In silico and Genotypic Analysis of Single Nucleotide Polymorphisms (SNPs) of Protein Tyrosine Phosphatase Non-Receptor Type 1 (PTPN1) in Type 2 Diabetes Mellitus (T2DM)

BS Thesis

Submitted By

Manal Khan 280683

Faiqa Hafeez 278995

Heba Khan 280689

Namra Tariq 279701

BS Applied Biosciences

Supervisor

Prof. Dr. Attya Bhatti

Co-supervisor

Prof. Dr. Peter John

Submission Date

30th May 2022

Atta Ur Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology (NUST)

In silico and Genotypic Analysis of Single Nucleotide Polymorphisms (SNPs) of Protein Tyrosine Phosphatase Non-Receptor Type 1 (PTPN1) in Type 2 Diabetes Mellitus (T2DM)

BS Thesis

Submitted By

Manal Khan, Faiqa Hafeez, Heba Khan, and Namra Tariq

BS Applied Biosciences

Supervisor

Prof. Dr. Attya Bhatti

Co-supervisor

Prof. Dr. Peter John

Submission Date

30th May 2022

Atta Ur Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology (NUST)

THESIS ACCEPTNCE CERTIFICATE

Certified that the contents and form of the thesis entitled **"In silico and Genotypic Analysis of Single Nucleotide Polymorphisms (SNPs) of Protein Tyrosine Phosphatase Non-Receptor Type 1 (PTPN1) in Type 2 Diabetes Mellitus (T2DM)"** submitted by UG students: Manal Khan, Heba Khan, Faiqa Hafeez, and Namra Tariq, have been found satisfactory for the requirement of the degree.

Signature: _____

Name of Supervisor: Dr. Attya Bhatti

Date: _____

Signature (HoD): _____

Date: _____

Signature (Principal):

Date: _____

AUTHOR'S DECLARATION

We, Manal Khan, Heba Khan, Faiqa Hafeez, and Namra Tariq, hereby state that our BS FYP Thesis titled "In silico and Genotypic Analysis of Single Nucleotide Polymorphisms (SNPs) of Protein Tyrosine Phosphatase Non-Receptor Type 1 (PTPN1) in Type 2 Diabetes Mellitus (T2DM)" is our own work and has not been submitted previously by us for taking any degree from this university National University of Sciences and Technology (NUST) or anywhere else in the country or world.

At any time, our statement is found to be incorrect even after our graduation, the university has the right to withdraw our BS degree.

Student Names: Manal Khan, Heba Khan, Faiqa Hafeez, and Namra Tariq

Date: _____

CERTIFCATE FOR PLAGIARISM

It is to confirm that this BS thesis titled "In silico and Genotypic Analysis of Single Nucleotide Polymorphisms (SNPs) of Protein Tyrosine Phosphatase Non-Receptor Type 1 (PTPN1) in Type 2 Diabetes Mellitus (T2DM)" by Manal Khan, Heba Khan, Faiqa Hafeez, and Namra Tariq has been examined by me. I undertake the following:

a) The thesis has significant new work and/or knowledge as compared to already published elsewhere. No sentence, table, equation, diagram, paragraph, or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b) The work presented is original and is owned by the authors (i.e., there is no plagiarism). No ideas, processes, results, or words of others have been presented as Authors' own work.
c) There is no fabrication of data or results which have been compiled or analyzed.
d) There is no falsification by manipulating research materials, processes, or changing or omitting of data or results such that the research is not accurately represented in the research record.

Date: _____ **Students' Signatures:**

Date: _____ **Signature and Stamp of Supervisor**

DEDICATION

"This thesis is dedicated to our supervisors at ASAB, and our parents for their persistent guidance and relentless support, enabled us to accomplish this project."

