COMPUTATIONAL PREDICTION OF DAMAGING NSSNPS IN PRSS12 GENE, RESPONSIBLE FOR AUTOSOMAL RECESSIVE NON-SYNDROMIC MENTAL RETARDATION.

[1]BUSHRA ATA HASHMI, [2]MUHAMMAD AYAZ KHAN, [3]MUHAMMAD AKHLAQ, [4]MUZAMMIL AHMAD KHAN

[1,2] Gomal Centre of Biochemistry and Biotechnology, Gomal University, D.I.Khan, Khyber-Pakhtoonkhuwa, Pakistan
[3] Department of Pharmaceutics, Faculty of Pharmacy, Gomal University, Dera Ismail Khan, 29050, KPK, Pakistan
[4] Interim Translational Research Institute, Hamad Medical Corporation, Doha, Qatar

CORRESPONDENCE: MUZAMMIL AHMAD KHAN E-mail: muzammilgandapur1983@gmail.com

ABSTRACT

Various forms of genetic variations exist in the human genome which ranges from large chromosomal anomalies to single nucleotide variation. Single nucleotide polymorphism (SNP) may occur within the coding as well as non-coding regions of genes. SNPs that are present in the coding region do not always change the amino acid sequence of protein, and that due to degeneracy of genetic code. Therefore SNPs in coding region are of two types, synonymous and non-synonymous. Synonymous SNPs encode same amino acid and thus do not affect protein sequence while non-synonymous SNPs change the amino acid sequence with in protein.

Single nucleotide polymorphism is a variation of one base substitution which when appears in coding exons (nsSNPs) leads to single amino acid variation. These variations may lead to the functional consequences of protein and ultimately results in a disease phenotype. In the present computational study, various softwares were employed for functional and structural analysis of nsSNPs in the protein coding exons of PRSS12 (MIM# 606709) gene to determine its deleteriousness. Mutation in this gene causes Autosomal Recessive Non-Syndromic Mental Retardation. The gross bioinformatics analysis predicted seven most deleterious nsSNPs in five candidate exons i.e. 1, 6, 7, 9 and 11. Exon 9 and 11 contained two pathogenic SNPs while exon 1, 6, and 7 was surrounding one damaging nsSNP each. This study will assist the molecular geneticists to selectively sequence the candidate disease associated exons inspite of screening whole gene.

INTRODUCTION

With the rapid advancement in next generation technologies, such as SNP based microarray and next-generation sequencing, massive amount of SNP data is generated everyday which update the genome databases on daily basis. Single Nucleotide Polymorphism (SNP) is the most prevalent genetic variation in the human genome which may occur in coding and non-coding regions of a gene or in the intergenic regions (regions between genes). Coding SNPs do not always change the amino acid sequence of protein due to degeneracy of genetic code, therefore SNPs in the coding region are of two types i.e, synonymous and non- synonymous. Synonymous SNPs do not affect protein sequence while non-synonymous SNPs change the amino acid sequence with in a protein[1]. The SNPs that are not present in the protein coding regions may also have the possibility to affect the gene splicing, transcription factor binding or RNA degradation and gene expression. Its quite hectic job to determine the pathogenicity of numerous nsSNP in a single gene, with in a limited time and limited resources. The process is made easier by various online bioinformatics tools to predict the phenotypic effect of single nucleotide variation. These softwares are based on sequence homology, thermodynamic properties and protein structure[2].

PRSS12, protease serine 12 (neurotrypsin or motopsin), is a 'erine protease and belongs to the family of trypsin protease. Cytogenetically; PRSS12 is mapped on 4q28.1 and consist of 13 exons which encodes 875 amino acid long protein. Expression studies has reported this gene in the nerve tissues therefore its defect has been shown to cause autosomal recessive nonsyndromic mental retardation type 1 (MRT1)[3].

In the present study; computational analysis of nsSNPs in PRSS12 gene were performed to determine their phenotypic effects. The analysis determined 7 deleterious alleles in various exons. Most of these alleles were targeting SRCR domain of PRSS12 protein. The study will aid molecular biologist in targeted sequencing of PRSS12 gene in the linkage study of MR patient.

