International Annals of Science
ISSN: 2456-7132
Volume 8, Issue 1, pp. 96-113, 2020
DOI: https://doi.org/10.21467/ias.8.1.96-113

RESEARCH ARTICLE

Novel Mutations within *PRSS1* Gene that Could Potentially Cause Hereditary Pancreatitis: Using Bioinformatics Approach

Mujahed I. Mustafa[1*], Abdelrahman H. Abdelmoneim[1], Nafisa M. Elfadol[1], Soada A. Osman[1], Tebyan A. Abdelhameed[1], and Mohamed A. Hassan[1,2]

[1] Department of Biotechnology, Africa City of Technology, Khartoum, Sudan
[2] Department of Bioinformatics, DETAGEN Genetics Diagnostic Center, Kayseri, Turkey

* Corresponding author email: mujahedimustafa@gmail.com

Received: 19 July 2019 / Revised: 04 October 2019 / Accepted: 31 October 2019 / Published: 03 November 2019

ABSTRACT

Hereditary pancreatitis (HP) is a rare heterogeneous disease with partial penetrance identified by frequent episodes of severe abdominal pain, often showing in young aged children. It is complicating by chronic pancreatitis, and high rate of pancreatic cancer (up to 40-50%). The aim of this work was to classify the most deleterious mutation in *PRSS1* gene and to predict their influence on the functional and structural level by a variety of bioinformatics analysis tools. The raw data of *PRSS1* gene were recovered from SNP database, and further used to examine a deleterious effect using SIFT, PolyPhen-2, PROVEAN, SNAP2, SNPs&GO, PHD-SNP, PANTHER and P-Mut. The functional analysis predicted that two SNPs "rs1366278558 and rs767036052" have a deleterious effect at functional level. Additionally, we submitted them to I-mutant 3.0, and MUPro respectively to investigate their effect on structural level; the two tools revealed that; two mutations have a dramatic decrease of the protein stability, thus suggesting that the M1R and L4P mutations of PRSS1 gene could destabilize the amino acid interactions causing functional abnormalities of PRSS1 protein. The 3D structure of PRSS1 was predicted by RaptorX and modeled using UCSF Chimera to compare the differences between the native and the mutant amino acids. From the comparative analysis at the functional and structural level, these two SNPs "M1R and L4P" have a deleterious effect and thus could be used as diagnostic markers to predict HP. These findings can be used as a platform to develop large-scale studies in the future.

Keywords: Hereditary pancreatitis, Pancreatic Cancer, Bioinformatics Approach, Single-nucleotide polymorphisms, *PRSS1*.

1 Introduction

Hereditary pancreatitis (HP) is a rare heterogeneous disease with partial penetrance characterized by recurrent severe abdominal pain that manifest on juvenile age children. Progression to chronic pancreatitis is common with high rate of pancreatic malignancy (up to 40-50%) with almost 80% penetrance and flexible fluency.[1-5] The first warning sign appears early before 10 years old, upon which they will be mainly complaining of pancreatic pain (>70%). morphological variations as pancreatic calcifications are diagnosed at early age of 22-25 years. Exocrine and endocrine pancreatic deficiency happened in 34% and 26% of patients.[6, 7] The first family documented to have hereditary pancreatitis was described in 1952.[8, 9] after that about 100 families have been reported [4], which they were encountered by physicians in different countries [3, 10-16]. HP is a progressive inflammatory disease in which pancreatic secretory parenchyma is destroyed and replaced by fibrous tissue [17], as the disease progress it will eventually lead to malnutrition and diabetes.[18] Several studies show that; patients with HP have a significantly elevated risk of developing pancreatic cancer when compared with the overall people up to 40-50%[1, 16, 19-22], Also cigarette Smoking were found to rise the possibility factor for developing pancreatic malignancy in Patients suffering from Hereditary Pancreatitis.[23, 24]

Currently, there is no effective medical treatment for HP, but it can be managed by pancreatic enzyme replacement therapy along with analgesics to control the pain. In addition, endoscopic retrograde cholangiopancreato-graphy (ERCP) and surgical intersections are reserved for complicated cases.[25-31]

HP has mainly been related with mutations in the serine protease 1 gene (PRSS1). It has been mapped to chromosome 7 q35.[5, 20] Among all genes that was mentioned in the literature to have associations with hereditary pancreatitis PRSS1 gene is the most reported one [1, 5-7, 15, 26, 30, 32-36], PRSS1 gene are found to encode human cationic trypsinogen. Furthermore Most high penetrance PRSS1 modifications will increase intra pancreatic trypsin activity. [37] Interestingly, mutations in PRSS1 gene may protect against the disease.[38, 39] In several studies (R122H), mutation in PRSS1 gene was strongly associated with hereditary pancreatitis [6, 25, 32, 40, 41]. The genetic makeup of the mutation in *PRSS1* gene and full pathogenesis of by which it causes the disease is still unclear.[5] In vitro functional and characterization studies, is a highly demanding task in terms of workload, time and financial cost. For these reasons, computational analysis is an appropriate alternative that is more rapid and low-cost approach, which is why it has been used to study many types of inheritance diseases in the past years [42-44] to enrich our knowledge of the ways mutations could affect protein structure and function. The main objective of this work was to classify the most damaging SNPs that could be used as diagnostic markers. An extensive in silico approach using multiple software was used in this study, our result can be used as a platform to develop larger-scale studies in the future.

2 Materials and Methods

2.1 Data mining

The raw data of *PRSS1* gene were retrieved from NCBI website by selecting the gene view from the main result page then downloading the resulted table in to an excel sheet[45], furthermore the protein reference sequence was collected from Uniprot[46]

2.2 Functional Investigation of Damaging SNPs

2.2.1 SIFT

Is the first in silico method for functional analysis, which calculates whether an amino acids alteration change protein function, or not. SIFT scores < 0.05 are expected to be damaging altered amino acid, otherwise it considered to be tolerant.[47, 48] from the single protein tools we choose sift sequence, then we inserted our reference sequence with the original parameters unchanged and download the resulting values. The disadvantage about this tool is the inability to provide result for two sequence variations with the identical position in the same time.

2.2.2 Polyphen-2

It's a trained machine learning to predict the possible effect of amino acid replacement on protein function and structure, by calculating Position Specific Independent Count (PSIC) for each SNP at the time. It provides accurate predictions with three possible output whether probably damaging (values are more rapidly to one), possibly damaging or benign (values are varieties from zero to 0.95.) [49, 50] we inserted the amino acid sequences after preparing it as the web site specification in to the batch query area on the site.

2.2.3 PROVEAN

It's very fast and accurate online in silico functional analysis tool that calculates whether specific amino acid replacement has an effect on the biological function of a protein depending on the alignment-based score. PROVEAN probability has two possibilities, deleterious or neutral with cutoff -2.5.[51] we choose (provean protein) choice from the web side and then inserted the reference sequence along with the amino acid variations without changing the parameter. Like SIFT, it can help the researcher identify mistakes in preparing the amino acid variations.

2.2.4 SNAP2

It is a trained functional analysis tool that differentiates between effect and neutral SNPs by taking a variety of features into validation. It

consumes more time to work with the result than other tools and it got an accuracy of 83%, with two expectations, effect (positive score) or neutral (negative score). But it still considered an important and substantial enhancement over other methods. [52]we directly run the prediction after entering the reference protein sequence.

2.2.5 SNPs & GO

It is a trained machine learning based on the technique to precisely calculate the deleterious associated alterations from protein sequence. SNPs&GO collects in unique framework information derived from protein sequence, evolutionary information, and function as coded in the Gene Ontology terms. SNPs&GO performs other prediction methods (PHD-SNP and PANTHER) [53] we uploaded both the reference sequence and the amino acid variations then submit them and downloaded the result . The website provides friendly user environment with fast and reliable result.

2.2.6 P-Mut

It is a web-based tool for the explanation of SNPs alternates on proteins; it is characterized by fast and precise calculation. The mutations can be predicted to be either Neutral or disease causing. [54] we analyses the sequence and the variations through (analyzing mutations) link on the web side using PMut2017 predictor.

2.3 Stability Investigation

2.3.1 I-Mutant 3.0

It's a structural analysis online tool for the routine analysis of protein stability by considering the single-site alterations. Negative I-Mutant scores are expected to decrease the protein stability, otherwise (positive) it considered to increase it.[55] It is relatively easy to use website. We selected the (protein sequence) category that is related to (Prediction of protein stability changes upon single point mutations) then we inserted the reference sequence, the position and the residue and highlighted the (DDG value and binary classification method) before finally submitting these inputs.

2.3.2 MUPro

It is a structural analysis online tool for the calculation of protein stability variations upon arbitrarily SNPs. The value of the energy change is expected, and assurance mark between -1 and one for evaluating the assurance of the expectation is calculated. A score < 0 means the mutant decreases the protein stability; conversely, a score > 0 means the mutant increases the protein stability.[56]

2.4 ConSurf server

It is a web server offers evolutionary conservation summaries for proteins of known structure in the protein data bank. ConSurf spot the parallel amino acid sequences and run multi alignment methods. The conserved amino acid across species detects its position using specific algorisms.[57, 58]

2.5 BioEdit

It is a software package proposed to stream a distinct program that can run nearly any sequences operation as well as a few basic alignment investigations.[59] The FASTA format sequences of PRSS1 protein were retrieved from UniProt and used as an input to locate and determine if the SNPs are located at conserved sites or not. Through ClustalW choice in the accessory applications.

2.6 GeneMANIA

It is a method to identify protein function and gene - gene interactions; it integrates multiple genomics and proteomics data to create reliable information about the function of unknown proteins, although some time it fails to know the functions of some proteins.[60] we inserted the gene name then we download the relevant result into an excel sheet.

2.7 3D Clustering Analysis

2.7.1 Mutation3D

It is a functional calculation and visualization online tool for investigating the three-dimensional plan of amino acid alterations on protein models and structures. The input formats were the gene name and the SPNs of interest. [61]

2.8 Biophysical Validation & Visualization analysis

2.8.1 Project HOPE

It is a webserver to search protein 3D structures by bringing together structural information from several sources such as UniProt database. The main aims for the submissions in Project HOPE are to analysis and confirm results that we had it earlier. [62] we upload the sequence without the identifications line then we defined the positions with the related amino acids. The advantage in using Project HOPE is the detailed information provided for each SNPs variation, but the main disadvantage would be the delay in the results that sometimes occurs for hours.