ACKNOWLEDGEMENTS

First and foremost, we will like to praise Allah Almighty, the most beneficent, and the most merciful for his blessings given to us in completing our thesis. We would like to express our deepest gratitude to our supervisor, Prof. Dr. Attya Bhatti and Co-supervisor, Prof. Dr. Peter John for their indispensable guidance and intangible support throughout this project. We would also like to extend our sincere thanks to PhD Scholar, Bibi Sabiha, who profusely provided her knowledge and expertise in conducting our in-silico work. Furthermore, we are grateful to Dr. Jahangir Sarwar Khan, HoD of the surgical ward, Holy Family Hospital (Rawalpindi) for his generous support in executing our project.

We would like to acknowledge our professors and lab in charge at ASAB for the knowledge and skills they have shared with us.

Lastly, we thank our friends and family, especially our parents for their constant care, support, and love for believing in us and keeping our spirits high during this process.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A1C	Glycated Hemoglobin
AMPK	AMP-Activated Protein Kinase
AR	Androgen Receptor
ARMS	Amplification Refractory Mutation System
ATP	Adenosine Triphosphate
DDG	Free Energy Change
DM	Diabetes Mellitus
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic Acid
EGR1	Early Growth Response Protein 1
ER	Endoplasmic Reticulum
FPG	Fasting Plasma Glucose Test
GLUT	Glucose Transporter
GTP	Guanosine Triphosphate
HLA	Human Leukocyte Antigen
HNF1A	Hepatocyte Nuclear Factor 1
IDF	International Diabetes Federation
IL	Interleukins
IR	Insulin Receptor
IRS	Insulin Receptor Substrate
JAK2	Janus Kinase 2 Gene
MAPK	Mitogen Activated Protein Kinase
NF-KB	Nuclear Factor Kappa B
NO	Nitic Oxide
PCR	Polymerase Chain Reaction

PhD-SNP	Predictor of Human Deletarious Single Nucleotide Polymorphism
Phyre-2	Protein Homology/Analogy Recognition Engine V 2.0
PI3K	Phosphatidyl Inositol3- Kinase
PolyPhene	Polymorphism Phenotyping
PROVEAN	Protein Variation Effect Analyzer
PTK	Protein Tyrosine Kinase
PTM	Post Translational Modification
PTP	Protein Tyrosine Phosphatase
PTPN1	Protein Tyrosine Phosphatase Non-Receptor Type 1
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
RPG	Random Plasma Glucose Test
SIFT	Sorting Intolerant from Tolerant
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machine
T2DM	Type 2 Diabetes Mellitus
TBE	Tris Borate EDTA
TE	Tris EDTA
TM	Template Modelling
UTR	Untranslated Region
UV	Ultra Violet
WHO	World Health Organization

Abstract

Type 2 Diabetes Mellitus (T2DM) is one of the leading incommunicable and rapidly growing health afflictions worldwide; a condition challenging to treat and costly to manipulate. A global estimation of diabetes by the International Diabetes Federation (IDF) in 2021 stipulates those 537 million adults lying in the age group of 20-79 are currently living with diabetes. This number is anticipated to ascent to 643 million by 2030 and 783 million by 2045. The risk factors associated with T2DM includes hormonal, environmental, and genetic.

The aim of the study was to replicate intronic SNPs *(rs3787345* and *rs718049)* of PTPN1 gene within the Pakistani population that has been reported earlier in Caucasian and French population via ARMS PCR. In silico analysis was carried out for the prediction of most deleterious missense SNPs in PTPN1 gene.

Results of our in-silico study revealed that the 16 missense SNPs that can dysregulate the function of PTPN1. The genotypic analysis of the two selected intronic SNPs *rs3787345* and *rs718049* manifested the presence of heterozygous genotype (CT) in the sample study of *rs718049* while *rs3787345* requires further investigation. Our results support the existence of functionally significant intronic SNPs in T2DM patients. Further studies having substantial sample size are required to authenticate our findings.

