International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 3, No. 3 (2014): 388-392
Research Article
Open Access

ISSN: 2320-0634

Mutation in (*GJB3* and *GJB4*) Genes involved in Deafness in two Sudanese Families Using Next Generation Sequencing

Mohamed A. Salih[1], Amina I. Dirar[2,*], Nada O. Ibrahim[1], Maison H. Osman[1,3], Solima M.A. Sabeel[1], Nuha A. Yusif[3], Alsadeg A. Yusif[4], Gusai H. Abdelsamad[3] and Atif A. Elagib[6]

[1] *Department of Bioinformatics, Africa City of Technology, Sudan.*
[2] *Faculty of Pharmacy, University of Khartoum, Sudan.*
[3] *Antalya Medical Center Khartoum, Sudan.*
[4] *Faculty of Medicine, University of Khartoum, Sudan.*
[5] *Department of Microbiology, University of Bahri, Sudan.*
[6] *Department of Immunology and Biotechnology, Tropical Medicine Research Institute, National Center of Research, Sudan.*

*Corresponding author: Amina I. Dirar; e-mail: aminadirar2007@gmail.com

Received: 28 April 2014 Accepted: 15 May 2014 Online: 28 May 2014

ABSTRACT

Next-generation sequencing (NGS) technology becomes the premier tool in genetic and genomic analysis by offering high-throughput sequencing.The identification of causer variant of genetic disease became a great challenge. Various *in silico* bioinformatics tools have been used to predict deleterious effect of mutation associated with hearing impairment among Sudanese patients. Our aim was to explore whether other genes, than those reported single nucleotide polymorphisms (SNPs), associated with deafness are found among Sudanese patients using multiple algorithms tools. NGS data of two Sudanese families were analyzed *in silico*. The potentially functional nonsynonymous single nucleotide polymorphisms (nsSNPs) and their effect on protein was predicted by Polymorphism Phenotype (PolyPhen) and Sort Intolerant from Tolerant (SIFT) softwares, respectively. Protein stability change was calculated using I-Mutant 2.0. University of California, San Francisco (UCSF) chimera software was used to compare between the 3D structure of wild and mutant type of proteins. Our analysis showed deleterious nsSNP's in *GJB3* and *GJB4* genes ranging from 0.00-0.05 (SIFT), >1.50 (Polyphen) and decrease protein activity (I-mutant), and accordingly the candidate genes were selected as a causative variant of hearing impairment in two Sudanese families. In conclusion, NGS provides data to facilitate the discovery of variants associated with hearing impairment. The accuracy of deleterious nsSNPs predicted can be increased by combining different computational methods as SIFT, Polyphen2 and I-Mutant 2.0.

Keywords: Next generation sequencing; nonsynonymous single nucleotide polymorphisms; *GJB3*; *GJB4*; deafness

INTRODUCTION

Hearing loss is the most prevalent neurosensory disorder in humans [1]. It affects individuals of all ages and at any time. In developed countries, cases of isolated deafness are caused by genetic, environmental and idiopathic causes almost equally [2]. Deafness could be syndromic or non-Syndromic. The later can be transmitted in an autosomal recessive, autosomal dominant or X-linked mode of inheritance [3]. In 2012, WHO released new estimates on the magnitude of disabling hearing loss reaching 360 million persons (5.3% of the world's population) [4]. The prevalence of disabling hearing loss is highest in South Asia, Asia Pacific and Sub-Saharan Africa, affecting child development, education, social integration, and quality of life with a substantial impact on public health [5].