2.8.2 Displaying Amino Acid Mutations

For this task we used UCSF Chimera, It's for visualization and investigation of SNPs at the molecular level. Protein in Pdb format can be viewed by UCSF Chimera to scan the native and the mutant amino acids to observe the alterations that occur. [63] we study the structural changes mainly through (structure editing and rotamers) tools in the chimera software .

2.9 Variant Effect Predictor (VEP)

The Ensembl Variant Effect Predictor software provides toolsets for an organized approach to annotate and aid prioritization of variants in both large-scale sequencing projects and smaller analysis studies.[64]

3 Results and Discussion

3.1 Results

Data related to the total number of SNPs in different regions of *PRSS1* gene was retrieved from dbSNP database with the distribution of SNPs in coding and non-coding regions of PRSS1 gene (figure 1). Out of 911 SNPs there are 506 SNPs, contained 339 nsSNPs, 133 synonymous, 17 frame shit and 17 nonsense, with 26 in the 3'-UTR region and 11 in the 5'-UTR region (figure 2).

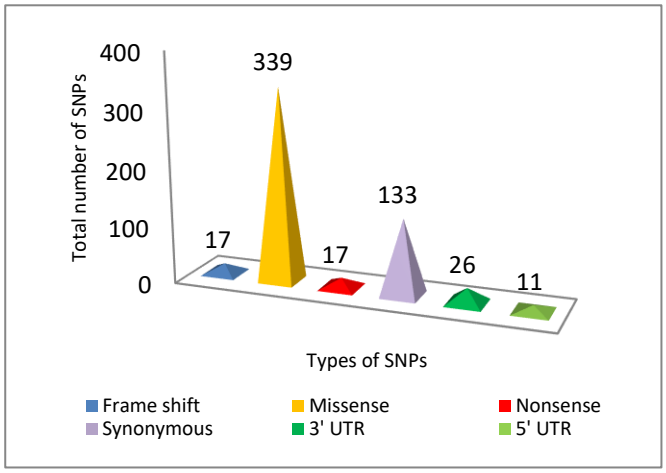

Figure 2: *Graphic representation of PRSS1 gene workflow.*

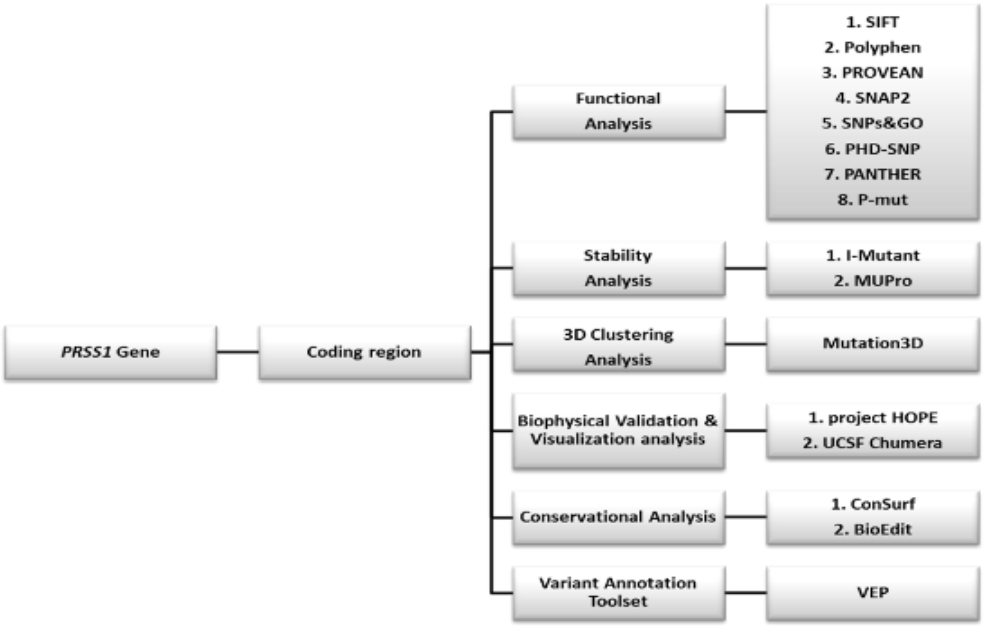

Figure 1: *the distribution of SNPs in coding and non-coding regions of PRSS1 gene.*

These 339 nsSNPs were used to examine the deleterious effect on the associated protein. To identify the most deleterious nsSNPs, we toke the combined deleterious results from 8 functional tools, only the highly deleterious SNPs by SIFT, PolyPhen, PROVEAN, SNAP2, SNPs&GO, PHD-SNP, PANTHER and P-Mut, meet the criteria as illustrated in (Table 1,2 and 3). Remarkably, only two SNPs (M1R and L4P) give positive results with all the tools.

Table 1: *Shows the affected SNPs that investigate by several online tools (*SUB: Substitutions (variants))*

dbSNP rs#	SUB	SIFT		Polyphen		PROVEAN		SNAP2	
		prediction	Score	prediction	Score	Prediction	Score	prediction	Score
rs1366278558	M1R	Damaging	0	probably damaging	1	Deleterious	-4.183	effect	51
rs780969708	C30W	Damaging	0	probably damaging	1	Deleterious	-9.449	effect	66
rs769459903	P36R	Damaging	0	probably damaging	1	Deleterious	-7.82	effect	59
rs138464021	G49D	Damaging	0	probably damaging	1	Deleterious	-6.273	effect	90
-	G49V	Damaging	0	probably damaging	1	Deleterious	-8.061	effect	87
-	L52F	Damaging	0	probably damaging	1	Deleterious	-3.628	effect	32
rs149246646	I53N	Damaging	0	probably damaging	1	Deleterious	-6.266	effect	68
rs778570468	W57G	Damaging	0	probably damaging	1	Deleterious	-11.76	effect	83
rs1338646513	W57C	Damaging	0	probably damaging	1	Deleterious	-11.77	effect	63
rs1192452565	V58G	Damaging	0	probably damaging	1	Deleterious	-6.083	effect	73
rs370761165	A61E	Damaging	0	probably damaging	1	Deleterious	-4.397	effect	80
rs372411481	G83R	Damaging	0	probably damaging	1	Deleterious	-7.011	effect	40
-	G83W	Damaging	0	probably damaging	1	Deleterious	-7.169	effect	5
rs1268805560	H96Y	Damaging	0	probably damaging	1	Deleterious	-5.652	effect	72
rs1209409723	H96Q	Damaging	0	probably damaging	1	Deleterious	-7.515	effect	78
rs1454816504	Y99C	Damaging	0	probably damaging	1	Deleterious	-8.117	effect	72
rs1323769980	D107Y	Damaging	0	probably damaging	1	Deleterious	-8.448	effect	94
rs1426710453	L113H	Damaging	0	probably damaging	1	Deleterious	-6.567	effect	79
rs144403091	V123M	Damaging	0	probably damaging	1	Deleterious	-2.664	effect	21
rs749518244	L128Q	Damaging	0	probably damaging	1	Deleterious	-5.698	effect	49
-	L128P	Damaging	0	probably damaging	1	Deleterious	-6.649	effect	74
rs768673799	P129H	Damaging	0	probably damaging	1	Deleterious	-8.254	effect	45
rs748208676	I141N	Damaging	0	probably damaging	1	Deleterious	-6.392	effect	89
rs1164996242	S142P	Damaging	0	probably damaging	1	Deleterious	-4.777	effect	86
rs1164331073	W144C	Damaging	0	probably damaging	1	Deleterious	-12.47	effect	80
rs1221038304	G145R	Damaging	0	probably damaging	1	Deleterious	-7.632	effect	91
rs1172272446	C160R	Damaging	0	probably damaging	1	Deleterious	-11.07	effect	28
rs778796800	C160Y	Damaging	0	probably damaging	1	Deleterious	-10.15	effect	22
-	C160F	Damaging	0	probably damaging	1	Deleterious	-10.15	effect	41
rs200973660	C171Y	Damaging	0	probably damaging	1	Deleterious	-10.21	effect	80
-	C171S	Damaging	0	probably damaging	1	Deleterious	-9.272	effect	78
rs756821075	Y175N	Damaging	0	probably damaging	1	Deleterious	-8.298	effect	75
-	Y175H	Damaging	0	probably damaging	1	Deleterious	-4.616	effect	66
rs1217657614	D194V	Damaging	0	probably damaging	1	Deleterious	-8.36	effect	85
rs763907908	C196G	Damaging	0	probably damaging	1	Deleterious	-11.15	effect	93
rs1412477456	G201R	Damaging	0	probably damaging	1	Deleterious	-7.194	effect	87
rs1288010897	G201V	Damaging	0	probably damaging	1	Deleterious	-8.04	effect	83
rs1289842951	P203H	Damaging	0	probably damaging	1	Deleterious	-8.323	effect	65
rs747422004	V205D	Damaging	0	probably damaging	1	Deleterious	-5.836	effect	81
rs1366495669	S215P	Damaging	0	probably damaging	1	Deleterious	-4.618	effect	93
rs1481112469	C220W	Damaging	0	probably damaging	1	Deleterious	-10.13	effect	90
rs1164573795	P226S	Damaging	0	probably damaging	1	Deleterious	-7.318	effect	63
rs1309672836	Y229H	Damaging	0	probably damaging	1	Deleterious	-4.582	effect	85