1. Introduction

1.1 Diabetes Mellitus

Type 2 Diabetes Mellitus (T2DM), is one of the leading incommunicable and rapidly growing health afflictions worldwide; a condition challenging to treat and costly to manipulate. A global estimation of diabetes by the International Diabetes Federation (IDF) in 2021 stipulates that those 537 million adults lying in the age group of 20-79 are currently living with diabetes. This number is anticipated to ascent to 643 million by 2030 and 783 million by 2045 (IDF, IDF Diabetes Atlas, 2021). People with diabetes mellitus are at an elevated chance of developing a scope of complications, for example; nephropathy, cardiovascular diseases, visual impairment, and retinal changes that can prompt early death. In addition, it forces crucial clinical and monetary weights. Hereditary responsiveness and ecological impacts appear to be the main reasons for the occurrence of the disease. Anyhow, due to drastic inclination towards heftiness, lack of physical activity, and type 2 diabetes has been lately noticed. It indicates that obesity and physical inactivity might comprise the primary purposes behind the spiraling load of diabetes in the developed world (Asif, 2014).

1.2 Epidemiology

The onset of non-insulin dependent diabetes mellitus is slow; hence its determination of true time in an individual can be a difficult task. Type 2 diabetes reports for approximately 90% of all cases of diabetes, and the number of cases is predicted to increase with time. It was reported by IDF in 2021 that globally 1 in every 10 individuals is living with diabetes and over 3 in 4 adults suffering from it belong to countries with low or middle incomes. Moreover, a total of 6.7 million deaths were reported in the past year; it was also discovered that 547 million adults having impaired glucose tolerance are considered to be at a high risk of developing type 2 diabetes. Next, is the western Pacific where a total of 206 million people is affected by diabetes and the number is expected to rise to 260 million by 2045. In southeast Asia, 90 million adults are reported with diabetes and that number would rise to 152 million by 2045. Other regions such as Middle East, North Africa, Europe, Africa, South and Central America, North America