MATERIALS AND METHODS

For prioritization of candidate exons in PRSS12 gene, the SNP data was acquired from Ensembl Genome Browser (April 2013 release)[4]. The biological effect for these nsSNPs was simulated by using SIFT (Sorting Intolerant From Tolerant)[5,6], Polymorphism Phenotyping 2 (PolyPhen 2)[7,8], SNAP (screening for non-acceptable polymorphisms)[9,10] and Mutation taster[11,12] softwares [Consult Wang et al., 20132 for detailed description of these soft wares]. The data were fine tuned and commonly predicted pathogenic nsSNPs by all softwares were proposed for evaluating the effect of protein stability by using I-Mutant 2.0[13], which is a support vector machine (SVM)-based program employed for automated prediction of protein stability changes upon missense mutation[14]. Afterwards; evaluation of amino acid

conservation, for evolutionary importance of natural selection, was determined by doing multiple protein sequence alignment for eight different species viz. Homo sapiens, Musmusculus, Gallus gallus, Daniorerio, Chimpanzee, Fugo, Anole lizard, and Orangutan. The biological sequence alignment was performed by using Bioedit software[15]. Structural variation in the PRSS12 protein, due to nsSNP, was evaluated by using PSIPRED (protein structure prediction)[16], a protein secondary structure prediction tool[17].

RESULTS AND DISCUSSION

Although experimental techniques are the most comprehensive and precise methods in molecular biology to distinguish the functional SNPs from neutral ones. But it is not feasible in terms of time and cost to perform the laboratory experiments for all SNPs in human genome (or even in a single gene) and elucidate their functional consequences.

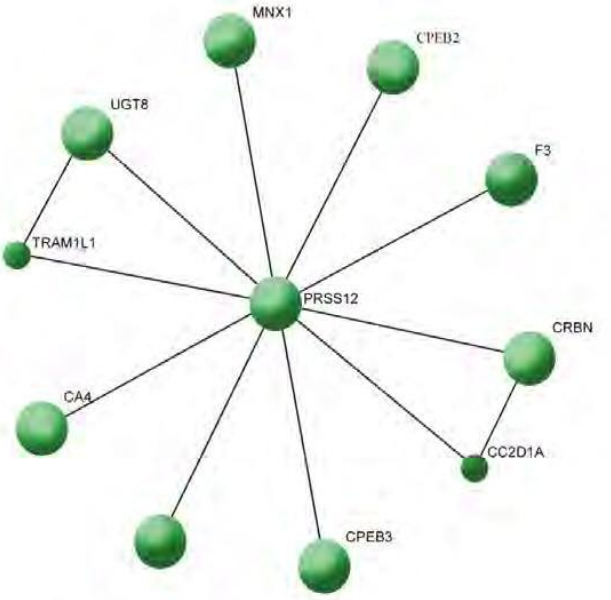
Table I:
The predicted pathogenic variants in PRSS12 gene distributed in different exons

ID	Alleles	Amino acid	Amino acid coordination	Exonic Location	Location in Domain	BioEdit Alignment	Exypred3D
rs142551296	G/C	F/L	678	11	Trypsin Domain	Highly conserved	Structurally changed
rs139381330	C/T	A/T	549	9	SRCR Domain	Highly conserved	Structurally changed
rs201005601	G/T	A/D	547	9	SRCR Domain	Highly conserved	Structurally changed
rs200738521	T/C	E/G	459	7	SRCR Domain	Highly conserved	Structurally changed
rs376308511	A/G	C/R	455	7	SRCR Domain	Highly conserved	Structurally changed
rs138640682	C/T	R/Q	426	6	SRCR Domain	Highly conserved	Structurally changed
rs72903215	G/A	P/L	147	1	Kringle Domain	Highly conserved	Structurally changed

Computational prediction has made it convenient for scientists and researches to prioritized the most probably damaging nsSNP2. In the present bioinformatics study, various softwares were used to predict the functional consequences of nsSNPs in the coding region of PRSS12 gene. This gene had 109 nsSNP in total (April 2013 release) which were distributed in 13 exons. After bioinformatics analysis, 7 nsSNPs were predicted to be the most deleterious that can affect the protein structure & function and may lead to disease phenotype. These prioritized nsSNPs are distributed among exons 1, 6, 7, 9 and 11 (see table I). Most of these pathogenic variants were detected in the SRCR domain but one each was also observed in Trypsin and Kringle domains. The clear evidence of disease association was not observed by locating these SNPs in Human Gene Mutation Database.