Table 2: *List of SNPs analyzed for disease association by three online servers (*RI: Reliability Index)*

dbSNP rs#	Mutation	PHD-SNP Prediction	RI	Probability	SNP &Go Prediction	RI	Probability	PANTHER Prediction	RI	Probability
rs1366278558	M1R	Disease	0	0.514	Disease	3	0.668	Disease	9	0.95
rs780969708	C30W	Disease	6	0.825	Disease	7	0.848	Disease	10	0.99
rs769459903	P36R	Disease	4	0.708	Disease	1	0.558	Disease	5	0.734
rs138464021	G49D	Disease	8	0.917	Disease	7	0.862	Disease	8	0.88
-	G49V	Disease	8	0.913	Disease	7	0.852	Disease	8	0.902
-	L52F	Disease	7	0.841	Disease	6	0.783	Disease	9	0.954
rs149246646	I53N	Disease	8	0.888	Disease	7	0.844	Disease	8	0.891
rs778570468	W57G	Disease	8	0.897	Disease	6	0.787	Disease	10	0.98
rs1338646513	W57C	Disease	9	0.942	Disease	7	0.849	Disease	10	0.993
rs1192452565	V58G	Disease	6	0.818	Disease	4	0.695	Disease	7	0.848
rs370761165	A61E	Disease	8	0.891	Disease	5	0.762	Disease	6	0.779
rs372411481	G83R	Disease	3	0.67	Disease	3	0.667	Disease	4	0.715
-	G83W	Disease	5	0.756	Disease	4	0.687	Disease	9	0.952
rs1268805560	H96Y	Disease	8	0.879	Disease	5	0.741	Disease	4	0.695
rs1209409723	H96Q	Disease	7	0.867	Disease	5	0.728	Disease	3	0.673
rs1454816504	Y99C	Disease	7	0.867	Disease	5	0.756	Disease	9	0.927
rs1323769980	D107Y	Disease	9	0.933	Disease	7	0.829	Disease	9	0.973
rs1426710453	L113H	Disease	8	0.905	Disease	5	0.765	Disease	9	0.935
rs144403091	V123M	Disease	0	0.511	Disease	0	0.507	Disease	4	0.681
rs749518244	L128Q	Disease	5	0.753	Disease	4	0.7	Disease	7	0.869
-	L128P	Disease	5	0.765	Disease	4	0.7	Disease	8	0.902
rs768673799	P129H	Disease	3	0.66	Disease	1	0.545	Disease	7	0.845
rs748208676	I141N	Disease	6	0.794	Disease	6	0.823	Disease	7	0.87
rs1164996242	S142P	Disease	6	0.812	Disease	7	0.853	Disease	6	0.817
rs1164331073	W144C	Disease	9	0.929	Disease	6	0.805	Disease	8	0.918
rs1221038304	G145R	Disease	7	0.866	Disease	6	0.812	Disease	9	0.966
rs1172272446	C160R	Disease	5	0.757	Disease	8	0.876	Disease	10	0.987
rs778796800	C160Y	Disease	3	0.674	Disease	7	0.834	Disease	10	0.992
-	C160F	Disease	5	0.76	Disease	7	0.866	Disease	10	0.99
rs200973660	C171Y	Disease	8	0.925	Disease	7	0.827	Disease	10	1
-	C171S	Disease	8	0.895	Disease	5	0.772	Disease	10	1
rs756821075	Y175N	Disease	7	0.835	Disease	4	0.698	Disease	7	0.873
-	Y175H	Disease	3	0.661	Disease	2	0.589	Disease	6	0.806
rs1217657614	D194V	Disease	8	0.904	Disease	7	0.872	Disease	8	0.898
rs763907908	C196G	Disease	6	0.816	Disease	5	0.743	Disease	10	0.999
rs1412477456	G201R	Disease	8	0.882	Disease	7	0.839	Disease	7	0.849
rs1288010897	G201V	Disease	8	0.906	Disease	7	0.855	Disease	7	0.84
rs1289842951	P203H	Disease	7	0.835	Disease	4	0.701	Disease	10	0.979
rs747422004	V205D	Disease	7	0.831	Disease	6	0.775	Disease	8	0.903
rs1366495669	S215P	Disease	8	0.878	Disease	7	0.827	Disease	6	0.804
rs1481112469	C220W	Disease	9	0.928	Disease	7	0.855	Disease	9	0.967
rs1164573795	P226S	Disease	7	0.871	Disease	6	0.787	Disease	9	0.933
rs1309672836	Y229H	Disease	6	0.795	Disease	1	0.573	Disease	5	0.742

Table 3: *Show a List of SNPs investigated by P-Mut*

dbSNP rs#	Amino Acid change	P-mut Prediction	P-mut Score
rs1366278558	M1R	Decrease	0.78 (88%)
rs767036052	L4P	Decrease	0.70 (86%)

Table 4: *Structural Investigation predicted by using I-Mutant v3.0 and MUPro*

dbSNP rs#	Amino Acid change	SVM2 Prediction Effect	RI	DDG Value Prediction	MUPro Prediction	MUPro Score
rs1366278558	M1R	Decrease	1	-0.62	Decrease	-1.0183
rs767036052	L4P	Decrease	8	-1.66	Decrease	-1.6239

*SVM: Support Vector Machine. *DDG Value: free energy changes value.

Further additional analysis were held for only these 2 SNPs.we submitted them to I-mutant 3.0, and MUPro respectively to investigate their effect on the structural level; the two tools revealed that, there is a dramatic decrease in the protein stability, thus suggesting that the M1R and L4P mutations of PRSS1 could destabilize the amino acid interactions causing functional abnormalities of PRSS1 protein. (Table 4)

3.2 Discussion

Two mutations were predicted to have a potential major impact on the structure and function of the *PRSS1* protein by using different computational analysis tools (Figure 2). The approaches used were grounded on different characteristics and limitations that unmask the pathogenicity and deliver evidences about the influence of each mutation. In the past years in silico analysis has been done for several inherited diseases and tumor associated genes along with other disorders. [43, 65, 66] in this study we used computational in silico analysis of *PRSS1* gene to study nsSNP effect on the PRSS1 protein that could lead to the development of Hereditary Pancreatitis . These SNPs were submitted to SIFT, PolyPhen, PROVEAN, and SNAP2 sever; we found 141 SNPs to be damaging by SIFT. In PolyPhen2, the results showed that 170 SNPs were found to be damaging (54 possibly damaging and 116 probably damaging showed deleterious). In PROVEAN server our result showed that 203 SNPs were predicted to be deleterious. While in SNAP2 server the result showed that 172 SNPs were predicted to be deleterious. The alterations in calculation abilities and result are likely duo to the fact that every prediction algorithm uses different sets of sequences and alignments. we submitted the four positive combined SNPs results from SIFT, PolyPhen-2, PROVEAN and SNAP2 (Table 1) to be analyzed further by disease related softwares: SNPs & GO, PHD-SNP, PANTHER(Table2), P-Mut and MUPro servers. There was (85, 95 and 90) SNPs founded to be disease related by SNPs & GO, PHD-SNP and PANTHER servers respectively. While MUPro servers revealed different and unique results of only 2 disease related SNPs as shown in (Table3). We selected only the four positive combined results (disease-causing SNPs). (Figure 3) illustrate the relation between these soft wares. Additionally, we performed analysis by Mutation3D; our result shows that: (L4P) located in a domain, which indicates its vital significance (Figure 4).

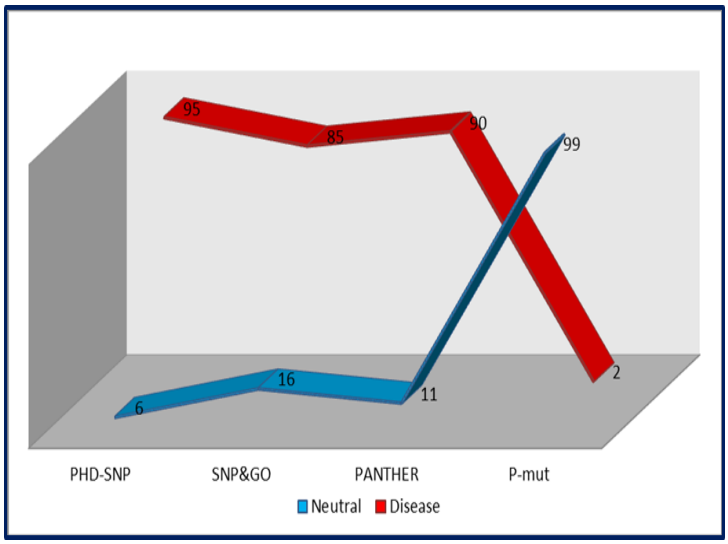

Figure 3: *Shows Illustration of damaging mutations predicted by numerous of tools.*

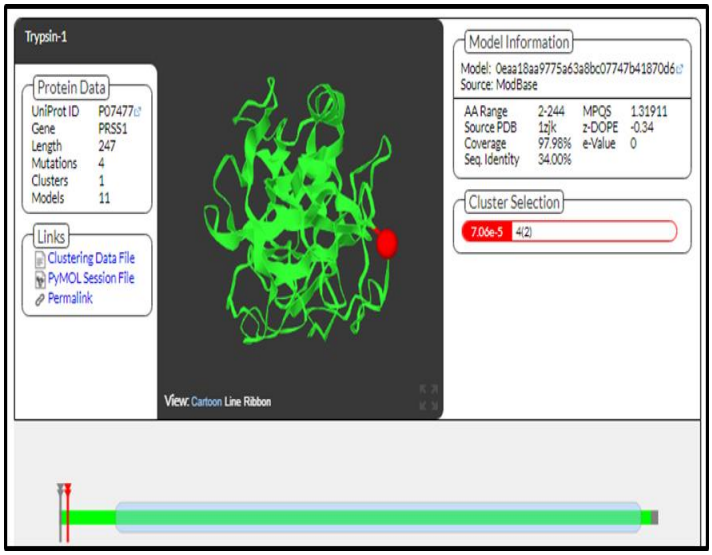

Figure 4: *Screenshot for (L4P) shows high-possibility SNP in its domain.*

The 3D protein structure analysis enables mapping of amino acid substitutions and, therefore, RaptorX was used to make a 3D structure model for PRSS1 protein (Figure 5).

