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



# 'BUSHRAATA HASHMl,1MUHAMMAD AYAZ KHAN, 'MUHAMMAD AKHLAQ, 'MUZAMMILAHMAD KHAN

'·' Gomal Centre of Biochemistry and Biotechnology, Gomal University, D.1.Khan, Khyber-Pakhtoonkhuwa, Pakistan ' Department of Pharmaceutics, Faculty of Pharmacy. Gomal University. Dera Ismail Khan, 29050, KPK. Pakistan

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

**CORRESPONDENCE: MUZAMMIL AHMAD KHAN** [E-mail:muzammllgandapur1983@gmall.com](mailto:muzammllgandapur1983@gmall.com)

## t\8 IR 1



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 ac:id 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 inprotein.

Single nucleotide polymorphism is a variation of one base ubstitution which when appears in coding exons (nsSNPs) leads to single amino acid variation. These variations may lead to the fu,ictlonal consequences of protein and ultimately reslilts in a disease phenotype. In the present computational study, various softwares were employed for functional and structural analysis of nsSNPs In the proteln coding exons of PRSS12 (MIM# 606709) gene to determine its deleteriousness. Mutation in this gene causes Autosomal Recessive Non-Syndromic Mental Retardation. The gross blolnformatics 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 inspiteofscreening whole gene.

# I IRODUC11O

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 alwayschange 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. SynonymousSNPs do not aff-ect protein sequence while non-synonymou SNPs change theamino acid sequence with In a protein'. 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'.

PRSSl 2, protease serine 12 (neurolrypsln or motopsln), is **a** 'erlne protease

and belongs to the family of trypsin protease. Cytogenetically; PRSS12 fs 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 retardatfon type 1 (MRT1)1•

In the present 5tudy;computational analysis of nsSNPs in PRSSl 2 gene were performed to determine their phenotypic effects. The analysisdetermined 7 deleterious alleles in various exons. Most of these alleles were targeting SRCR domain of PRSSl 2 protein.The study will aid molecularbiologist In targeted sequencing of PRSSl 2 gene in the linkagestudy of MR patient.

# MAILRI LSANUM[JH D

Fo,r prioritization of candidate exons in PRSS12 gene, the SNP data was acquired from Ensembl Genome Browser (April 2013 release)'.The biological effect for these nsSNPs was simulated by using SIFT(Sorting Intolerant From Tolerant)'•, Polymorphism Phenotyping 2 (PolyPhen 2)", SNAP (screening for non-acceptable polymorphisms)'-" and Mutation taster'"' 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", which is a support vector machine (SVM)-based program employed for automated prediction of protein stability changes upon missense mutation". 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 Bloedlt software". Structural variation in the PRSS12 protein, due to nsSNP, was evaluated by using PSIPRED (protein structure prediction)'', a protein secondary structure prediction tool''.

**RLSULI DISCU !:-1 N**

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 interms of timeand cost to perform the laboratoryexperiments for all SNPs in human genome (or even in a single gene) and elucidate their functional conse• quences.