Connexins are the major proteins of gap junctions and are important in the key process of intercellular communication in most metazoan cell types. Connexins family includes many genes for example *GJB2*, *GJB3*, *GJB4* and *GJB6*. Several connexins are known to be expressed in the cochlea and involved in deafness [6] Distinct dominant mutations in the same connexin molecules have been demonstrated to underlie either skin disease or deafness or, indeed, both disorders [7]. Mutations in the gap junction β2 (*GJB2*) gene encoding the connexin 26 protein, are responsible for approximately 50% of autosomal recessive non-syndromic hearing loss in many populations [8]. Unlike *GJB2* none of the *GJB3* mutations associated with skin disease are associated with genetic hearing loss. In previous study, *GJB4* gene was classified as a probable non-causative variant in Sensorineural Hearing Impairment (SNHI) [9]. To date, more than 100 genes are associated with deafness, and approximately 50 genes have been identified to cause non-syndromic hereditary hearing impairment [9]. Recently, deafness-associated genes identified so far, showed that mutations in *GJB2* are the most common cause of monogenic hearing impairment worldwide [9]. Mutations in *GJB3* were identified in families with autosomal dominant and recessive non-syndromic hearing loss (NSHL) and erythrokeratodermia variabilis (EKV) [7]. Due to the vast genetic heterogeneity of deafness, high-throughput sequencing known as Next-Generation Sequencing (NGS) is the ideal tool, to detect and map rare variants [5, 10]. Millions of fragments of DNA from a single sample are sequenced in unison which allows an entire genome to be sequenced in less than one day [11].

MATERIALS AND METHODS

Two deaf Sudanese patients were selected for this study, and based on data taken from these patients, no family history of deafness was reported among other members. In this work, blood specimen in EDTA tubes and clinical data were collected with ethical approval from Africa City of Technology, ethical committee and informed consent of these patients were obtained. DNA was extracted using (Qiagen Flexigen), in accordance with the manufacturer's instructions.

Employing NGS method, two patient samples originating from Sudan were screened using the targeted sequencing approach, then 80 genes associated with sensorineural hearing impairment (according to previously published literature) [9] were selected as the target genes. In this work one microgram of genomic DNA was fragmented to approximately 100 base pairs using an Illumina sample preparation kit from Otogenetic Company USA, followed by end-repair, adenylation and adapter ligation for library generation. All protein-coding regions and exon-intron, 3'-UTR, 5'-UTR of the target genes were then captured using hybridization with designed nucleotide probes. Final captured target DNA (read) was then sequenced using the Illumina HiSeq 2000, the amplified sequencing features in this platform are generated by bridge PCR [12, 13], after immobilization in the array, all the molecules are sequenced in parallel by means of sequencing by synthesis. There is a good advantage of using this approach (targeted re-sequencing), it can produce a significantly reduced amount of genetic data as compared with exome sequencing.

Bioinformatics processing and data analysis

In this study the decision of choosing the right set of SNPs to be screened was a critical one. One possible way to overcome this problem was to prioritize SNPs according to their structural and functional significance using different bioinformatics prediction tools, SNPs identification was performed and we compared the SNPs to find out whether it concerns a novel or an already known mutation using a SNPs-database server (dSNPs) [14, 16]. In order to check the effect of the mutation on the protein structure and function, SIFT and PolyPhen-2 softwares were used, respectively.

Sorting Intolerant From Tolerant [SIFT]

SIFT [14] was used to perform protein conservation analysis and predict the phenotypic effect of amino acid substitutions. SIFT is based on the premise that protein evolution is correlated with protein function. Variants that occur at conserved alignment positions are expected to be tolerated less than those that occur at diverse positions. The algorithms construct a multiple sequence alignment of proteins with the query sequence that belongs to the same group. The program generates alignments with a large number of homologous sequences and assigns scores to each residue. The scores range from 0.0-1.0. The lower the tolerance index (TI) of a particular amino acid substitution, the larger is its likely impact [15]. An nsSNP with a TI score of (≤ 0.05) is considered to be deleterious, and a score of (> 0.05) is considered to be tolerant.