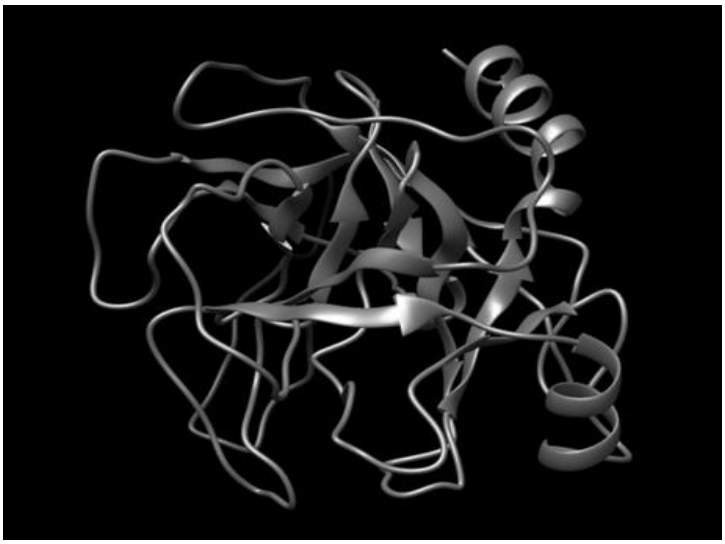

Figure 5: *The 3D structure of PRSS1 protein model was generated by using RaptorX.*

To supports and matches the results acquired from different computational tools, we used UCSF Chimera. (Figure 6-7) shows the differences between native and mutant amino acids illustrated here in the green and red boxes. The schematic structures of the native amino acids are in the left side and the mutant ones are in the right side. The backbone, which is the same for each amino acid, is colored red and the side chain, unique for each amino acid is colored black, the 3D wide type residues colored green and mutant ones colored red, while the protein is colored dark gray. Project HOPE server was used to submit the two most deleterious nsSNPs (M1R) and (L4P). (rs1366278558): (M1R): Methionine changed to Arginine at position 1. As showed in (Figure 6) this may decrease the protein stability which disrupts the amino acid interactions. (rs767036052):(L4P): Leucine residue changed to Proline at position 4. As showed in (Figure 7) the altered remain is smaller; this may cause loss of interactions.

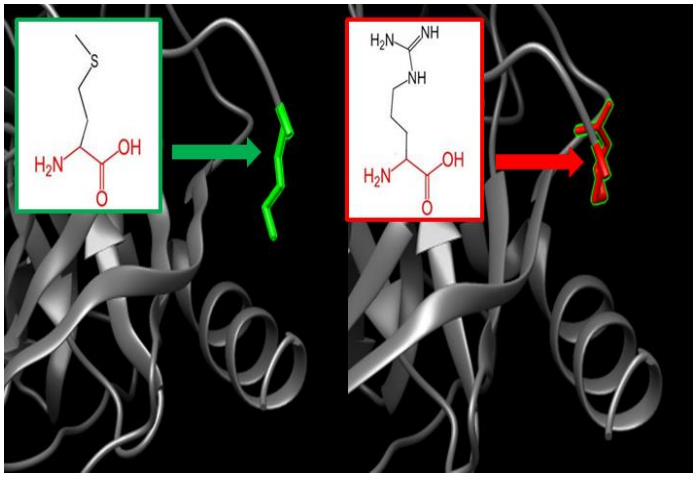

Figure 6: *Shows Methionine changes to Arginine at position 1.*

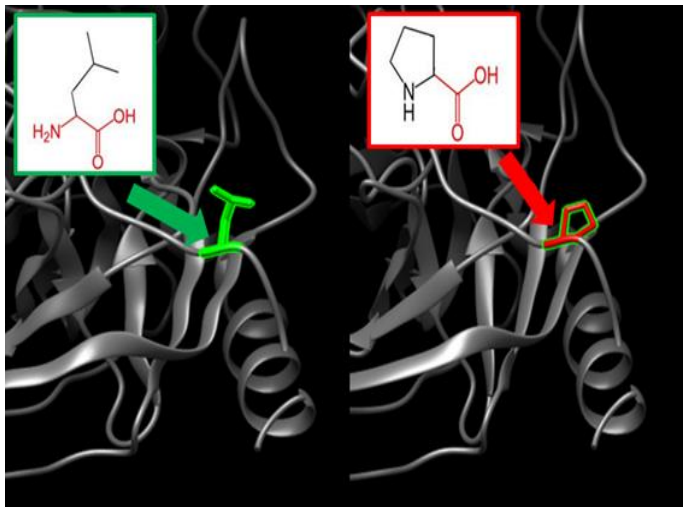

Figure 7: *Shows Leucine changes to Proline at position 4.*

We also used ConSurf to flag the SNPs that are sited at highly conserved amino acid positions, which has a tendency be more damaging than SNPs that are sited at non- preserved positions. Our ConSurf analysis unmasked that (L4P) mutation was found in highly conserved site and expected to have a high influence on *PRSS1*

protein structure and function as illustrated in (Figure 8). To confirm our findings on (M1R & L4P) mutations, we used BioEdit (version 7.2.5) where Alignment of 10 amino acid arrangements of *PRSS1* confirm their conservation and hence significance as evident in (Figure 9).

GeneMANIA revealed that *PRSS1* has many vital functions: blood microparticle, cobalamin metabolic process, extracellular matrix organization, extracellular structure organization, serine hydrolase activity, serine-type endopeptidase activity, serine-type peptidase activity. The associated genes contribute to accomplish similar function were demonstrated by GeneMANIA. on (Figure 10) and (Table 5-6). The VEP annotates variants using a wide range of reference data. That include transcripts, regulatory regions, and frequencies from previously observed variants, citations, clinical significance information, and predictions of biophysical consequences of variants , and that what makes VEP give accurate results [64].

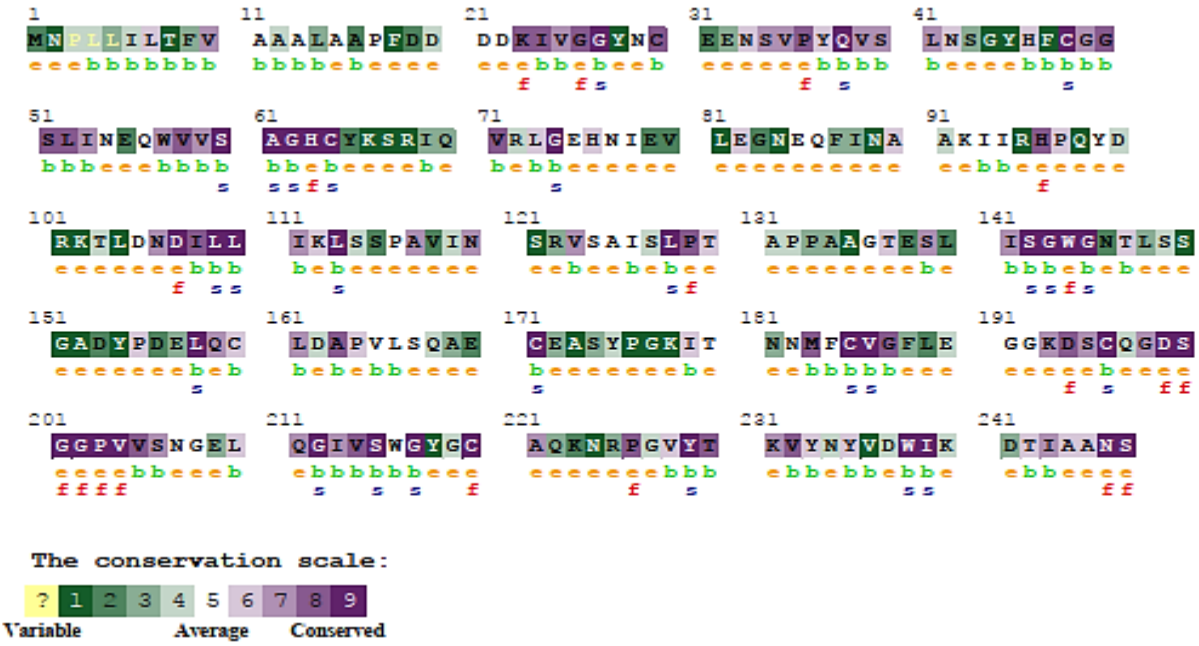

Figure 8: *Shows the conserved amino acids across species in PRSS1 protein were determined using Consurf. (e) An exposed residues according to the neural-network algorithm via an orange letter. (b) Residues predicted to be buried are demonstrated via a green letter. (f) A predicted functional residues (highly conserved and exposed) are indicated with a red letter. (s) A predicted structural residues (highly conserved and buried) that are demonstrated with a blue letter. (?) Insufficient data- the calculation for this site was performed on less than 10% of the sequences are demonstrated via a yellow letter.*

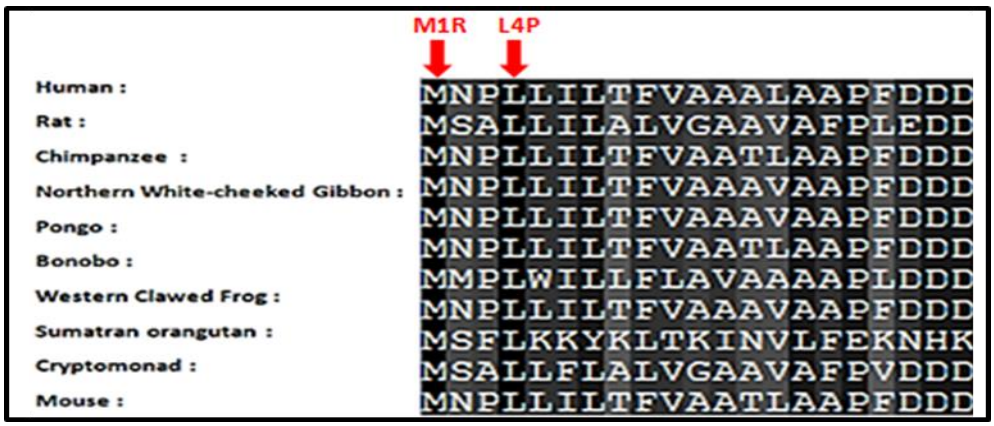

Figure 9: *Alignment of 10 amino acid sequences of PRSS1 demonstrating that the residues predicted to be mutated (indicated by red arrows) are evolutionarily conserved across species.*

