

Prioritization of Disease Prone Exons in *INPP5E* Gene, Associated With Joubert Syndrome, by *in silico* Analysis of Non-Synonymous SNPs

Afia Gul,¹ Muhammad Ayaz Khan,¹ Muhammad Zubair,¹ Iftikhar Ahmad² and Muzammil Ahmad Khan^{1,3*}

¹Gomal Centre of Biochemistry and Biotechnology, Gomal University, D.I. Khan, Khyber-Pakhtoonkhuwa, Pakistan

²Department of Community Medicine, Gomal Medical College, D.I. Khan, Khyber-Pakhtoonkhuwa, Pakistan

³Interim Translational Research Institute, Hamad Medical Corporation, Doha, Qatar

Abstract :- In the present computational study, various softwares have been employed for functional and structural analysis of non-synonymous single nucleotide polymorphism (nsSNP) in the protein coding exons of *INPP5E* (MIM# 613037) gene to determine its deleteriousness. Mutation in this gene causes Joubert syndrome (MIM # 213300). The overall bioinformatics analysis predicted eight most deleterious nsSNPs in the four candidate exons *i.e.* 2, 3, 4, and 7. Out of these eight damaging nsSNPs three each are present in 2nd and 4th exons, while 3rd and 7th exons contain one damaging nsSNP each. This study will assist the molecular geneticists to selectively sequence the candidate disease associated exons that will contain the deleterious nsSNPs, inspite of screening the whole gene.

Key words: Joubert Syndrome, *INPP5E*, nsSNP, insilico analysis, disease prone exons, molecular geneticist.

INTRODUCTION

One of the primary challenges of current genomics research is to explore the genomic variations in the human population. Genomic variations underlying the human genetic disorders are of key interest in current molecular medicine and pharmacogenomics. There are numerous forms of genetic variation in the human genome, ranging from large chromosomal anomalies to single nucleotide variation. The most simplest and common form of these variations is the substitution of single nucleotide, termed as single nucleotide polymorphism (SNP) (Wu and Rui, 2012). SNP may prevail in coding region of gene (protein coding part), non-coding region (intron, promoter) or in the intergenic region (regions between genes). The effect of SNP is exclusively dependent upon its genetic position. SNPs present in the promoter, regulatory region and intron influence on the gene expression, while SNP in coding region affects function of encoding protein. SNP variant which produces same polypeptide chain is named as

synonymous SNP and that which encodes different protein is termed as non-synonymous SNP (ns-SNP). It is estimated that half of all human genetic diseases are the consequence of nsSNP variants. Therefore, these nsSNPs with drastic phenotypic consequences are usually considered as deleterious nsSNPs, in contrast to tolerant nsSNPs with no phenotypic changes (Teng *et al.*, 2009).

The *INPP5E* gene, which encodes inositol polyphosphate-5-phosphatase, is a 644 amino acids long protein and is involved in removing phosphate from inositol phosphates or inositol phospholipids from its 5' position (Kisseleva *et al.*, 2000). Cytogenetically *INPP5E* is positioned on 9q34.3 locus. Sub-cellularly; *INPP5E* protein is localized in cytoplasm, cytoskeleton, cilium axoneme and Golgi apparatus. Mutational study has revealed that this gene leads to Joubert syndrome; a clinically and genetically heterogeneous group of disorders characterized by midbrain-hindbrain malformation and various associated ciliopathies that include retinal dystrophy, nephronophthisis, liver fibrosis and polydactyly (Saraiva and Baraitser, 2009). Defect in *INPP5E* also causes a syndromic disorder associated with mental retardation, truncal obesity, retinal dystrophy and micropenis syndrome (MORMS) (Jacoby *et al.*, 2009).

* Correspondence author: muzammilgandapur1983@gmail.com
0030-9923/2015/0003-0601 \$ 8.00/0
Copyright 2015 Zoological Society of Pakistan

Prime objective of the present study is to determine the biological effect of reported alleles in *INPP5E* genes. The study also aims to determine their rate of prevalence in multiple exons of *INPP5E* to get an idea about the most probable disease associated exons. To accomplish this work, nsSNP data was taken from Ensembl Genome Browser and analyzed for biological effect, amino acid conservation, protein stability and protein secondary structure prediction by using various online bioinformatics tools. The computational analysis identified 8 damaging alleles distributed among exons 2, 3, 4 and 7.