Polymorphism Phenotype [PolyPhen-2]

PolyPhen-2 [16] is a computational tool for identification of potentially functional nsSNPs in coding region. Predictions are based on a combination of phylogenetic, structural and sequence annotation information characterizing a substitution and its position in the protein. For a given amino acid variation, PolyPhen-2 performs several steps: (a) extraction of sequence based features of the substitution site from the UniProt *database*, (b) calculation of profile scores for two amino acid variants, (c) calculation of structural parameters and contacts of a substituted residue. PolyPhen-2 scores are classified as 'benign', 'possibly damaging' or 'probably damaging'. We submitted the query in the form of chromosome co-ordinate with single mutational position. PolyPhen-2 is based on the search for three-dimensional protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases. Then, it calculates scores for each of two variants, and computes their difference. The higher the score

difference, the higher functional impact a particular amino acid substitution is likely to have. PolyPhen-2 scores of 0.00-1.50 are considered to be benign, 1.50-1.99 are "Possibly damaging" and >2.0 are "Probably damaging" [15, 17].

I-Mutant 2.0

I-Mutant 2.0 [18] is a support vector machine (SVM)-based tool. We used the sequence-based version of I-Mutant 2.0 that classifies the prediction into three classes: neutral mutation ($-0.5 \leq DDG \leq 0.5$ kcal/mol), large decrease (≤ -0.5 kcal/mol) and large increase (>0.5 kcal/mol). The output file shows the predicted free energy change (DDG) which is calculated from the unfolding Gibbs free energy change of the mutated protein minus the unfolding free energy value of the native protein (kcal/mol) [19].

3-D structure modeling

We used UCSF chimera software version 1.8 algorithms to determine the 3D structure model of *GJB3* and *GJB4* proteins. Comparison was done between the wild type and the mutant type 3D structures [20].

Table 1. Represents high score nsSNPs (deleterious) in *GJB3* and *GJB4* detected in two Sudanese patients, that were analyzed by computational methods SIFTS and PolyPhen-2.

Polyphen Score	SIFT Score	dbSNP ID	Substitution	Allele	Gene Name
4.116	**0**	**rs1805063:T**	**W32R**	**C/T**	***GJB3***
2.188	**0**	**rs3738346:C**	**E204A**	**A/C**	***GJB4***

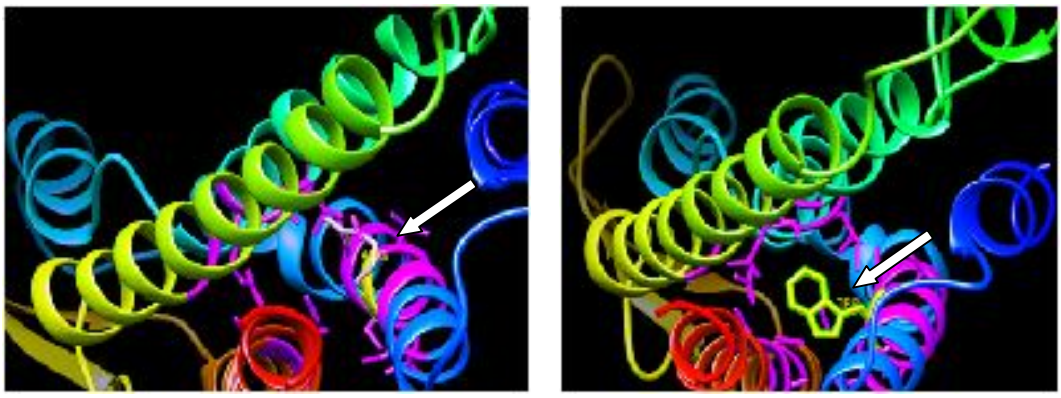
RESULTS AND DISCUSSION

Deleterious mutation Analysis

In the first patieint, 18 nsSNPs were predicted to be deleterious by both softwares, SIFT and PolyPhen-2 with a score ranging from 0.00-0.05 (SIFT) and > 1.50 (Polyphen). In the second patient, 17 nsSNPs were deleterious by both softwares. High scoring SNP's were found in *GJB3* and *GJB4* genes and were selected as a causative variant of hearing impairment among these patients, respectively (Table 1).