**Tab!H:**

The predicted pathogenic variunts in PRSS12 gene distributed in

**different exous**

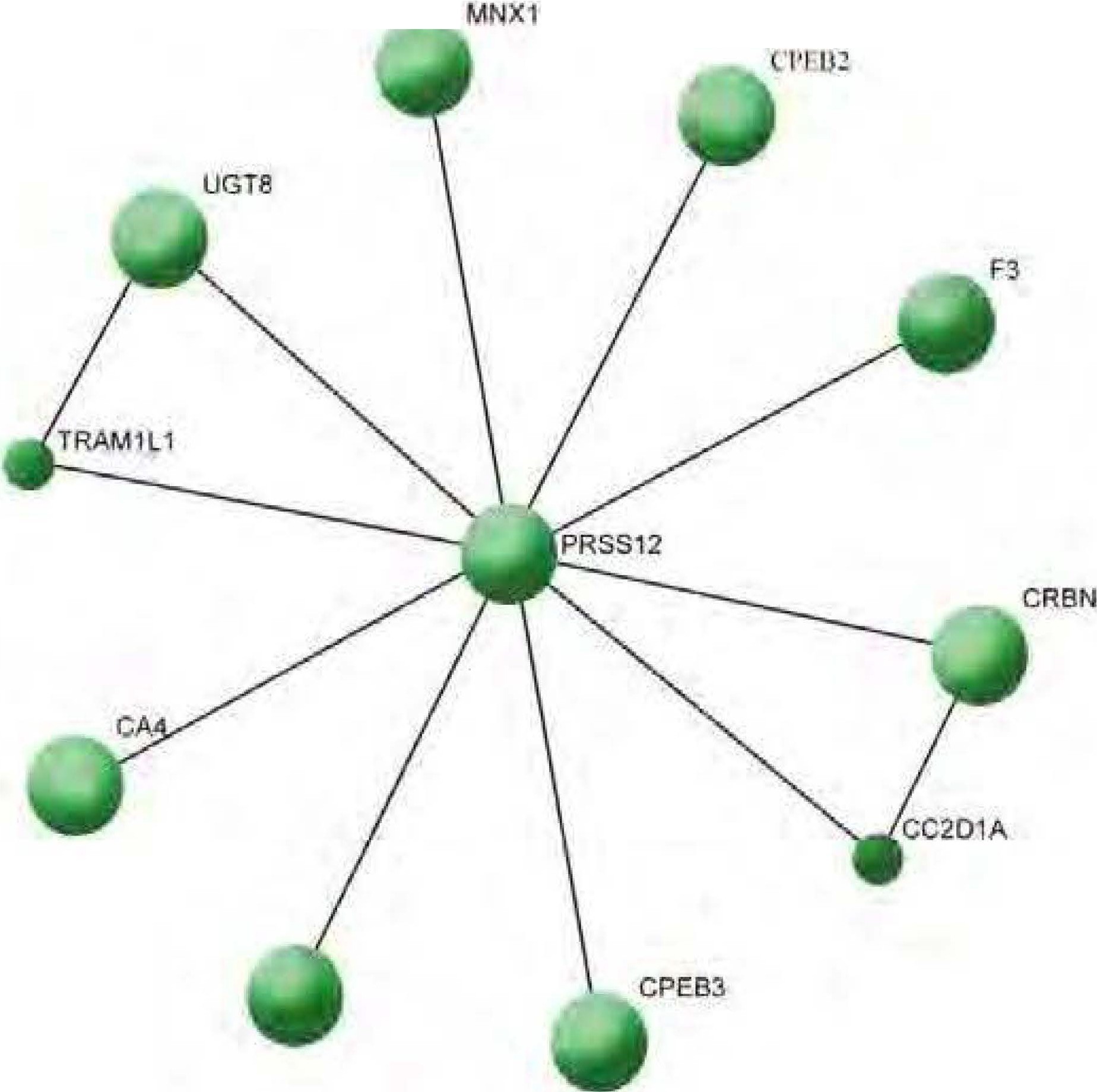
I ll1CUbiC t, ... ,.'11

rnh\11 A1iilll\1,i(!\d 1,,1111o11 In 81v1Mh

lff 1\0.i.l .,.-i,1 1..0\IJ\hU:11111n Oo1ru-fo U mqil a.,,.,,Jlo

Expression studies also revealed its presence in Leydig cell, of testis which explains that it could possibly be involved in fertility. String matching predicted F3 (coagulation factor 111), CC2D1 A and CRBN as strong protein interactor of PRSS12 (see figure I). This interaction study Indicates the possible role of PRSS12 In slgnal transduction pathway.

