kcal/mol). The reliability index (RI) value was computed only when the sign of stability change was predicted and evaluated from the output of the server¹². The input FASTA sequences of proteins, along with the residual changes, were provided by the server for the analysis of DDG value (kcal/mol).

SNPs and GO:

SNPs and GO, a server predicts human disease-related mutations in proteins with functional annotations (<http://snps-and-go.biocomp.unibo.it/snps-and-go/>). An accurate method based on support vector machines, to predict disease related mutations from the protein sequence, scoring with accuracy=82% and Matthews correlation coefficient=0.63. It collected an unique framework information derived from protein sequence, protein sequence profile, and protein function. The server also implemented PhD-SNP method that take in input different subsets of SNPs & GO's input features. SNP&GO was trained on a set of more than 33000 mutations and tested with cross validation procedure over sets in which similar proteins were kept in the same dataset also for the calculation of the LGO score, as derived from the GO database. The output predicted the query protein mutation was disease or not disease related¹³.

PROVEAN:

PROVEAN, Protein Variation Effect Analyzer tool predicts the impact of an amino acid substitution or indel on the protein function (<http://provean.jcvi.org/index>).

php/). This algorithm allowed balanced separation between the deleterious and neutral substitutions with the help of a threshold value. The prediction steps of PROVEAN involved (i) clustering of BLAST hits is performed by CD-HIT with a parameter of 75% global sequence identity (ii) top 30 clusters of closely related sequences form the supporting sequence set (iii) a delta alignment score is computed for each supporting sequence. The overall accuracy for binary classification of protein variants (deleterious or neutral, classification property is being deleterious) was 79.5% for UniProt human protein variations. A query sequence was provided in FASTA format and the default threshold value was -2.5. A score of < -2.5 indicated that the variant was deleterious, and > -2.5 score was considered as neutral¹⁴.

MutPred:

MutPred, a web server predicts the query amino acid substitution (AAS) whether disease-associated or neutral in human (<http://mutpred.mutdb.org/>). The server also predicts the molecular cause of disease from the given AAS. The server is mainly based on Sorting Intolerant from Tolerant (SIFT) method and a database of 14 gain/loss of from structural and functional properties. The training data set was updated to contain 39,218 disease-associated mutations from HGMD and 26,439 putatively neutral substitutions from Swiss-Prot. The updated version of the server is MutPred1.2 and consists of standard code related to evolutionary conservation. The output of MutPred contains a general score (g), cores with $g > 0.5$ and $p < 0.05$ are referred to as actionable hypotheses, $g > 0.75$ and $p < 0.05$ are referred to as confident hypotheses and $g > 0.75$ and $p < 0.01$ are referred to as very confident hypotheses¹⁵. The output of MutPred showed the prediction of deleterious mutation and also five features included loss of stability, gain of disorder, gain of catalytic residue, loss of helix and gain of loop.

RESULTS AND DISCUSSION:

The most deleterious nonsynonymous SNP (nsSNPs) involved in disease mechanism was determined by using *in silico* analysis. The powerful methods focused on the screening of most deleterious nsSNPs in the query genes. In our study, multiple servers were used in the USH2A gene and flow chart of methodology was shown in Fig 1. Our *in silico* can reveal the use of different algorithms as

powerful servers for prioritizing the deleterious SNPs from the gene dataset. The *in silico* investigation on USH2A gene reported to have a total of 39321 coding SNPs, of which 2227 were nsSNPs (missense) and 252 were synonymous SNPs validated by 1000 Genomes dataset. Out of these SNPs, only 15 nsSNPs non-synonymous coding were selected for further *in silico* analysis.

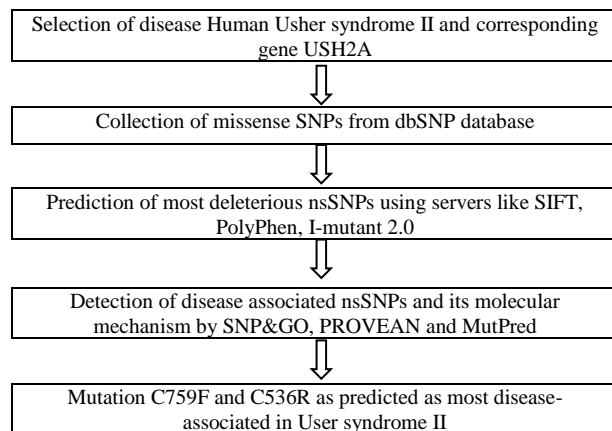


Figure 1: Flowchart of in silico screening of deleterious nsSNPs from dataset in human USH2A gene.