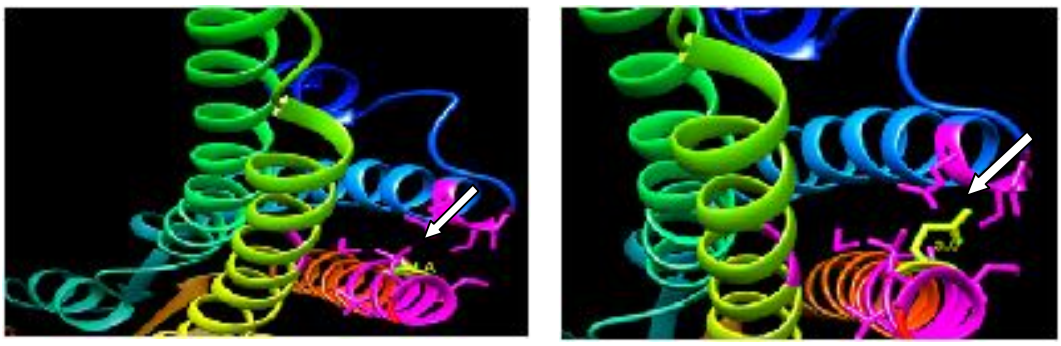
Modeling deleterious nsSNPs

Single amino acid mutations can significantly alter protein structure thereby disturbs stability. The protein sequences of *GJB3* and *GJB4* retrieved from NCBI [21] were submitted into CPH model server [22] to obtain the wild type 3D protein structure from Chimera software. Then it was compared with the mutant type 3D structure (Figures 1 and 2).



(A) Wild type **(B) Mutant type**

Figure 1. Models of wild type and mutant structures (cartoon shape) of *GJB3*. (A) Structure of wild type amino acid Treptophan (TRP) in "stick" (yellow color) at position 32 of *GJB3*. (B) Structure of mutant amino acid Arginine (ARG) in "stick" (yellow color) at position 32 of *GJB3*.



(A) Wild type **(B) Mutant Type**

Figure 2. Models of wild type and mutant structures (cartoon shape) of *GJB4*. (A) Structure of wild type amino acid Glycine (GLY) in "stick" (Yellow color) at position 204 of *GJB4*. (B) Structure of mutant amino acid Alanine (ALA) in "stick" (Yellow color) at position 204 of GJB4.

Table 2. I-Mutant 2.0 result

T	pH	RI	Stability	NEW	WT	Position	Gene name
25	7	6	Decrease	R	W	32	***GJB3***
25	7	7	Decrease	A	E	204	***GJB4***

WT: Amino acid in Wild-Type Protein; T: Temperature in Celsius degrees; NEW: New Amino acid after Mutation; pH: -log[H+]; RI: Reliability Index

Protein stability analysis

The protein stability change due to a single point mutation was predicted using support vector machine-based tool I-Mutant 2.0. The candidate nsSNPs with ID's, rs1805063 (W32R) and rs3738346 (E204A) were submitted as protein sequences input data to I-Mutant 2.0 software and showed decrease protein stability (Table 2).

Next generation sequencing (NGS), though it is an expensive technique, it provides greater understanding of comparative biological studies through whole-genome sequencing. Mutation events that occur in gene-coding regions give rise to different phenotypic variation and thus NGS provides a fast, affordable, and thorough way to determine the genetic cause of a disease. Computational methods (SIFT, PolyPhen-2, and I-Mutant 2.0) provide an accurate powerful tools in understanding the prediction of the phenotypic effect of nsSNPs for Mendelian inheritance disease [19, 23].

A design of a 3D model of *GJB3* and *GJB4* native proteins were compared with mutated modeled proteins using UCSF Chimera. I-Mutant showed decreased stability in *GJB3* and *GJB4* genes. We choose deleterious nsSNPs of *GJB3* gene namely rs1805063:T (W32R/NCBI data), and of *GJB4* gene namely rs3738346:C (E204A) based on scoring and clinical data, respectively. Unexpectedly, the position of the nsSNP mutation in *GJB3* gene was found to be R32W in Expasy database.