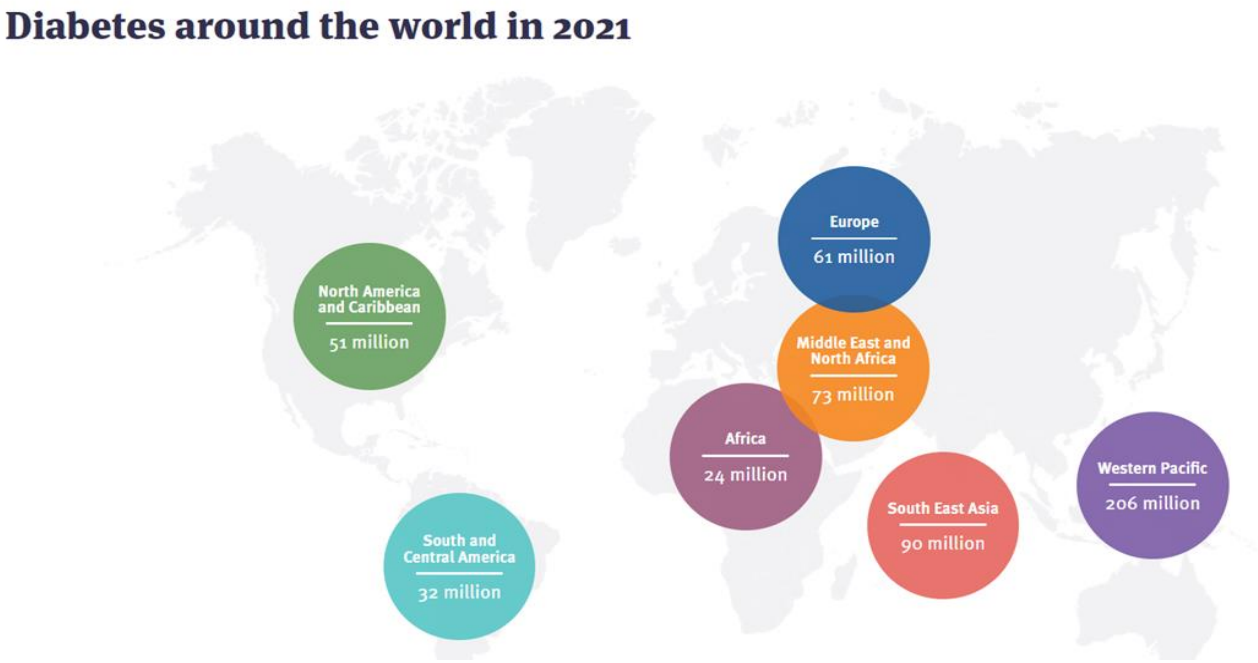


Figure 1: Diabetes around the world in 2021 (IDF, IDF Diabetes Atlas, 2021).

and Caribbean have millions of people living with diabetes and their numbers are also expected to elevate at a rapid rate by 2045 (WHO, 2021).

Pakistan lies in Southeast Asia and according to a report published by IDF in 2021, 33 million adults were living with diabetes resulting in a drastic rise in the number. Therefore, leading it to become a country with the highest number of diabetic patients in the world after, China which has 141 million, and India having 74 million. Furthermore, 11 million Pakistani adults suffer from impaired glucose tolerance placing them at high level of chance of developing type 2 diabetes (IDF, IDF Diabetes Atlas, 2021).

More than 26.9% of the Pakistani population is left untreated with diabetes which leads to them facing lethal consequences such as suffering from kidney failure, heart attack, stroke, blindness etc. which issues an increase in healthcare expenses and poor quality of life (IDF, IDF Diabetes Atlas, 2021).

1.3 Etiology

Type 2 diabetes mellitus (T2DM) represents 90% instances of diabetes. In T2DM, the reaction to insulin is decreased and this is characterized as insulin resistance. In this state insulin is unproductive, and at first it is encountered by an expansion in insulin production in order to keep up with glucose homeostasis. However, after some time there is a decline in insulin production due to the destruction of the beta cells of pancreas, resulting in T2DM. It is

commonly detected in people above the age of 45, but recently it is being observed in children and adolescents because of unhealthy lifestyles and deficient dietary intake (Goyal & Jialal, 2021).

1.4 Pathophysiology

Diabetes Mellitus Type 2 is a globally prevalent metabolic disease that is characterized by the inability of pancreatic beta cells to secrete insulin and the failure of tissues to respond to insulin. Secretion and action of insulin inside the body are the two factors that need to be highly regulated because an imbalance will affect tissues sensitive to insulin like the liver, skeletal muscles, adipose tissues, and beta cells of the pancreas. Disturbance in the functioning of these tissues will cause an abnormal increase in blood glucose level (Galicia-Garcia et al., 2020). Therefore, both beta cell dysfunction and insulin resistance are major contributors to the pathogenesis of diabetes.