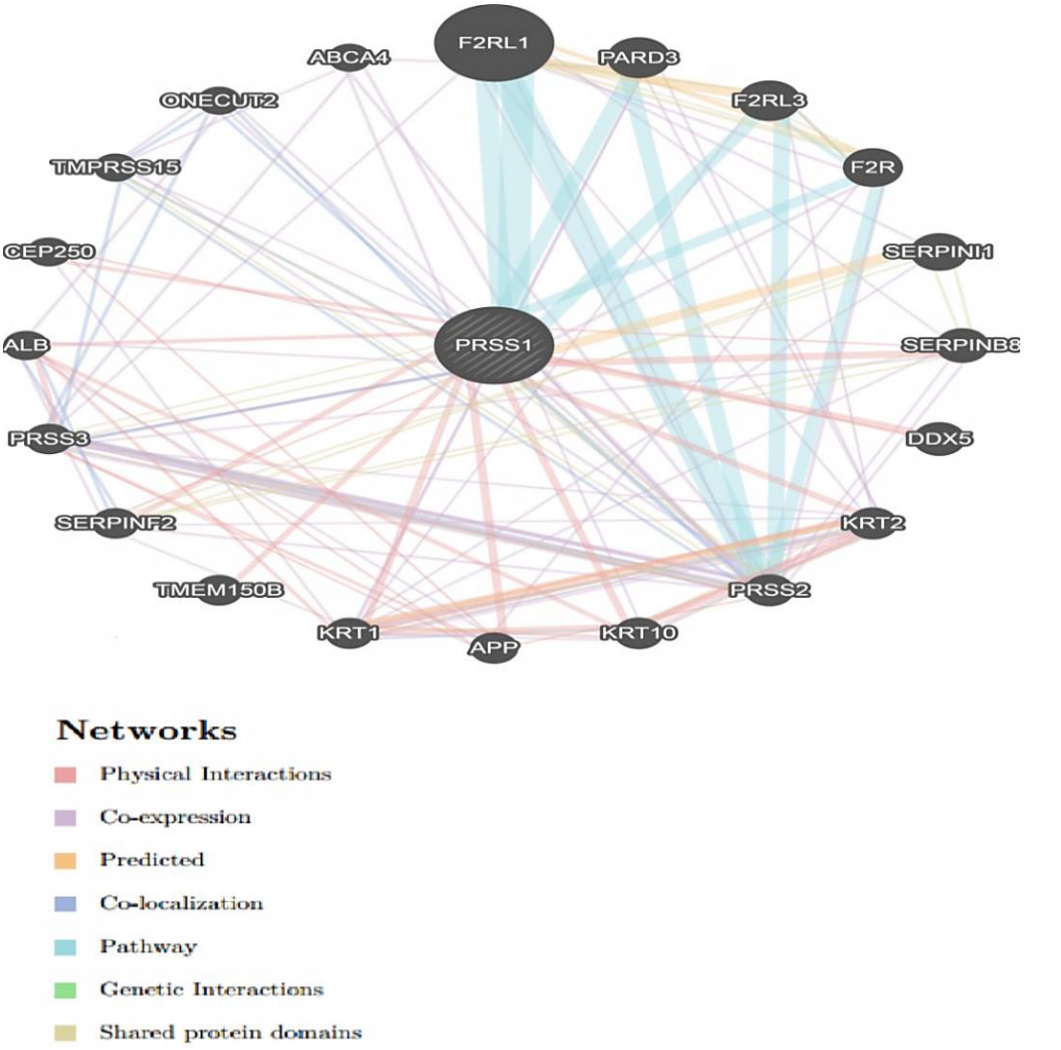

Figure 10: *Shows the relations between PRSS1 and its associated genes.*

Table 5: *PRSS1 gene functions and its appearance in network and genome*

Function	FDR	Genes in network	Genes in genome
serine-type endopeptidase inhibitor activity	0.001356361	4	44
renal system process involved in regulation of systemic arterial blood pressure	0.001356361	3	13
regulation of hemostasis	0.001356361	4	54
regulation of coagulation	0.001356361	4	56
regulation of blood coagulation	0.001356361	4	54
regulation of wound healing	0.003496886	4	74
platelet activation	0.006856052	5	211
endopeptidase inhibitor activity	0.007599469	4	99
endopeptidase regulator activity	0.007599469	4	102
peptidase inhibitor activity	0.007599469	4	101
blood microparticle	0.008676071	4	108
regulation of endopeptidase activity	0.009340491	5	251
regulation of peptidase activity	0.009856636	5	258
peptidase regulator activity	0.013382385	4	128
renal system process	0.013666888	3	45
endocrine process	0.013666888	3	42
negative regulation of endopeptidase activity	0.013666888	4	140
regulation of blood vessel size	0.013666888	3	44
regulation of tube size	0.013666888	3	44
negative regulation of peptidase activity	0.013666888	4	142
regulation of systemic arterial blood pressure	0.013666888	3	43

platelet alpha granule lumen	0.01586832	3	48
vascular process in circulatory system	0.028313203	3	59
platelet alpha granule	0.029997483	3	61
secretory granule lumen	0.030241005	3	62
serine-type endopeptidase activity	0.033116705	3	67
killing of cells of other organism	0.033116705	2	10
exocytosis	0.033116705	4	191
disruption of cells of other organism	0.033116705	2	10
glomerular filtration	0.037840129	2	11
renal filtration	0.037840129	2	11
vesicle lumen	0.040192639	3	76
enzyme inhibitor activity	0.040192639	4	214
protein kinase C-activating G-protein coupled receptor signaling pathway	0.040192639	2	12
cytoplasmic membrane-bounded vesicle lumen	0.040192639	3	76
positive regulation of collagen metabolic process	0.044298773	2	13
positive regulation of collagen biosynthetic process	0.044298773	2	13
regulation of blood pressure	0.044298773	3	81
platelet degranulation	0.044771122	3	82
positive regulation of ERK1 and ERK2 cascade	0.04525757	3	83
negative regulation of multicellular organismal process	0.045315882	4	230
positive regulation of coagulation	0.048108433	2	15
negative regulation of hydrolase activity	0.048108433	4	242
collagen metabolic process	0.048108433	3	87
positive regulation of hemostasis	0.048108433	2	15
positive regulation of blood coagulation	0.048108433	2	15
positive regulation of multicellular organismal metabolic process	0.048108433	2	15
multicellular organismal macromolecule metabolic process	0.049591958	3	91
regulation of endocrine process	0.051022071	2	16
regulation of collagen biosynthetic process	0.051022071	2	16
tissue homeostasis	0.051393559	3	94
zymogen activation	0.052534202	3	97
regulation of collagen metabolic process	0.052534202	2	17
multicellular organismal metabolic process	0.052534202	3	97
positive regulation of release of sequestered calcium ion into cytosol	0.052534202	2	17
serine-type peptidase activity	0.052961891	3	98
serine hydrolase activity	0.055058507	3	101
cellular metal ion homeostasis	0.055058507	4	264
positive regulation of Ras protein signal transduction	0.055058507	2	18
collagen biosynthetic process	0.057743853	2	19
multicellular organismal homeostasis	0.057743853	3	105
regulation of multicellular organismal metabolic process	0.057743853	2	19
positive regulation of small GTPase mediated signal transduction	0.057743853	2	19
cobalamin metabolic process	0.0619817	2	20
fibrinolysis	0.0619817	2	20
cellular cation homeostasis	0.063339568	4	284
regulation of ERK1 and ERK2 cascade	0.063981924	3	111
metal ion homeostasis	0.064852029	4	288
cellular ion homeostasis	0.066330871	4	292
extracellular matrix organization	0.066330871	4	292
extracellular structure organization	0.066330871	4	293
ERK1 and ERK2 cascade	0.07491573	3	120
positive regulation of calcium ion transport into cytosol	0.094036985	2	26
positive regulation of cytosolic calcium ion concentration	0.094313684	3	131

*FDR: false discovery rate is greater than or equal to the probability that this is a false positive.

Table 6: *The gene co-expression, shared domain, and interaction with PRSS1gene network*

Gene 1	Gene 2	Weight	Network group
F2R	*PRSS1*	0.025958601	Co-expression
PRSS2	*PRSS1*	0.03579626	Co-expression
PRSS2	*F2R*	0.025450852	Co-expression
KRT1	*KRT2*	0.021397633	Co-expression
PRSS3	*PRSS1*	0.033738803	Co-expression

PRSS3	*PRSS2*	0.033134628	Co-expression
ABCA4	*PRSS1*	0.020958284	Co-expression
ABCA4	*PRSS2*	0.021833802	Co-expression
ABCA4	*PRSS3*	0.020297663	Co-expression
F2R	*F2RL1*	0.01723061	Co-expression
SERPINF2	*KRT1*	0.005162225	Co-expression
PRSS3	*PRSS2*	0.030308522	Co-expression
ABCA4	*PARD3*	0.018756393	Co-expression
SERPINI1	*F2RL1*	0.018079963	Co-expression
PRSS2	*PRSS1*	0.010663025	Co-expression
KRT10	*PRSS1*	0.004751287	Co-expression
KRT1	*KRT2*	0.017388888	Co-expression
PRSS3	*PRSS2*	0.016354393	Co-expression
ALB	*PRSS1*	0.00978991	Co-expression
TMPRSS15	*PRSS1*	0.009836646	Co-expression
TMPRSS15	*PRSS2*	0.02349308	Co-expression
ONECUT2	*PRSS1*	0.011885054	Co-expression
ONECUT2	*PRSS2*	0.026482629	Co-expression
ONECUT2	*ALB*	0.023143126	Co-expression
ONECUT2	*TMPRSS15*	0.023315145	Co-expression
KRT2	*PARD3*	0.00407993	Co-expression
KRT2	*F2RL3*	0.002086929	Co-expression
KRT1	*PARD3*	0.008348665	Co-expression
KRT1	*KRT2*	0.001395255	Co-expression
SERPINF2	*PRSS2*	0.020101048	Co-expression
PRSS3	*PRSS2*	0.037385598	Co-expression
KRT2	*F2RL1*	0.004147036	Co-expression
KRT2	*PARD3*	0.004057049	Co-expression
PRSS2	*SERPINB8*	0.013921468	Co-expression
KRT1	*PARD3*	0.007636325	Co-expression
KRT1	*F2RL3*	0.001890512	Co-expression
KRT1	*KRT2*	0.000903638	Co-expression
SERPINF2	*KRT2*	0.001538189	Co-expression
PRSS3	*SERPINB8*	0.012880304	Co-expression
PRSS3	*PRSS2*	0.01267024	Co-expression
PRSS3	*PRSS2*	0.008046081	Co-expression
ALB	*SERPINF2*	0.012912146	Co-expression
PRSS3	*PRSS2*	0.059784386	Co-expression
ABCA4	*PRSS1*	0.021043219	Co-expression
PRSS2	*PRSS1*	0.017638749	Co-expression
PRSS3	*PRSS1*	0.010804671	Co-expression
PRSS3	*PRSS2*	0.008761909	Co-expression
ABCA4	*TMPRSS15*	0.014334416	Co-expression
KRT10	*KRT2*	0.013258844	Co-expression
KRT1	*KRT2*	0.012602167	Co-expression
KRT1	*KRT10*	0.010437634	Co-expression
PRSS3	*PRSS2*	0.020688048	Co-expression
APP	*F2R*	0.007284051	Co-expression
KRT1	*SERPINI1*	0.011799906	Co-expression
PRSS3	*KRT2*	0.01028535	Co-expression
PRSS3	*PRSS2*	0.056584958	Co-expression
SERPINB8	*F2RL1*	0.007467608	Co-expression