A previous study was done in Kenyan and Sudanese school children with non-syndromic autosomal recessive deafness (NSARD) caused by variants of the *GJB2* gene. Compared to many other ethnic groups, deafness-associated variants of the coding region of *GJB2* are rare in Sudan and Kenya, suggesting a causative role of other genetic or epigenetic factors [24]. Our result totally agree with their suggestion that *GJB3* and *GJB4* could be the candidate genes responsible for hearing impairment in Sudanese. Additionally, another study suggests that E204A may be strongly associated with deafness [25]. Based on these predictions, it can be hypothesized that E204A mutation has pathogenic effect on Connexins' activity leading to hearing loss.

W32R found in the first patient caused by a single nucleotide SNP (C/T) leading to the substitution of Arginine to Tryptophan at position 32, this mutation was predicted to be damaging by SIFT and PolyPhen-2. Through homology modeling, it was predicted that the W32R residue contributes to an increase in structural stabilization of the Connexin. Another study detected a mutation in a Leucine 34-to-Proline (L34P) substitution in the first transmembrane helix [26]. Mutation W32R is only two nucleotides away from L34P, in the first transmembrane helix. We suggest that this missense mutation have pathogenic effects and that the mutant protein could not localize gap junctions between adjacent cells. Impact of single amino acid substitution on protein stability remains one of the most promising setbacks in protein science. Therefore relationship between nsSNPs and their location in protein structure must be analyzed to determine the effect of nsSNPs mutation associated with non-syndromic hearing impairment. Further genetic studies must be conducted to link other variant in these genes to non-syndromic hearing impairment in Sudanese.

CONCLUSION

Next generation sequencing (NGS) provides both qualitative and quantitative sequence data to facilitate the discovery of genes, regulatory element and functional rare variants associated with hearing impairment in patients. The accuracy of deleterious nsSNPs predicted can be increased by combining different computational methods as SIFT, PolyPhen-2 and I-Mutant 2.0. We hope our work will provide an empirical guideline for researchers to prioritize the known nsSNPs on the basis of molecular analysis. It is obvious from the results that the deployment of *in silico* tools for application in biomedical research is highly effective and has a great impact on the ability to uncover the cause of genetic variation in different genetic diseases. Important for future studies is an increase in selection of sample size in order to obtain more information regarding possible novel genes and their mutations involved in the hearing process.

Acknowledgments

This project was supported by Africa City of Technology (ACT) and Sudanese National Society for Deaf (SNSD). We thank Prof. Taha Ahmed Talaat (SNSD), who helped in diagnosis and allowed to work in his group, as we thank all the families and patients for their cooperation in this study.

REFERENCES

1. Rudnicki, A. and K.B. Avraham (2012). microRNAs: the art of silencing in the ear. EMBO Mol Med. 4(9): 849-859.
2. Rehm, H.L., Williamson, R. E., Kenna, M. A., D.P. Corey, and B.R. Korf, eds. Understanding the Genetics of Deafness: A Guide for Patients and Families. 2003, Harvard Medical School Center for Hereditary Deafness.
3. Al-Qahtani, M.H., Baghlab, I., Chaudhary,A. G., et al. (2009). Spectrum of GJB2 Mutations in a Cohort of Nonsyndromic Hearing Loss Cases from the Kingdom of Saudi Arabia. Genet Test Mol Biomarkers. 14(1): 79-83.
4. WHO. Deafness and hearing loss. 2013; Available from: http://www.who.int/mediacentre/factsheets/fs300/en/.
5. Brownstein, Z., Y. Bhonker, and K.B. Avraham (2012). High-throughput sequencing to decipher the genetic heterogeneity of deafness. Genome Biol. 13(5): 245.
6. Alexandrino, F., Oliveira, C. A., Reis, F. C et al. (2004). Screening for mutations in the GJB3 gene in Brazilian patients with nonsyndromic deafness. J Appl Genet. 45(2): 249-54.