So far only one deletion mutation is reported in PRSS12 gene which lead to autosomal recessive nonsyndromic intellectual disability3. Expression studies also revealed its presence in Leydig cells of testis which explains that it could possibly be involved in fertility. String matching predicted F3 (coagulation factor III), CC2D1A and CRBN as strong protein interactor of PRSS12 (see figure I). This interaction study indicates the possible role of PRSS12 in signal transduction pathway.

Figure I:
The graphical presentation of PRSS12 protein interactors



CONCLUSION:

The Current bioinformatics analysis of nsSNPs in PRSS12 gene has determined thirteen most probably damaging nsSNPs. Therefore it is speculated that the present insilico study will assist the molecular geneticists to selectively sequence the most probable candidate exons that contain the most deleterious nsSNPs, inspite of screening whole gene. Based on the current analytical work, 5 exons viz, 1, 6, 7, 9 and 11 are prioritized as candidate and may be regarded as mutational prone exon.

REFERENCES

1. Wu J, Rui J. Prediction of Deleterious Nonsynonymous Single-Nucleotide polymporphism for human disease. Scientific World Journal. 2012; doi: 10.1155/2013/675851.
2. Wang M, Sun Z, Akutsu T, Song J. Recent Advances in Predicting Functional Impact of Single Amino Acid Polymorphisms: A Review of Useful Features, Computational Methods and Available Tools. Current. Bio informatics. 2013; 8: 161-176.
3. Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Augé J, Bacq D, Briault S, Vekemans M, Munnich A, Attié-Bitach T, Sonderegger P, Colleaux L. Truncating neurotrypsin mutation in autosomal recessive nonsyndromic. Mental Retardation. Science, 2002; 29; 298 (5599): 1779-1781.

4. Ensembl Genome Browser retrieved from URL: http://asia.ensembl.org/index.html
5. Sorting Intolerant From Tolerant (SIFT) retrieved from URL: http://sift.jcvi.org/
6. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. Nature. Protocols. 2009; 4: 1073–1081.
7. Polymorphism Phenotyping 2 (PolyPhen 2) retrieved from URL: http://genetics.bwh.harvard.edu/pph2/
8. Adzhubei I., Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using Poly Phen-2. Curr. Protoc. Hum. Genet. 2013; Chapter 7:Unit7.20.
9. SNAP retrieved from URL: http://www.broad.mit.edu/mpg/snap/
10. Bromberg Y, Rost B. SNAP: predict effect of nonsynonymous polymorphisms on function. Nucleic. Acids. Research, 2007; 35: 3823–3835.
11. Mutation taster retrieved from URL: http://www.mutationtaster.org/
12. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nature. Methods. 2010; 7: 575–576.
13. I-Mutant 2.0 retrieved from URL: http://folding.biofold.org/i-mutant/i-mutant2.0.html
14. Capriotti E, Fariselli, P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucl. Acids. Res. 2005; 33: W306-W310 doi:10.1093/nar/gki375.
15. Hall TA. BioEdit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl acid symposium series No., 1999; 41: 95- 98.
16. PSIPRED retrieved from URL: http://bioinf.cs.ucl.ac.uk/psipred/
17. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol., 1999; 292: 195-202.

CONFLICT OF INTEREST:

None declared

ACKNOWLEDGEMENT:

Acknowledgement goes to the Provincial Government of Khyber-Pakhtunkhwa for awarding Laptops to Bushra Ata Hashmi in Naway Sahar Laptop scheme. MAK is sponsored by HEC, Pakistan through IPFP program and funded a start up research grant under grant no PM-IPFP/HRD/HEC/2011/346.