MATERIALS AND METHODS

For prioritization of candidate exons in *INPP5E* gene, SNP alleles data were obtained from Ensembl Genome Browser (April 2013 release). Prediction of biological effect after single amino acid substitution was performed by using SIFT (Sorting Intolerant From Tolerant) (Kumar *et al.*, 2009), Polymorphism Phenotyping 2 (PolyPhen 2) (Adzhubei *et al.*, 2013), SNAP (screening for non-acceptable polymorphisms) (Bromberg and Rost, 2007) and Mutation taster (Schwarz *et al.*, 2010) softwares (See Wang *et al.*, 2013 for understanding the principle of above softwares). Commonly predicted deleterious nsSNPs were submitted for evaluating effect of protein stability by using I-Mutant 2.0 software, which is a support vector machine (SVM)-based program used for the automated prediction of protein stability changes upon missense mutation (Capriotti *et al.*, 2005). Afterwards; evaluation of Amino Acid conservation during the course of evolution was determined by aligning the protein sequences of eight different species *viz.*, *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Danio rerio*, Chimpanzee, Fugo, Anole lizard, and Orangutan. This biological sequence alignment was performed by using bioedit software (Hall, 1999). Protein structural variation, due to nsSNP, was assessed through PSIPRED (protein structure prediction), which determine protein secondary structure prediction (Jones, 1999). Figure 1 summarizes methodology in the form of flow sheet diagram. The authenticity of employed bioinformatics tools are already validated through

the mutational study in human genetic disorders (Shaiq *et al.*, 2012).

RESULTS AND DISCUSSION

INPP5E gene harbour 633 SNPs in total, out of which 70 are nsSNPs, 53 are synonymous SNPs (sSNPs), 21 are splice site variants and 221 are intronic variants (Ensembl Genome Browser, April 2013 release). The data already contained analyzed result of SIFT and Polyphen 2 softwares for the biological significance of nsSNPs. Mutation Taster, SNAP and I-Mutant 2, mutually identified 26 nsSNPs as the most damaging ones. Then, the evolutionary conservation of substituted amino acid, due to nsSNP, found 24 nsSNPs (out of 26) occurring in the highly conserved region. Afterwards; effect on the protein secondary structure, arisen due to single amino acid substitution, determined 8 nsSNPs (out of 24) which reduces the stability of original secondary structure formed by wild type protein (Table I). These 8 nsSNPs are distributed among the four different protein encoding exons (2, 3, 4, and 7) of *INPP5E* gene, and were regarded as the plausible candidate disease associated exons because of harboring damaging alleles of nsSNP. Out of these four disease associated exons, exon 2 contained three damaging nsSNPs (rs199873582, rs138068434, COSM225640), exon 3 harbor one (rs368235861), exon 4 harbor three (rs147531141, rs121918130, COSM135961) and exon 7 surrounded only one damaging nsSNP (rs374083402).

The nsSNPs may exert their effect on protein folding, stability, functional sites, protein expression, subcellular localization and protein interaction with other entities. The most precise and reliable method in distinguishing functional SNPs (damaging) from neutral (non damaging) ones are experimental techniques. It is not feasible in terms of time and cost to perform laboratory experiments for all nsSNPs in the targeted gene to elucidate their functional importance. In this situation computationally based prediction analysis is the only method of choice which is cost effective in terms of time.

INPP5E gene has 70 nsSNP (April 2013 release), distributed among 10 exons. Bioinformatics