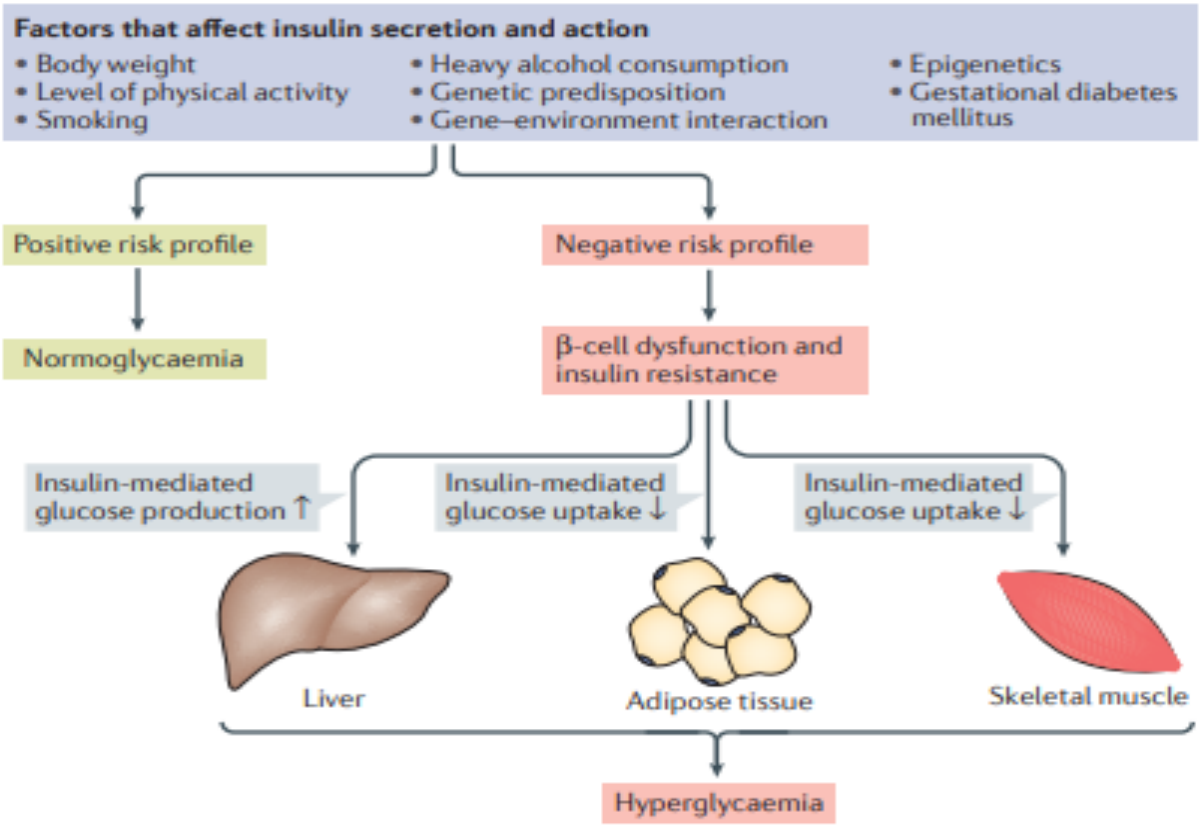


Figure 2: Pathophysiology of Diabetes Type 2 (Zheng et al., 2018a).

Under normal conditions insulin secretion from beta cells of pancreas upregulate glucose intake in skeletal muscles and adipose tissues and downregulate the production of glucose in liver. If there is resistance of insulin in muscles, adipose tissue or liver or beta cells are dysfunctional

hyperglycemia will occur, causing an increase in glucose level of blood. Factors listed above affect the action and secretion of insulin in diabetes mellitus.

1.5 Beta Cells Physiology and their Dysfunction

Beta cells of the pancreas are responsible for maintaining the glucose concentration in the blood via the release of insulin. They produce pre-pro insulin that undergoes conformational change in the endoplasmic reticulum to form proinsulin. This proinsulin is translocated from the endoplasmic reticulum (ER) to the Golgi apparatus where the mature form of insulin is prepared. This insulin is then stored in a granule until beta cells get the signal to release insulin (Cerf, 2013). This release is triggered mainly by high levels of glucose and some other factors like hormones, amino acids, and fatty acids. Glucose, one of the crucial regulators of insulin secretion from beta cells, causes stimulation of gene transcription, insulin synthesis, and insulin release from beta cells (Mahler & Adler, 1999).

Cellular regulation and proper functioning of beta cells depend on cellular regulation and integrity. However, it is evident that a complex interplay of environmental and molecular pathways is responsible for the dysfunction of beta cells. High levels of free fatty acids and glucose cause dysfunction by triggering ER stress. Similarly, lipid and glucose toxicity in obesity affect the functioning of beta cells by inducing metabolic and oxidative stress (Zheng et al., 2018b). Furthermore, persistent hyperglycemia increases proinsulin production, leading to the deposition of misfolded insulin in beta cells that increases the production of reactive protein folding mediated oxygen species. While ER stress disturbs calcium mobilization within cells, that stimulates proapoptotic signals including insulin degradation, recruitment of macrophages, and inflammation of beta cells (Galicia-Garcia et al., 2020).

