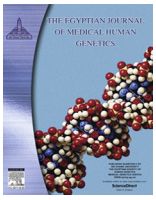
The Egyptian Journal of Medical Human Genetics 19 (2018) 333–344

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Medical Human Genetics

journal homepage: www.sciencedirect.com



Original article

# Pathogenic predictions of non-synonymous variants and their impacts: A computational assessment of *ARHGEF6* gene

Yashvant M. Khimsuriya, Jenabhai B. Chauhan *

*Post Graduate Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand, India*

ARTICLE INFO

*Article history:*
Received 21 March 2018
Accepted 8 May 2018
Available online 23 August 2018

*Keywords:*
Computational methods
*ARHGEF6*
Intellectual disability
Missense mutation

ABSTRACT

*Introduction:* ARHGEF6, a key member and activator of RhoGTPases family that is involved in G-Protein Coupled receptor (GPCR) pathway and stimulate Rho dependent signals in the brain, and mutations in this gene can cause intellectual disability (ID) in Human. Therefore, we aimed to study the consequences of *ARHGEF6* non-synonymous mutations by using advanced computational methods.
*Methods:* Classification of the genetic mutations in *ARHGEF6* gene was performed according to Ensembl Genome Database and data mining was done using ensemble tools. The functional and disease effect of missense mutations, and pathogenic characteristics of amino acid substitutions of *ARHGEF6* were analyzed using eleven diversified computational tools and servers.
*Results:* Overall, 47 *ARHGEF6* non-synonymous (NS) variants were predicted to be deleterious by SIFT, Polyphen2 and PROVEAN scores. Above that, SNPs&GO and PhD SNP were further graded 21 customarily pathogenic NS-variants. Protein stability analysis resulted in the significant change in terms of $\Delta\Delta G$ of most identified NS-variants, except K609I. Seven variants were analyzed to be located on most potential domain RhoGEF/DH, whereas the remaining 14 were distributed on CH, SH3, PH and BP domains. Furthermore, pathogenic effects of mutations on protein was presented with different parameters using MutPred2 and PROJECT HOPE. Additionally, STRING network data predicted GIT2 and PARVB as most interacted partners of ARHGEF6.
*Conclusion:* These findings can be supportive of genotype-phenotype research as well as the development in pharmacogenetics studies. Finally, this study revealed a significance of computational methods to figure out highly pathogenic genomic variants linked with the structural and functional relationship of ARHGEF6 protein.
© 2018 Ain Shams University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

The Rho-guanine nucleotide exchange factor 6 (ARHGEF6) protein is known for its involvement in the Rho GTPase cycle, which mediates the organization of cytoskeleton, cell shape, and motility. It is identified as third responsible X-linked intellectual disability (XLID) gene, after Oligophrenin 1 (*OPHN1*) and P21-protein activated kinases 3 (*PAK3*) [1]. It is also known as PAK-interacting exchange factor, alpha (*αPIX*) and *COOL2*. ARHGEF6 is 87.5 kDa protein of 776 amino acids, which belongs to a family of cytoplasmic proteins (RhoGTPases) that activate the Ras-like family of Rho proteins by exchanging bound GDP for GTP. ARHGEF6 in complex with BIN2 and GIT2, forms a complex with G-proteins and stimulate Rho-dependent signals [2]. It also acts as a Ras-related C3 botulinum toxin substrate 1 (RAC1) guanine nucleotide exchange factor. The RhoGTPases are critical regulators of the actin cytoskeleton, where they often mediate signaling from the external environment. In the central nervous system, their function has been linked to axonal growth, development of dendritic arborizations and spine morphogenesis [3,4].

As an activator protein, ARHGEF6 plays a significant role in the cellular mechanisms of Rho-GTPases. The biological mechanisms through which *ARHGEF6* mutations causes the intellectual disability are still not well recognized, although defective plasticity of synaptic networks have been previously proposed. However, several studies reported that this protein is primarily expressed in neuropil regions of the hippocampus and the deregulations can alter neuronal connectivity and impaired synaptic function and cognition [5]. ARHGEF6 located in dendritic spines regulate spine

Peer review under responsibility of Ain Shams University.
* Corresponding author at: Associate Professor & Head, Post Graduate Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand, Gujarat 388121, India. Fax & Tel.: +91 2692 229 189.
*E-mail address:* jenabhaichauhan@aribas.edu.in (J.B. Chauhan).

https://doi.org/10.1016/j.ejmhg.2018.05.002
1110-8630/© 2018 Ain Shams University. Production and hosting by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

morphogenesis by acting through downstream activation of p21-activated kinase (PAK3). The mutations in both of these genes could induce intellectual disability [6]. A recent study reported that αPIX promotes dendritic Golgi translocation in hippocampal neurons [7].

The human *ARHGEF6* gene, spanning over 22 exons is located on X-chromosome at sub-band q26.3. In humans, rare and common genomic mutations of *ARHGEF6* are constantly being diagnosed with the help of conventional as well as high throughput sequencing technologies [8]. But, the evaluation and correlation of these broad spectrum of clinical phenotypes and their connection with molecular alterations of *ARHGEF6* gene are not yet well examined.