PRSS2	*F2RL1*	0.009436603	Co-expression
PRSS3	*F2RL1*	0.0091086	Co-expression
PRSS3	*PRSS2*	0.006767724	Co-expression
PRSS2	*SERPINB8*	0.005437602	Co-expression
CEP250	*KRT1*	0.007838975	Co-expression
ALB	*APP*	0.015414801	Co-expression
PRSS2	*F2RL1*	0.011420856	Co-expression
KRT10	*KRT2*	0.031353086	Co-expression
KRT1	*SERPINB8*	0.017290482	Co-expression
KRT1	*KRT2*	0.030572487	Co-expression
ALB	*SERPINF2*	0.011540175	Co-localization
F2R	*F2RL3*	0.004708032	Co-localization
PRSS2	*PRSS1*	0.025921002	Co-localization
KRT1	*KRT10*	0.025425598	Co-localization
PRSS3	*PRSS1*	0.023983192	Co-localization
PRSS3	*PRSS2*	0.023682926	Co-localization
ALB	*SERPINF2*	0.016614223	Co-localization
TMPRSS15	*PRSS1*	0.02293591	Co-localization
TMPRSS15	*PRSS2*	0.022705158	Co-localization
TMPRSS15	*PRSS3*	0.022577295	Co-localization
ONECUT2	*PRSS1*	0.022678863	Co-localization
ONECUT2	*PRSS2*	0.022438433	Co-localization
ONECUT2	*PRSS3*	0.02270408	Co-localization
ONECUT2	*TMPRSS15*	0.026521208	Co-localization
F2RL1	*PRSS1*	0.31954107	Pathway
PARD3	*PRSS1*	0.15361263	Pathway
F2RL3	*PRSS1*	0.1483753	Pathway
F2R	*PRSS1*	0.12154161	Pathway
F2R	*F2RL3*	0.03525233	Pathway
PRSS2	*F2RL1*	0.31954107	Pathway
PRSS2	*PARD3*	0.15361263	Pathway
PRSS2	*F2RL3*	0.1483753	Pathway
PRSS2	*F2R*	0.12154161	Pathway
F2RL1	*PRSS1*	0.7365974	Pathway
CEP250	*PRSS1*	0.08479679	Physical Interactions
CEP250	*DDX5*	0.021656184	Physical Interactions
ALB	*PRSS1*	0.0910383	Physical Interactions
ALB	*KRT10*	0.0910383	Physical Interactions
ALB	*KRT1*	0.04481957	Physical Interactions
ALB	*PRSS3*	0.0910383	Physical Interactions
DDX5	*PRSS1*	0.3375174	Physical Interactions
KRT10	*KRT2*	0.37723482	Physical Interactions
KRT1	*KRT2*	0.38693836	Physical Interactions
KRT1	*KRT10*	0.20642798	Physical Interactions
KRT2	*PRSS1*	0.28754097	Physical Interactions
KRT10	*PRSS1*	0.24892193	Physical Interactions
KRT10	*KRT2*	0.15666465	Physical Interactions
KRT1	*PRSS1*	0.24077669	Physical Interactions
KRT1	*KRT2*	0.15153825	Physical Interactions
KRT1	*KRT10*	0.13118546	Physical Interactions
SERPINB8	*PRSS1*	0.4951709	Physical Interactions
SERPINF2	*PRSS1*	0.28534	Physical Interactions

APP	*PRSS1*	0.41637108	Physical Interactions
PRSS3	*APP*	0.13214332	Physical Interactions
TMEM150B	*PRSS1*	0.6281027	Physical Interactions
ALB	*APP*	0.0232307	Physical Interactions
CEP250	*APP*	0.0232307	Physical Interactions
SERPINB8	*PRSS1*	0.18267727	Physical Interactions
APP	*PRSS2*	0.030401716	Physical Interactions
SERPINF2	*PRSS1*	0.0573582	Physical Interactions
PRSS3	*APP*	0.05769704	Physical Interactions
F2RL3	*F2RL1*	0.5863869	Predicted
F2R	*F2RL1*	0.17300643	Predicted
KRT1	*KRT2*	0.15433232	Predicted
SERPINI1	*PRSS1*	1	Predicted
F2RL3	*F2RL1*	0.09312839	Shared protein domains
F2R	*F2RL1*	0.09312839	Shared protein domains
F2R	*F2RL3*	0.09312839	Shared protein domains
SERPINB8	*SERPINI1*	0.028828265	Shared protein domains
PRSS2	*PRSS1*	0.008043572	Shared protein domains
KRT10	*KRT2*	0.008737453	Shared protein domains
KRT1	*KRT2*	0.024351321	Shared protein domains
SERPINF2	*SERPINI1*	0.028828265	Shared protein domains
SERPINF2	*SERPINB8*	0.028828265	Shared protein domains
PRSS3	*PRSS1*	0.010362946	Shared protein domains
PRSS3	*PRSS2*	0.010437737	Shared protein domains
TMPRSS15	*PRSS1*	0.006456545	Shared protein domains
TMPRSS15	*PRSS2*	0.006503143	Shared protein domains
F2RL3	*F2RL1*	0.00354377	Shared protein domains
F2R	*F2RL1*	0.00354377	Shared protein domains
F2R	*F2RL3*	0.00354377	Shared protein domains
SERPINB8	*SERPINI1*	0.027777778	Shared protein domains
PRSS2	*PRSS1*	0.009871521	Shared protein domains
KRT10	*KRT2*	0.010121632	Shared protein domains
KRT1	*KRT2*	0.018882168	Shared protein domains
KRT1	*KRT10*	0.011125098	Shared protein domains
SERPINF2	*SERPINI1*	0.027777778	Shared protein domains
SERPINF2	*SERPINB8*	0.027777778	Shared protein domains
PRSS3	*PRSS1*	0.009871521	Shared protein domains
PRSS3	*PRSS2*	0.011186602	Shared protein domains

Table 7: *Maintenance outline of amino acids in PRSS1*

Residues position	Residues	CS Score normalized	Color	Residues Variety	B/E
1	M	0.211	7.3	M,N,P	e
4	L	0.361	6.3	M,S,L,I,H,F	b

CS Score normalized (1-4= variable, 5= average, 6-9= conserved).
B/E: Buried (b) or Exposed (e) residue.

Table 8*: Shows variants consequences, impact and biotype features by VEP tool*

Uploaded variation	SNP	Location	Consequence	IMPACT	BIOTYPE
rs1366278558	M/R	7:142749486-142749486	start lost	HIGH	protein coding
rs1366278558	-	7:142749486-142749486	upstream gene variant	MODIFIER	retained intron
rs1366278558	-	7:142749486-142749486	non coding transcript exon variant	MODIFIER	retained intron

rs1366278558	M/R	7:142749486-142749486	start lost	HIGH	protein coding
rs1366278558	M/R	7:142749486-142749486	upstream gene variant	MODIFIER	protein coding
rs1366278558	M/R	7:142749486-142749486	non coding transcript exon variant	MODIFIER	retained intron
rs1366278558	M/R	7:142749486-142749486	start lost	HIGH	protein coding
rs1366278558	M/R	7:142749486-142749486	regulatory region variant	MODIFIER	open chromatin region
rs767036052	L/P	7:142749495-142749495	missense variant	MODERATE	protein coding
rs767036052	L/P	7:142749495-142749495	upstream gene variant	MODIFIER	retained intron
rs767036052	L/P	7:142749495-142749495	non coding transcript exon variant	MODIFIER	retained intron
rs767036052	L/P	7:142749495-142749495	missense variant	MODERATE	protein coding
rs767036052	L/P	7:142749495-142749495	upstream gene variant	MODIFIER	protein coding
rs767036052	L/P	7:142749495-142749495	non coding transcript exon variant	MODIFIER	retained intron
rs767036052	L/P	7:142749495-142749495	missense variant	MODERATE	protein coding
rs767036052	L/P	7:142749495-142749495	regulatory region variant	MODIFIER	open chromatin region
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	start lost	HIGH	protein coding
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	start lost	HIGH	protein coding
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	non coding transcript exon variant	MODIFIER	retained intron
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	upstream gene variant	MODIFIER	retained intron
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	non coding transcript exon variant	MODIFIER	retained intron
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	upstream gene variant	MODIFIER	protein coding
rs1366278558	M/R	CHR_HSCHR7_2_CTG6:142790203-142790203	start lost	HIGH	protein coding
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	missense variant	MODERATE	protein coding
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	missense variant	MODERATE	protein coding
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	non coding transcript exon variant	MODIFIER	retained intron
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	upstream gene variant	MODIFIER	retained intron
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	non coding transcript exon variant	MODIFIER	retained intron
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	upstream gene variant	MODIFIER	protein coding
rs767036052	L/P	CHR_HSCHR7_2_CTG6:142790212-142790212	missense variant	MODERATE	protein coding

The predicted variants consequences are shown in (Tables 8), VEP reported regulatory consequences for many variants, including 6 variants within a coding region, 2 variants within a non-coding region, 8 variants within upstream gene, 8 variants within noncoding transcript exon and 6 variant within start lost codon. in general any mutations within a coding region will likely affect the protein function, while regulatory variants within non-coding genomic regions can greatly affect the expression of protein[67, 68], the SNPs in the upstream, 5'UTR region might affect transcription or translation process[69]

In the light of our work, we agree with previous studies linking (P36R and V123M) with chronic pancreatitis [70] . We also support the previous findings relating these mutations to certain forms of hereditary pancreatitis. (P36R and V123M) may be associated with pancreatic cancer along other types of cancer and it has also been related in the past to familial Peutz-Jeghers syndrome [72-75], so this study can be used as a platform to develop large scale studies in the future in relation

to these disease . This study is the first computational analysis of *PRSS1 gene* which was based on functional and structural analysis while all earlier studies [31, 76] foxed on frequency and Whole exome sequencing. Furthermore, this study revealed two novel mutations (M1R, and L4P) that had a possible functional influence, which means that these SNPs could be used as diagnostic biomarkers for HP. Further wet lab studies are needed to confirm these results.