Highly regulated insulin release is very crucial for various metabolic processes in the body and for that functional integrated beta cells are required. In a morbific state, the process demonstrated above can get activated in beta cells, causing failure in cell-to-cell communication, loss of integrity, and poor regulation of glucagon and insulin release from the pancreas (Cerf, 2013). Any disruption in the insulin gene expression, insulin folding, or precursor of insulin synthesis will cause an insulin release defect that is the primary cause of beta cell dysfunction and the basis of diabetes type 2.

1.6 Insulin Resistance

A decreased or impaired response of insulin sensitive cells to the secreted insulin at a systemic level is insulin resistance (Wondmkun, 2020). There are three main categories of insulin

resistance: impaired response of target cells, diminished secretion of insulin by the pancreas or the presence of an antagonist in plasma. Persistent increases in glucose and saturated free fatty acids are the risk factors for insulin resistance. A high level of fat in circulation either due to diet or lipolysis leads to a competition between glucose and fat for cellular uptake. And increased level of lipid and glucose cause glucolipotoxicity, a situation that can compromise the function and structure of beta cells by inducing oxidative stress (Amiel, 2011). In addition, inflammatory cytokines and hormones that are secreted by adipose tissue can interact with genetic factors responsible for insulin resistance pathway. Initially, beta cells try to compensate IR by secreting more insulin or by increasing their number. But overtime, due to excessive workload, their capacity to secrete insulin decreases and various stress factors induce apoptosis in beta cells, which is characteristic of T2D.

1.7 Usual Insulin Functioning

In response of increased blood glucose level, insulin is produced and secreted by beta cells in pancreas. This insulin then binds to insulin receptor and induces the translocation of GLUTE-4 receptor, normally present in adipose tissue, skeletal muscles and cardiac tissues, to the cell membrane. GLUTE-4 uptakes the glucose from blood stream and this glucose is then either stored in form of glycogen in the tissue or catabolized to form ATP in cells (Reed et al., 2021).

1.8 Diagnosis

Clinically diabetes is usually diagnosed by using glycated hemoglobin (A1c) test or fasting plasma glucose test (FPG). In some instances, healthcare professionals may use random plasma glucose test (RPG).

- Glycated Hemoglobin Test

  This test provides you with an average of blood sugar level over past two to three months. A person is not required to starve before this test. This diagnosis procedure is also known as glycosylated hemoglobin, HbA1C and hemoglobin A1C. Different factors need to be considered while using this method of diagnosis like age and anemia because this test does not give accurate result on anemic people. Test results are indicated in percentages and higher percentage indicates higher level of blood sugar.

- Fasting Plasma Glucose Test

Test is performed after overnight fasting or at least after fasting of 8 hours. Fasting means person will not eat or drink anything except for the sips of water. This test provides average of blood glucose level instantly.

- Random Plasma Glucose Test

  In some cases, this test is performed when person have symptoms of diabetes but do not want to wait hours of fasting for diagnostic test. This procedure does not require any prior fasting and can be performed at any time.

- Gestational Glucose Diagnosis

  Following tests are performed on pregnant women to know how their body deals with the glucose.

- Oral glucose tolerance test

  Oral glucose tolerance test/ Glucose tolerance test is performed after overnight fasting. Firstly, clinician draws your blood and then you drink sugar containing liquid. After that your blood is drawn and examined for sugar level after every 1 to 2 hours.

  This test can also be performed on people who are not pregnant. But it is not preferable for diabetes diagnosis because it is expensive and less convenient method as compared to other diagnosis procedures.

Result*	A1C Test	Fasting Blood Sugar Test	Glucose Tolerance Test	Random Blood Sugar Test
Diabetes	6.5% or above	126 mg/dL or above	200 mg/dL or above	200 mg/dL or above
Prediabetes	5.7 – 6.4%	100 – 125 mg/dL	140 – 199 mg/dL	N/A
Normal	Below 5.7%	99 mg/dL or below	140 mg/dL or below	N/A

Figure 2: Diagnostic Tests and their results for Diabetes (CDC, Diabetes Tests, 2021).