Figure I:

Thegr:1pl1ic:1I11reseutatio11of PRS 121>ro1ein interactors

T')'l) Ul H1t;hl)'

II

S1ruc1t1rull}

rs-142551196 *GI*

rsll•J&mo *err*

FIL

,vr

o/8

S:-iY

Domain rnl'B<"rved

5RCR lli>:J,ly

9

UQmain cpru;ern d

SRC'll IU!;hly

changed S!Nl.!IUr.tlh· thIDJ:!Cd

SllUciumlly

u1rKl$b<ll

ts:?.0071!>"-SZI

r,,J763U Sll

r,IJ MOo l

Gri

Tl<'

Ar!J

OT

,O

I F,O

C!R

R/Q

*541*

4

455

ti

l)()m11111 ,X.1}1;-C(\(I

SRl'R H1shly

7

1">1111111111 OOIL'!tfVCd

KrR Highly

*1*

Dom:un COlatt\lt:d

.SRr'R H• hly

6

Oommn t:(tf!St"t">1Cd

Ch.Dinged S1nu:1urally dum1wd Stru n1rhlly th;.ugcd

Sl1m:-n1mUy

clum1;cd

**CON LUSIO** :

The Current bioinformatics analysis of nsSNPs in PRSSl 2 gene has determined thirteen most probably damaging nsSNPs.Therefore itis speculated that the present insilico study will assist the molecular

r>72\ltJ3215

,,,11.

GIA

1q7

Kringk lhghl)

I

Domain COflii('f'\ d

Slru,;,:hually

changed

geneticists to selectively sequence the most probable candidate exons that contain the most deleterious nsSNPs, inspite of screening wholegene. Based on the current analytical work, 5 exons viz, 1, 6,7, 9 and 11 are prioritized as candidate and may be regarded as

Computational prediction hasmade it convenient for scientists and



researches to prioritized the most probably damaging nsSNP2. In the present bioinformatics study,variol1s softwares were used to predict the functional consequences of nsSNPs in the coding region of PRSS12 gene. This gene had 109 nsSNPin total (April 2013 release) which were distributed in 13 exons. After bloinformatics analysis, 7 nsSNPs were predicted to be the most deleteriousthat can affect the protein structure & function and may lead to disease phenotype. These prioritized nsSNPs are distributed among exons 1,6, 7,9and 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 evidenceof disease association was not observed by locating Lhese SNPs inHuman Gene Mutation Database.

So far only one deletion mutation is reported in PRSS12 gene which lead to autosom<1I recessive nonsyndromic intellectual disability3\_

mutational prone exon.

## RHCRE CE.S

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.

1. Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, Bacq D,

Briault S, Vekemans M, Munnich A, Attie-Bitach T, Sonderegger P, Colleaux L Truncating neurotrypsln mutation in autosomal recessive nonsyndromic. Mental Retardation. Science, 2002; 29; 298 (5599):1779-1781.

1. Ense m bl Genome Browser retrieved from URL: <http://asia.ensembl.org/index.html>



S. Sorting Intolerant From Tolerant (SIFT) retrieved from URL:

<http://sift.jcvl.org/>

1. 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.
2. Polymorphism Phenotyping 2 (PolyPhen 2) retrieved from URL:

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

1. Adzhubei l.,Jordan OM, Sunyaev SR. Predicting functional effect of human missense mutations using Poly Phen-2. Curr. Protoc. Hum.Genet. 2013; Chapter 7:Unit7.20.
2. SNAP retrieved from URL: <http://www.broad.mit.edu/mpg/> snap/
3. Bromberg Y, Rost B. SNAP: predict effect of nonsynonymous polymorphisms on function. Nucleic. Acids. Research, 2007; 35: 3823-3835.
4. Mutation taster retrieved from URL: <http://www.mutationraster.org/>
5. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations.Nature.Methods.201O; 7:575-576.
6. I-Mutant 2.0 retrieved from URL: <http://folding.biofold.org/1-> mutant/i-mutant2.0.1,tm1

14, Capriotti E, Fariselli, P, Casadio R. I-Mutant2.0:predicting stability

changes upon mutation from the protein 5equenceor structure. Nud. Acids. Res. 2005; 33: W306-W310 doi:10.1093/nar/gki375.

1. 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.

1. PSIPRED retrieved from URL:<http://bioinf.cs.ucl.ac.uk/psipred/>
2. Jones OT. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol., 1999; 292: 195- 202.

CUi',,;FllCT OF l1 TF.RES T·

None declared

**CKNOWLEOG[M[NT:**

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.