4 Conclusion

A total of two SNPs was predicted to have potential responsibility for the functional and structural alterations of PRSS1 gene. It is predicted from comparison of the results between various bioinformatics analysis tools; Out of a total of 911 SNPs in the PRSS1 gene, 506 were nsSNPs; out of 506 nsSNPs, two were found to be the most deleterious nsSNPs (M1R and L4P) by eight functional analysis tools. Stability analysis results showed a dramatic decrease of the protein stability. These two SNPs may assist as diagnostic biomarkers for the prognosis of HP and may be used as a platform to develop large-scale studies in the future.

5 Declarations

5.1 Data Availability

All data underlying the results are available as part of the article and no additional source data are required.

5.2 Competing Interests

The authors declare that there is no conflict of interest regarding the publication of this work.

How to Cite this Article:

M. Mustafa, A. Abdelmoneim, N. Elfadol, S. Osman, T. Abdelhameed, and M. Hassan, “Novel Mutations within PRSS1 Gene that Could Potentially Cause Hereditary Pancreatitis: Using Bioinformatics Approach”, *Int. Ann. Sci.*, vol. 8, no. 1, pp. 96-113, Nov. 2019. doi:10.21467/ias.8.1.96-113

References

[1] R. M. Charnley, "Hereditary pancreatitis," *World J Gastroenterol*, vol. 9, pp. 1-4, Jan 2003.

[2] W. D. Jackson, "Pancreatitis: etiology, diagnosis, and management," *Curr Opin Pediatr*, vol. 13, pp. 447-51, Oct 2001.

[3] L. N. Dai, Y. W. Chen, W. H. Yan, L. N. Lu, Y. J. Tao, and W. Cai, "Hereditary pancreatitis of 3 Chinese children: Case report and literature review," *Medicine (Baltimore)*, vol. 95, p. e4604, Sep 2016.

[4] L. Le Bodic, J. D. Bignon, O. Raguenes, B. Mercier, T. Georgelin, M. Schnee, *et al.*, "The hereditary pancreatitis gene maps to long arm of chromosome 7," *Hum Mol Genet*, vol. 5, pp. 549-54, Apr 1996.

[5] P. Dyrla, T. Nowak, J. Gil, C. Adamiec, M. Bobula, and M. Saracyn, "[Hereditary pancreatitis]," *Pol Merkur Lekarski*, vol. 40, pp. 113-6, Feb 2016.

[6] V. Rebours, P. Levy, and P. Ruszniewski, "An overview of hereditary pancreatitis," *Dig Liver Dis*, vol. 44, pp. 8-15, Jan 2012.

[7] N. Howes, M. M. Lerch, W. Greenhalf, D. D. Stocken, I. Ellis, P. Simon, *et al.*, "Clinical and genetic characteristics of hereditary pancreatitis in Europe," *Clin Gastroenterol Hepatol*, vol. 2, pp. 252-61, Mar 2004.

[8] S. K. Lee, "[Hereditary pancreatitis]," *Korean J Gastroenterol*, vol. 46, pp. 358-67, Nov 2005.

[9] K. L. Raphael and F. F. Willingham, "Hereditary pancreatitis: current perspectives," *Clin Exp Gastroenterol*, vol. 9, pp. 197-207, 2016.

[10] A. Masamune, K. Kikuta, S. Hamada, E. Nakano, K. Kume, A. Inui, *et al.*, "Nationwide survey of hereditary pancreatitis in Japan," *J Gastroenterol*, vol. 53, pp. 152-160, Jan 2018.

[11] J. Mora, L. Comas, E. Ripoll, P. Goncalves, J. M. Queralto, F. Gonzalez-Sastre, *et al.*, "Genetic mutations in a Spanish population with chronic pancreatitis," *Pancreatology*, vol. 9, pp. 644-51, 2009.

[12] S. Raty, A. Piironen, M. Babu, H. Pelli, J. Sand, S. Uotila, *et al.*, "Screening for human cationic trypsinogen (PRSS1) and trypsinogen inhibitor gene (SPINK1) mutations in a Finnish family with hereditary pancreatitis," *Scand J Gastroenterol*, vol. 42, pp. 1000-5, Aug 2007.

[13] J. R. Sibert, "Hereditary pancreatitis in England and Wales," *J Med Genet*, vol. 15, pp. 189-201, Jun 1978.

[14] H. Goebell, R. Ammann, and W. Creutzfeldt, "History of the European pancreatic club: the first 40 years 1965-2005. The development of the European pancreatic club as a scientific society," *Pancreatology*, vol. 5 Suppl 1, pp. 1-15, 2005.

[15] S. E. Applebaum-Shapiro, R. Finch, R. H. Pfutzer, L. A. Hepp, L. Gates, S. Amann, *et al.*, "Hereditary pancreatitis in North America: the Pittsburgh-Midwest Multi-Center Pancreatic Study Group Study," *Pancreatology*, vol. 1, pp. 439-43, 2001.

[16] M. T. Joergensen, K. Brusgaard, D. G. Cruger, A. M. Gerdes, and O. B. Schaffalitzky de Muckadell, "Genetic, epidemiological, and clinical aspects of hereditary pancreatitis: a population-based cohort study in Denmark," *Am J Gastroenterol*, vol. 105, pp. 1876-83, Aug 2010.

[17] A. B. Atlas, S. R. Orenstein, and D. M. Orenstein, "Pancreatitis in young children with cystic fibrosis," *J Pediatr*, vol. 120, pp. 756-9, May 1992.

[18] A. Masamune, "Genetics of pancreatitis: the 2014 update," *Tohoku J Exp Med*, vol. 232, pp. 69-77, Feb 2014.

[19] C. Langner, "[Hereditary gastric and pancreatic cancer]," *Pathologe*, vol. 38, pp. 164-169, May 2017.

[20] A. B. Lowenfels, P. Maisonneuve, E. P. DiMagno, Y. Elitsur, L. K. Gates, Jr., J. Perrault, *et al.*, "Hereditary pancreatitis and the risk of pancreatic cancer.

International Hereditary Pancreatitis Study Group," *J Natl Cancer Inst*, vol. 89, pp. 442-6, Mar 19 1997.

[21] D. Malka, P. Hammel, F. Maire, P. Rufat, I. Madeira, F. Pessione, *et al.*, "Risk of pancreatic adenocarcinoma in chronic pancreatitis," *Gut*, vol. 51, pp. 849-52, Dec 2002.

[22] F. U. Weiss, "Pancreatic cancer risk in hereditary pancreatitis," *Front Physiol*, vol. 5, p. 70, 2014.

[23] A. B. Lowenfels, P. Maisonneuve, D. C. Whitcomb, M. M. Lerch, and E. P. DiMagno, "Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis," *Jama*, vol. 286, pp. 169-70, Jul 11 2001.

[24] A. B. Lowenfels, P. Maisonneuve, and D. C. Whitcomb, "Risk factors for cancer in hereditary pancreatitis. International Hereditary Pancreatitis Study Group," *Med Clin North Am*, vol. 84, pp. 565-73, May 2000.

[25] A. Mastoraki, A. Tzortzopoulou, S. Tsela, N. Danias, G. Sakorafas, V. Smyrniotis, *et al.*, "Hereditary pancreatitis: dilemmas in differential diagnosis and therapeutic approach," *J Gastrointest Cancer*, vol. 45, pp. 22-6, Mar 2014.

[26] M. R. Patel, A. L. Eppolito, and F. F. Willingham, "Hereditary pancreatitis for the endoscopist," *Therap Adv Gastroenterol*, vol. 6, pp. 169-79, Mar 2013.

[27] S. Kargl, M. Kienbauer, H. C. Duba, R. Schofl, and W. Pumberger, "Therapeutic step-up strategy for management of hereditary pancreatitis in children," *J Pediatr Surg*, vol. 50, pp. 511-4, Apr 2015.

[28] J. LaRusch, S. Solomon, and D. C. Whitcomb, "Pancreatitis Overview," in *GeneReviews((R))*, M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, *et al.*, Eds., ed Seattle (WA): University of Washington, Seattle

University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., 1993.

[29] E. D. Rivera Rivera, "[Pancreatitis, genes and islet cells auto transplant; updates and new horizons]," *Rev Gastroenterol Peru*, vol. 37, pp. 156-161, Apr-Jun 2017.

[30] S. Solomon, D. C. Whitcomb, and J. LaRusch, "PRSS1-Related Hereditary Pancreatitis," in *GeneReviews((R))*, M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, *et al.*, Eds., ed Seattle (WA): University of Washington, Seattle

University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., 1993.

[31] J. LaRusch, M. M. Barmada, S. Solomon, and D. C. Whitcomb, "Whole exome sequencing identifies multiple, complex etiologies in an idiopathic hereditary pancreatitis kindred," *Jop*, vol. 13, pp. 258-62, May 10 2012.

[32] J. Y. Kim, S. H. Choi, J. S. Ihm, S. J. Kim, I. J. Kim, and C. M. Kim, "[A case of R122H mutation of cationic trypsinogen gene in a pediatric patient with hereditary pancreatitis complicated by pseudocyst and hemosuccus pancreaticus]," *Korean J Gastroenterol*, vol. 45, pp. 130-6, Feb 2005.