7. Kelsell, D.P., W.L. Di, and M.J. Houseman (2001). Connexin mutations in skin disease and hearing loss. Am J Hum Genet. 68(3): 559-68.
8. Choi, S.Y., Lee, K. Y., Kim, H. J., et al. (2011). Functional evaluation of *GJB2* variants in nonsyndromic hearing loss. Mol Med. 17(5-6): 550-6.
9. Wu, C.C., Lin, Y. H., Lu, Y. C., et al. (2013). Application of massively parallel sequencing to genetic diagnosis in multiplex families with idiopathic sensorineural hearing impairment. PLoS One. 8(2): e57369.
10. Liu, D.J. and S.M. Leal (2010). A Novel Adaptive Method for the Analysis of Next-Generation Sequencing Data to Detect Complex Trait Associations with Rare Variants Due to Gene Main Effects and Interactions. PLoS Genetics. 6(10): e1001156.
11. Grada, A. and K. Weinbrecht (2013). Next-generation sequencing: methodology and application. J Invest Dermatol. 133(8): e11.
12. Fedurco, M., Romieu, A., Williams, S. et al. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 34(3): e22.
13. Adessi, C., Matton, G., Ayala, G., et al. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res. 28(20). e87
14. Ng, P., Sorting Intolerant Form Tolerant. Retrieved from SIFT: http://sift.bii.a-star.edu.sg/.
15. Masoodi, T.A., Shammari, S. A., Al-Muammar, M. N. et al. (2012), Screening and structural evaluation of deleterious Non-Synonymous SNPs of ePHA2 gene involved in susceptibility to cataract formation. Bioinformation. 8(12): 562-567.
16. PolyPhen-2 prediction of functional effects of human nsSNPs. 2013; Available from: http://genetics.bwh.harvard.edu/pph2/index.shtml.
17. Xi, T., I.M. Jones, and H.W. Mohrenweiser (2004). Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics. 83(6): 970-979.
18. I-Mutant2.0. Retrieved December 2013, from I-Mutant2.0: http://folding.biofold.org/i-mutant/i-mutant2.0.html.
19. Doss, C.G.P., Rajith, B., Garwasis, N., et al. (2012). Screening of mutations affecting protein stability and dynamics of FGFR1—A simulation analysis. Applied & Translational Genomics 1(1); 37–43.
20. UCSF Chimera an Extensible Molecular Modeling System. 2013; Available from: http://www.cgl.ucsf.edu/chimera/.
21. The National Center for Biotechnology Information. Available from: http://www.ncbi.nlm.nih.gov/.
22. Brunak, S.; Available from: http://www.cbs.dtu.dk/services/CPHmodels/.
23. Hicks, S., Wheeler, D. A. , Plon,S. E. et al. (2011). Prediction of Missense Mutation Functionality Depends on Both the Algorithm and Sequence Alignment Employed. Human Mutatation. 32(6): 661–668.
24. Gasmelseed, N.M., Schmidt, M., Magzoub, M. M., et al. (2004). Low frequency of deafness-associated *GJB2* variants in Kenya and Sudan and novel *GJB2* variants. Hum Mutat, 2004. 23(2): 206-207.
25. Lopez-Bigas, N., Melchionda, S., Gasparini, P., et al. (2002). A common frameshift mutation and other variants in *GJB4* (connexin 30.3): Analysis of hearing impairment families. Hum Mutat. 19(4): 458.
26. Gottfried, I., Landau, M., Glaser, F., et al. (2002). A mutation in *GJB3* is associated with recessive erythrokeratodermia variabilis (EKV) and leads to defective trafficking of the connexin 31 protein. Hum Mol Genet. 11(11): 1311-6.

*****