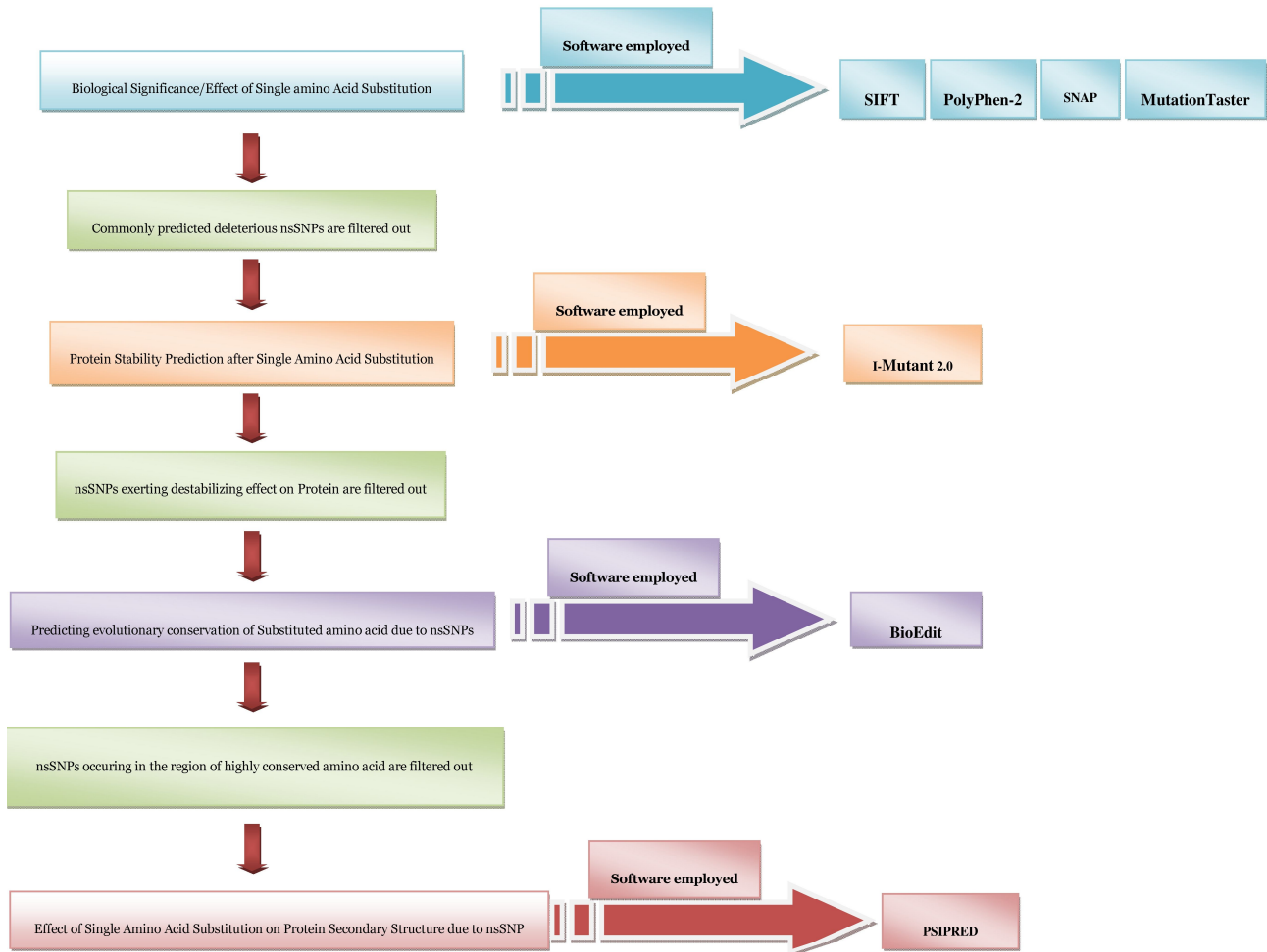


Fig. 1. Summary of the *in silico* method in the form of flow-sheet diagram to determine the damaging nsSNPs.

Table I.- Summary of deleterious non-synonymous SNPs, identified in *INPP5E* gene through computational analysis.

SNP ID	Allele	Exonic position of SNP	Amino acid polymorphism	Amino acid coordinates	Evolutionary importance of substituted amino acid	Wild type sequence		Mutated sequence	
						Protein secondary structure	Conf. value	Protein secondary structure	Conf. value
COSM225640	CC/TT	2	R/Q	297	Highly conserved	Coiled	6	Coiled	7
rs199873582	C/T	2	R/H	292	Highly conserved	Coiled	8	Coiled	9
rs138068434	C/T	2	G/R	282	Highly conserved	Coiled	1	Helix	6
rs368235861	C/T	3	G/R	337	Highly conserved	Strand	0	Strand	3
rs147531141	C/T	4	V/M	370	Highly conserved	Helix	2	Helix	3
COSM135961	C/A	4	G/C	369	Highly conserved	Helix	2	Helix	0
rs121918130	G/A	4	R/C	378	Highly conserved	Coiled	0	Coiled	2
rs374083402	A/G	7	L/P	493	Highly conserved	Helix	9	Helix	3

analysis predicted 8 deleterious nsSNPs which can affect the protein function and ultimately lead to disease phenotype. These prioritized nsSNPs are harbored by exon 2, 3, 4 and 7. Thus medical geneticist, working on *INPP5E* gene, will selectively sequence these prioritized exons for the increased risk of mutant allele prevalence.

Mutation in the *INPP5E* gene causes Joubert syndrome and MORMS syndrome. HGMD survey has found some nsSNP involved in the Joubert Syndrome, e.g. rs121918130 is reported by Bielas *et al.* (2009) in Joubert Syndrome (HGMD ID: CM095409). This indicates that the prioritized nsSNPs may have some role in disease etiology.

CONCLUSIONS

The current bioinformatics analysis of nsSNPs in *INPP5E* gene has determined 8 most probably damaging nsSNPs. Therefore it is speculated that the present *in silico* study will assist the molecular geneticists to selectively sequence the most probable candidate exons that contain the deleterious nsSNPs inspite of screening whole gene. Based on the current analytical work, four exons viz., 2, 3, 4 and 7 are prioritized as candidate disease associated exons and may also be regarded as mutational prone exons.