- Glucose Screening Test

Glucose challenge test is another name for this test. In this diagnosis healthcare professional draws your blood 1 hour after your intake of a liquid containing glucose. Fasting is not required for this test. If your blood sugar level is too high (135-140mg/dl) you need to have glucose tolerance test for confirmation of diabetes. (NIDDK, Diabetes Tests and Diagnosis, 2016)

1.9 Treatment Options

Improved lifestyle plays a key role in management of blood sugar level in the body. A person should prefer healthy eating options like diet having more of high fiber items like fruits, whole grains and less of sweets and starchy vegetables. Controlled Portion size with low calories is preferable. Furthermore, one can consult a dietitian to have a balanced healthy diet plan according to the preferences. An adult should indulge in physical activity for at least 30 minutes in most days of the week. It helps in losing weight that results in better control of cholesterol, blood sugar level and blood pressure (Mahler & Adler, 1999).

Medicines along with healthy lifestyle further ameliorates the situation of diabetes. The most commonly used medication is metformin, it decreases the glucose production in liver and peripheral insulin resistance. This drug also helps in weight loss. In some cases, patient is required to take combination of diabetic drugs because different drugs act differently to maintain blood sugar level. DPP-inhibitor, GLP-1 inhibitor agonist, Sulfonylureas and glinides are some other diabetic medications that are available in market (Reed et al., 2021).

Some diabetic patients need insulin therapy, different types of insulin depending on their bioavailability and half-life are available, for instance, long-acting insulin works for whole day or overnight to maintain glucose level in body. Physician decides the type of insulin suitable for the patient after monitoring (MayoClinic, 2021). For situations in which medication and lifestyle is not working, some fewer common treatments are also available. Weight loss surgery or bariatric surgery helps to lose a large amount of fat from the body resulting in improvement of blood sugar level. In some cases, patient do not need medication after this surgery (NIDDK, Insulin, Medicines, & Other Diabetes Treatments, 2022).

1.10 Risk Factors associated with Type 2 Diabetes Mellitus (T2DM)

The chances of getting T2DM by an individual depends on combination of multiple reasons such as genetic makeup and lifestyle. It is impossible to manipulate the factors such as your genes, family history, ethnicity, or age, however what you can do to prevent the onset of T2DM

is to take precautionary measures by involving physical activity, maintaining healthy weight, diet, and lifestyle.

1. **Endocrine connection with T2DM**

   The body is responsible for glucose production from the food we eat. The hormone, insulin is produced by pancreas that enables the uptake of glucose from bloodstream into the cells to provide energy. However, in T2DM the body is unable to efficiently utilize the insulin or at times the insulin is not being produced in sufficient quantity which leads to excess of glucose in blood (Anhalt, et al., 2022).

2. **Environmental connection with T2DM**

   - Diet: Variety of diet exist and are consumed in different parts of the world and as result different food groups are associated with the incidence of T2DM. However, normally plant food group tends to have lower risk of developing T2DM in comparison to meat, high density energy food is less protective than low density energy food, and sugary food items are responsible for obesity, hence increasing the risk of developing diabetes.
   - Physical activity: diabetes is reported to be in reverse proportion with leisure time physical activities as physically fit means your weight and glucose homeostasis is balanced and your cells are well sensitive to insulin but if you are physically inactive, chances of developing diabetes are increased.
   - Coffee, tea, alcohol, and smoking: intake of coffee and tea are reported to lower the risk of developing diabetes as they aid in improving the glycemic control in the human body. The effect of alcohol seems to be dependent on its dose, and it's reported that alcohols moderate consumption has chances of reducing the risk of T2DM by an approximate of 20% specifically in females and not in Asian populations. Moreover, the health impact of smoking is hazardous and that is why whether it is active or passive smoking, both tend to be at inclined risk of developing diabetes (Kolb & Martin, 2017).