[33] T. Jamer and B. Iwanczak, "[Genetic mutations as a cause of acute recurrent pancreatitis in children - case report and literature review]," *Dev Period Med*, vol. 20, pp. 228-234, 2016.

[34] A. Lal and D. R. Lal, "Hereditary pancreatitis," *Pediatr Surg Int*, vol. 26, pp. 1193-9, Dec 2010.

[35] C. Le Marechal, E. Masson, J. M. Chen, F. Morel, P. Ruszniewski, P. Levy, *et al.*, "Hereditary pancreatitis caused by triplication of the trypsinogen locus," *Nat Genet*, vol. 38, pp. 1372-4, Dec 2006.

[36] P. Simon, F. U. Weiss, M. Sahin-Toth, M. Parry, O. Nayler, B. Lenfers, *et al.*, "Hereditary pancreatitis caused by a novel PRSS1 mutation (Arg-122 --> Cys) that alters autoactivation and autodegradation of cationic trypsinogen," *J Biol Chem*, vol. 277, pp. 5404-10, Feb 15 2002.

[37] B. C. Nemeth and M. Sahin-Toth, "Human cationic trypsinogen (PRSS1) variants and chronic pancreatitis," *Am J Physiol Gastrointest Liver Physiol*, vol. 306, pp. G466-73, Mar 2014.

[38] J. M. Chen, C. Le Marechal, D. Lucas, O. Raguenes, and C. Ferec, ""Loss of function" mutations in the cationic trypsinogen gene (PRSS1) may act as a protective factor against pancreatitis," *Mol Genet Metab*, vol. 79, pp. 67-70, May 2003.

[39] D. C. Whitcomb, M. C. Gorry, R. A. Preston, W. Furey, M. J. Sossenheimer, C. D. Ulrich, *et al.*, "Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene," *Nat Genet*, vol. 14, pp. 141-5, Oct 1996.

[40] T. Y. Lee, H. C. Oh, M. H. Kim, S. Kwon, S. S. Lee, D. W. Seo, *et al.*, "[Three cases of hereditary pancreatitis in two households in the same family associated with R122H mutation in cationic trypsinogen gene]," *Korean J Gastroenterol*, vol. 49, pp. 395-9, Jun 2007.

[41] C. Hanck, A. Schneider, and D. C. Whitcomb, "Genetic polymorphisms in alcoholic pancreatitis," *Best Pract Res Clin Gastroenterol*, vol. 17, pp. 613-23, Aug 2003.

[42] M. I. Mustafa, T. A. Abdelhameed, F. A. Abdelrhman, S. A. Osman, and M. A. Hassan, "Novel Deleterious nsSNPs within MEFV Gene that Could Be Used as Diagnostic Markers to Predict Hereditary Familial Mediterranean Fever: Using Bioinformatics Analysis," *Adv Bioinformatics*, vol. 2019, p. 1651587, 2019.

[43] Z. Kosaloglu, J. Bitzer, N. Halama, Z. Huang, M. Zapatka, A. Schneeweiss, *et al.*, "In silico SNP analysis of the breast cancer antigen NY-BR-1," *BMC Cancer*, vol. 16, p. 901, Nov 18 2016.

[44] Y. H. Bakhit, M. O. Ibrahim, M. Amin, Y. A. Mirghani, and M. A. Hassan, "In Silico Analysis of SNPs in PARK2 and PINK1 Genes That Potentially Cause Autosomal Recessive Parkinson Disease," *Adv Bioinformatics*, vol. 2016, p. 9313746, 2016.

[45] D. A. Benson, I. Karsch-Mizrachi, K. Clark, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic Acids Res*, vol. 40, pp. D48-53, Jan 2012.

[46] "UniProt: the universal protein knowledgebase," *Nucleic Acids Res*, vol. 45, pp. D158-d169, Jan 4 2017.

[47] N. L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P. C. Ng, "SIFT web server: predicting effects of amino acid substitutions on proteins," *Nucleic Acids Res*, vol. 40, pp. W452-7, Jul 2012.

[48] *SIFT server*. Available: https://sift.bii.a-star.edu.sg/

[49] E. Capriotti and R. B. Altman, "Improving the prediction of disease-related variants using protein three-dimensional structure," *BMC Bioinformatics*, vol. 12 Suppl 4, p. S3, 2011.

[50] *PolyPhen-2 server*. Available: http://genetics.bwh.harvard.edu/pph2/

[51] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan, "Predicting the functional effect of amino acid substitutions and indels," *PLoS One*, vol. 7, p. e46688, 2012.

[52] M. Hecht, Y. Bromberg, and B. Rost, "Better prediction of functional effects for sequence variants," *BMC Genomics*, vol. 16 Suppl 8, p. S1, 2015.

[53] R. Calabrese, E. Capriotti, P. Fariselli, P. L. Martelli, and R. Casadio, "Functional annotations improve the predictive score of human disease-related mutations in proteins," *Hum Mutat*, vol. 30, pp. 1237-44, Aug 2009.

[54] V. Lopez-Ferrando, A. Gazzo, X. de la Cruz, M. Orozco, and J. L. Gelpi, "PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update," *Nucleic Acids Res*, vol. 45, pp. W222-w228, Jul 3 2017.

[55] E. Capriotti, P. Fariselli, and R. Casadio, "I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure," *Nucleic Acids Res*, vol. 33, pp. W306-10, Jul 1 2005.

[56] J. Cheng, A. Randall, and P. Baldi, "Prediction of protein stability changes for single-site mutations using support vector machines," *Proteins*, vol. 62, pp. 1125-32, Mar 1 2006.

[57] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, and N. Ben-Tal, "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids," *Nucleic Acids Res*, vol. 38, pp. W529-33, Jul 2010.

[58] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, *et al.*, "ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules," *Nucleic Acids Res*, vol. 44, pp. W344-50, Jul 8 2016.

[59] O. O. Oni, A. A. Owoade, and C. A. O. Adeyefa, "Design and evaluation of primer pairs for efficient detection of avian rotavirus," *Trop Anim Health Prod*, vol. 50, pp. 267-273, Feb 2018.

[60] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, *et al.*, "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function," *Nucleic Acids Res*, vol. 38, pp. W214-20, Jul 2010.

[61] M. J. Meyer, R. Lapcevic, A. E. Romero, M. Yoon, J. Das, J. F. Beltran, *et al.*, "mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome," *Hum Mutat*, vol. 37, pp. 447-56, May 2016.

[62] *project HOPE server*. Available: http://www.cmbi.ru.nl/hope

[63] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, *et al.*, "UCSF Chimera--a visualization system for exploratory research and analysis," *J Comput Chem*, vol. 25, pp. 1605-12, Oct 2004.

[64] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, *et al.*, "The Ensembl Variant Effect Predictor," *Genome Biol*, vol. 17, p. 122, Jun 6 2016.

[65] C. George Priya Doss, C. Sudandiradoss, R. Rajasekaran, P. Choudhury, P. Sinha, P. Hota, *et al.*, "Applications of computational algorithm tools to identify functional SNPs," *Funct Integr Genomics*, vol. 8, pp. 309-16, Nov 2008.

[66] G. Chandrasekaran, E. C. Hwang, T. W. Kang, D. D. Kwon, K. Park, J. J. Lee, *et al.*, "In silico analysis of the deleterious nsSNPs (missense) in the homeobox domain of human HOXB13 gene responsible for hereditary prostate cancer," vol. 90, pp. 188-199, Aug 2017.

[67] E. Rojano, P. Seoane, J. A. G. Ranea, and J. R. Perkins, "Regulatory variants: from detection to predicting impact," *Brief Bioinform*, Jun 8 2018.

[68] S. Dhamija and M. B. Menon, "Non-coding transcript variants of protein-coding genes - what are they good for?," *RNA Biol*, vol. 15, pp. 1025-1031, 2018.

[69] T. V. Tatarinova, E. Chekalin, Y. Nikolsky, S. Bruskin, D. Chebotarov, K. L. McNally, *et al.*, "Nucleotide diversity analysis highlights functionally important genomic regions," *Sci Rep*, vol. 6, p. 35730, Oct 24 2016.

[70] A. Schnur, S. Beer, H. Witt, P. Hegyi, and M. Sahin-Toth, "Functional effects of 13 rare PRSS1 variants presumed to cause chronic pancreatitis," *Gut*, vol. 63, pp. 337-43, Feb 2014.

[71] J. M. Chen, A. Piepoli Bis, L. Le Bodic, P. Ruszniewski, M. Robaszkiewicz, P. H. Deprez, *et al.*, "Mutational screening of the cationic trypsinogen gene in a large cohort of subjects with idiopathic chronic pancreatitis," *Clin Genet*, vol. 59, pp. 189-93, Mar 2001.

[72] S. Hansford, P. Kaurah, H. Li-Chang, M. Woo, J. Senz, H. Pinheiro, *et al.*, "Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond," *JAMA Oncol*, vol. 1, pp. 23-32, Apr 2015.

[73] A. W. Kurian, E. E. Hare, M. A. Mills, K. E. Kingham, L. McPherson, A. S. Whittemore, *et al.*, "Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment," *J Clin Oncol*, vol. 32, pp. 2001-9, Jul 1 2014.

[74] L. Wang, J. Huang, M. Jiang, and X. Zheng, "AFP computational secreted network construction and analysis between human hepatocellular carcinoma (HCC) and no-tumor hepatitis/cirrhotic liver tissues," *Tumour Biol*, vol. 31, pp. 417-25, Oct 2010.

[75] R. H. Hruban, M. I. Canto, and C. J. Yeo, "Prevention of pancreatic cancer and strategies for management of familial pancreatic cancer," *Dig Dis*, vol. 19, pp. 76-84, 2001.

[76] F. U. Weiss, M. Zenker, A. B. Ekici, P. Simon, J. Mayerle, and M. M. Lerch, "Local clustering of PRSS1 R122H mutations in hereditary pancreatitis patients from Northern Germany," *Am J Gastroenterol*, vol. 103, pp. 2585-8, Oct 2008.