Conflict of interest

None declared.

ACKNOWLEDGEMENTS

Acknowledgement goes to the Provincial Government of Khyber-Pakhtunkhwa for awarding Laptop to Afia Gul in Naway Sahar Laptop scheme. MAK is sponsored by HEC, Pakistan through IPFP program and funded a start up research grant.

REFERENCES

- ADZHUBEL, I., JORDAN, D.M. AND SUNYAEV, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.*, Chapter 7:Unit7.20.
- BIELAS, S.L., SILHAVY, J.L., BRANCATI, F., KISSELEVA, M.V., AL-GAZALI, L., SZTRIHAI, L., BAYOUMI, R.A., ZAKI, M.S., ABDEL-ALEEM, A., ROSTI, R.O., KAYSERILI, H., SWISTUN, D., SCOTT, L.C., BERTINI, E., BOLTSHAUSER, E., FAZZI, E., TRAVAGLINI, L., FIELD, S.J., GAYRAL, S., JACOBY, M., SCHURMANS, S., DALLAPICCOLA, B., MAJERUS, P.W., VALENTE, E.M. AND GLEESON, J.G., 2009. Mutations in INPP5E, encoding inositol polyphosphate-5-phosphatase E, link phosphatidyl inositol signaling to the ciliopathies. *Nat. Genet.*, **41**:1032-1036.
- BROMBERG, Y. AND ROST, B., 2007. SNAP: predict effect of nonsynonymous polymorphisms on function. *Nucl. Acids Res.*, **35**: 3823-3835.
- CAPRIOTTI, E., FARISELLI, P. AND CASADIO, R., 2005. I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucl. Acids Res.*, **33**: W306-W310 doi:10.1093/nar/gki375
- HALL, T.A., 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acid Symp. Ser.*, **41**: 95- 98.
- JACOBY, M., COX, J.J., GAYRAL, S., HAMPSHIRE, D.J., AYUB, M., BLOCKMANS, M., PERNOT, E., KISSELEVA, M.V., COMPÈRE, P., SCHIFFMANN, S.N., GERGELY, F., RILEY, J.H., PÉREZ-MORGA, D., WOODS, C.G. AND SCHURMANS, S., 2009. INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. *Nature Genet.*, **41**: 1027-1031.
- JONES, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. mol. Biol.*, **292**: 195-202.
- KISSELEVA, M.V., WILSON, M.P. AND MAJERUS, P.W., 2000. The isolation and characterization of a cDNA encoding phospholipid-specific inositol polyphosphate 5-phosphatase. *J. Biol. Chem.*, **275**: 20110-20116.
- KUMAR, P., HENIKOFF, S. AND NG, P.C., 2009. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, **4**: 1073-1081.
- SARAIVA, J.M. AND BARAITSER, M., 1992. Joubert syndrome: a review. *Am. J. med. Genet.*, **43**: 726-731.
- SCHWARZ, J.M., RÖDELSPERGER, C., SCHUELKE, M. AND SEELOW, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, **7**: 575-576.
- SHAIQ, P.A., KLAUSEGGER, A., LATIF, A., BAUER, J., QAMAR, R. AND RAJA, G.K., 2012. Missense mutation in LAMA3 associated with herlitz junctional epidermolysis bullosa in a Pakistani family. *Pakistan J. Zool.*, **44**: 1697-1702.
- TENG, S., MADEJ, T., PANCHENKO, A. AND ALEXOV, E., 2009. Modeling effects of human single nucleotide polymorphisms on protein interactions. *Biophys. J.*, **96**: 2178-2188.
- WANG, M., SUN, Z., AKUTSU, T. AND SONG, J., 2013. Recent advances in predicting functional impact of

single amino acid polymorphisms: A review of useful features, computational methods and available tools. *Curr. Bioinform.*, **8**: 161-176.

WU, J. AND RUI, J., 2012 Prediction of deleterious nonsynonymous single- nucleotide polymorphism for human disease. *Scient. World J.*, doi: 10.1155/2013/675851.

Electronic References:

Ensembl Genome Browser

URL: <http://asia.ensembl.org/index.html>

Sorting Intolerant From Tolerant (SIFT)

URL: <http://sift.jcvi.org/>

Polymorphism Phenotyping 2 (PolyPhen 2)

URL: <http://genetics.bwh.harvard.edu/pph2/>

SNAP

URL: <http://www.broad.mit.edu/mpg/snap/>

Mutation taster

URL: <http://www.mutationtaster.org/>

I-Mutant 2.0

URL: <http://folding.biofold.org/i-mutant/i-mutant2.0.html>

PSIPRED

URL: <http://bioinf.cs.ucl.ac.uk/psipred/>

(Received 14 September 2014, revised 31 January 2015)