The disease causing mutations may usually affect the size, charge and hydrophobicity value of each encoded amino acid variant, which can successively change the hydrogen bonding and conformational dynamics of the protein. Accordingly, the ability to better interpret the clinical complications of every mutation depends on identifying the real constructive pathogenic mutations from the correlated markers. Despite the fact that molecular validation of such mutations by *in vitro* and *in vivo* studies is more time consuming, and often requires technical expertise and huge expenses. The alternate approach to defeat this challenge is to examine the impact of each genetic variation using a recently developed advanced computational algorithms approaches. Different types of bioinformatics programs and servers have been designed to discover the consequences of genetic mutations on biophysical characteristics, structure and functional properties of proteins [9].

Therefore, we aimed this study to analyze pathogenic variants of *ARHGEF6* gene in exonic positions and to predict the structural and functional implications of ARHGEF6 protein by subjecting the gene sequences along with non-synonymous mutations to the various computational methods.

## 2. Methods and datasets

The Single Nucleotide Variants (SNVs) and protein sequence of the *ARHGEF6* gene (transcript ID:ENST00000250617.6) were obtained from NCBI dbSNP available at http://www.ncbi.nlm.nih.gov/SNP/ [10], NCBI protein (https://www.ncbi.nlm.nih.gov/protein/) [11] and Ensembl genome browser (https://asia.ensembl.org/index.html) [12]. The classification of all collected SNVs was done as non-coding and coding, depending on the variant nature and position. Only missense variants (non-synonymous) were chosen for further computational analysis because of their potential to disturb the structural conformation of proteins.

### 2.1. Functional prediction of missense variants

Damaging and deleterious effect of missense variants were predicted using the scores Sorting Intolerant from Tolerant (SIFT) (http://sift.jcvi.org), Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/) and Protein Variation Effect Analyzer (PROVEAN) (http://provean.jcvi.org/index.php) tools. SIFT is a sequence homology based tool which predicts the tolerated and deleterious SNVs and identifies the impact of amino acid substitution on protein functions. The results can be deleterious or tolerated substitutions demonstrating threshold $\leq 0.05$ score [13]. Polyphen2 is a sequence and structure evolutionary conservation based tool to classify damaging effect of amino acid substitutions and estimates position specific independent count (PSIC) score demonstrating 0.801–1.00 probably damaging index [14]. PROVEAN is a software to obtain pairwise sequence alignment (PSA) score and to identify non-synonymous variants [15]. Furthermore, we used Single Nucleotide Polymorphisms & Gene Ontology (SNPs&GO) (http://snps.biofold.org/snps-and-go/snps-and-go.html) and Predictor of human deleterious single nucleotide polymorphisms (PhD-SNP) (http://snps.biofold.org/phd13 9snp/phd-snp.html) those are Support Vector Machine (SVM)-based tools that used evolutionary information, protein sequence and functions to predict if a given mutation can be classified as disease-related or neutral [16,17].

### 2.2. Structural conformation and conservation analysis

The Consurf server available at http://consurf.tau.ac.il/, was used for high-throughput characterization and evolutionary conservation of amino acid positions based on the phylogenetic relationship between homologous sequences [18]. The degree of conservation of the amino-acid sites among 50 homologs with similar sequences was estimated. The conservation grades were then projected onto the molecular surface of the human *ARHGEF6* to reveal the stripes with highly conserved residues that are usually essential for biological function.

### 2.3. Prediction of disease related amino acid substitution by MutPred2

The MutPred2 (http://mutpred.mutdb.org/), a unique web based tool was developed to predict any amino acid substitution, whether pathogenic or benign [19]. Based on>50 different protein properties we can classify the inference of molecular mechanisms of pathogenicity. It uses SIFT, PSI-BLAST [20], and Pfam profiles [21] along with some structural disorder prediction algorithms, TMHMM [22], MARCOIL [23], and DisProt [24]. Random Forest (RF) classifier was used and obtained g-score for prediction of the probability and the p score for identification of structural and functional properties. As a result, by combining the scores of all programs, the accuracy of prediction ascend to a greater extent.

### 2.4. Structural analysis of ARHGEF6 protein and mutants

#### 2.4.1. Protein structure prediction and modeling

To succeed in dealing with the absence of crystal protein structure in databases, ARHGEF6 protein structure was built after subjecting the referenced amino acids sequence (NP_004831) to the I-TASSER (Iterative Threading ASSEmbly Refinement), a web based server available at https://zhanglab.ccmb.med.umich.edu/I-TASSER/. It uses basic templates from the PDB by multiple threading approaches and constructs full-length atomic models by iterative template fragment assembly simulations [25]. Then, it has predicted five top models, among which one best model was identified on the basis of confidence score, estimated TM-score and estimated root mean square deviation (RMSD) value. Further, the similar standards was also analyzed for structural deviation prediction for the c-alpha atoms of each amino acid residue in the mutant models. The selected model was eventually subjected for Gromacs energy minimization by Normal Mode Analysis Deformation and Refinement (NOMAD-Ref) Server available at http://lorentz.immstr.pasteur.fr/nomad-Ref.php, to remove disarrangement in the space of a collection of atoms [26]. This energy minimized model was used as a standard template to construct mutant models of ARHGEF6 (manually inserted mutated residues in the referenced protein sequence of *ARHGEF6*) by Modeller v9.19. This software applies comparative protein structure modeling by satisfaction of spatial restraints in the protein of interest. Likewise, RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) was used to check stereo-chemical properties of ARHGEF6 wild type models [27]. PyMol and Chimera programs were used to generate mutated models and visualize interactions of molecules [28].

Download English Version:

https://daneshyari.com/en/article/11019489

Download Persian Version:

https://daneshyari.com/article/11019489

Daneshyari.com