SIFT server estimated the consequence of nucleotide and amino acid substitution on gene and protein function. Out of the 15 screened nsSNPs, the prediction results showed 4 SNPs were deleterious. The results included SNPs C759F, W3521R and C536R with the tolerance score of 0.00 and R4674G showed score of 0.01. The remaining 11 SNPs showed negative prediction in SIFT and were removed from the dataset. In case of PolyPhen, the four screened SNPs showed prediction with probably damaging with high score. The two powerful servers SIFT and PolyPhen were proved with high accuracy in predicting the functional nsSNPs of given dataset based on gene sequence. With the above results, the accuracy of SIFT and PolyPhen was further validated through I-Mutant server. The results showed SNPs R4674G, W3521R and C536R with decrease in stability and C759F with increase in stability based on DDG score. Thus these servers predicted the role of four screened SNPs found to be most deleterious in USH2A gene and results shown in Table 1.

Table 1: nsSNPs screening from computational tools like SIFT, PolyPhen and I-mutant in USH2A gene.

SNP id	Amino Acid change	SIFT	POLYPHEN	I-mutant
rs80338902	C759F	Damaging 0.00	Probably damaging	Increase
rs80338904	R4674G	Damaging 0.01	Probably damaging	Decrease
rs111033264	W3521R	Damaging 0.00	Probably damaging	Decrease
rs111033273	C536R	Damaging 0.00	Probably damaging	Decrease

Table 2: Disease associated SNPs prediction from computational tools like SNP&GO, PROVEAN and MUTPRED in USH2A gene

SNP id	Amino Acid change	SNP&GO	PROVEAN	MUTPRED
rs80338902	C759F	disease	Deleterious -9.390	Pathological
rs80338904	R4674G	Neutral	Deleterious -4.308	Neutral
rs111033264	W3521R	Neutral	Deleterious -9.390	Pathological
rs111033273	C536R	disease	Deleterious -9.390	Pathological

SNP&GO server predicted the given query protein mutation related to disease or not. The results showed all the screened SNPs C759F and C536R predicted to be disease SNPs and W3521R, R4674G were predicted to be neutral. In case of PROVEAN server, the prediction showed four protein mutation C759F, W3521R, C536R and R4674G were deleterious with optimal score of -9 to -4. The final prediction was done by using MutPred server for all four protein mutations. The results showed three mutations C759F, W3521R and C536R were pathological and R4674G found to be neutral based on SIFT dataset as model and all results shown in Table 2. Hence, based on the comparison of effective servers, the SNPs C759F and C536R were screened and found to be most deleterious in human USH2A gene.

CONCLUSION:

In silico analysis, as now become a road map to define a standard disease specific SNP at molecular level. The current study on human USH2A gene and usherin deciphered the role of SNPs related to the Usher syndrome II. Based on the dataset of non-synonymous SNPs, the deleterious SNPs were prioritizing based on the results from powerful servers like SIFT, PolyPhen, I-mutant, SNP&GO, PROVEAN and MutPred. The analyses confirmed the disease associated mutations C759F and C536R positive for most deleterious among the dataset. Our study insights the impact of mutations in usherin related to Usher syndrome II. The future personalized medicine for the patients with mutations can be possible mode of treatment.

ACKNOWLEDGEMENTS:

The authors take this opportunity to thank the Management of VELS University for providing the facilities and encouragement to carry out this work.

CONFLICT OF INTEREST:

The author declares there is no conflict of interest.

REFERENCES:

1. Smith et al. Clinical diagnosis of the Usher syndromes. American Journal of Medical Genetics. 50; 1982:32-38.
2. Kimberling WJ et al. Localization of Usher syndrome type II to chromosome 1q. Genomics. 7; 1990: 245-249.
3. Weston MD et al. Gene mapping of Usher syndrome type IIa. localization of the gene to a 2.1-cM segment on chromosome

- 1q41. American Journal of Human Genetics. 56; 1995:216-223.
4. Pinckers AJLG, et al. Stable and progressive hearing loss in type 2A Usher syndrome. Annals Otolaryngology, Rhinology Laryngology. 105; 1996: 962-967.
5. Weston MD et al. Genomic Structure and Identification of Novel Mutations in Usherin, the Gene Responsible for Usher Syndrome Type IIa. The American Journal of Human Genetics. 66(4); 2000: 1199-1210.
6. Kimberling WJ et al. Localization of Usher syndrome type II to chromosome 1q. Genomics. 7; 1990:245-249.
7. Pinckers AJLG, et al. Stable and progressive hearing loss in type 2A Usher syndrome. Annals Otolaryngology, Rhinology Laryngology. 105(12); 1996: :962-967.
8. Hejtmancik JF et al. Clinical diagnosis of the Usher syndromes. Usher Syndrome Consortium. American Journal of Medical Genetics 50; 1994: 32-38.
9. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research. 31(13); 2003:3812-3814.
10. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols. 4(7); 2009: 1073-1081.
11. Adzhubei et al. A method and server for predicting damaging missense mutations. Nature Methods. 7(4); 2010: 248-249.
12. Capriotti E, Fariselli P and Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research. 33; 2005: W306-310.
13. Calabrese et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. Human Mutation. 30; 2009: 1237-1244.
14. Choi et al. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 7(10); 2012: e46688.
15. Li et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 25(21); 2009: 2744-2750.