3. **Genetic connection with T2DM**

   It has been observed that type 2 diabetes does not appear to have a direct inheritance pattern, but mostly the diabetic patients have at least one of their family members for instance parents or siblings to be suffering from diabetes. The chances of an individual to develop diabetes is influenced by an increase in the number of the family members

being diabetic. This could be to the shared genetic factors along with the lifestyle that is common among them (MedlinePlus, Type 2 Diabetes, 2017).

- TCF7L2- insulin secretion and production of glucose is affected by this gene
- PPARG- at position 12 of the chromosome there is a change from proline to arginine which results in a 20% increase in likelihood of getting T2DM
- KCNJ11- The gene encodes for ATP sensitive potassium channel that is involved in insulin regulation by beta cells.

Sr. No.	Locus	Chromosome	Risk Allele Frequency
1	NOTCH2	1	0.11
2	PROX1	1	0.5
3	IRS1	2	0.61
4	THADA	2	0.92
5	RBMS1/ITGB6	2	0.57
6	BCL11A	2	0.46
7	GCKR	2	0.62
8	IGF2BP2	3	0.29
9	PPARG	3	0.92
10	ADCY5	3	0.78
11	ADAMTS9	4	0.81
12	WFS1	4	0.27
13	ZBED3	5	0.26
14	CDKAL1	6	0.31
15	JAZF1	7	0.52
16	GCK	7	0.2
17	KLF14	7	0.55
18	DGKB/TMEM195	7	0.47
19	SLC30A8	8	0.75
20	TP31NP1	8	0.48
21	CDKN2A/B	9	0.79
22	TLE4	9	0.93
23	TCF7L2	10	0.25

24	HHDX	10	0.56
25	CDC123/CAMK1D	10	0.23
26	KCNQ1	11	0.61
27	KCNJ11/ABCC8	11	0.50
28	CENTD2	11	0.88
29	MTNR1B	11	0.30.
30	KCNQ1	11	0.52
31	HMGA2	12	0.10
32	TSPAN8/LGR5	12	0.23
33	OASL/HNF1A	12	0.85
34	PRC1	15	0.22
35	ZFAND6	15	0.56
36	FTO	16	0.56
37	HNF1B	17	0.43
38	DUSP9	X	0.12

Table 1: Thirty-eight genetic variants associated with Type 2 Diabetes at genome-wide significance (Ali, 2013).

1.11 Impact of study

The present study is based on in silico and genotypic analysis of nonsynonymous SNPs of the PTPN1 gene that will lead to personalized medicine which will allow for earlier diagnosis, risk assessments, and better treatments such as therapies for the patients. Personalized medicine and therapies outlined on the basis of each individual patient's genetic make-up will lead to better healthcare services and lower costs.

1.12 Objectives of the study

The objectives of the study are as follows

- Elucidate the possible role of SNPs of PTPN1 through in silico analysis.
- Determine the genotypic analysis of SNPs *rs3787345* and *rs718049* as a causative agent in disease etiology.

2. Literature Review

2.1 Diabetes

Type 2 Diabetes mellitus is one of the most common metabolic diseases in the world. Globally its prevalence is anticipated to ascent to 643 million by 2030. Diabetes is associated with a wide range of life-threatening health implications. Major features of diabetes include insulin resistance, hyperglycemia and dysfunction of beta cells (Teimouri, et al., 2022).

Insulin resistance is usually caused by the impairment of insulin receptor (IR). Normally, binding of insulin-to-insulin receptor which is a type of receptor tyrosine kinases, causes phosphorylation of its tyrosine residues. This process then phosphorylates several downstream targets for the activation of two signaling Cascades, phosphatidylinositol 3-kinase (PI3K) pathway and mitogen-activated protein kinase (MAPK), both pathways are responsible for mediating the function of insulin. Thus, phosphorylation plays crucial role in physiological functioning of insulin in target body tissues. Several phosphatases like inositol phosphatase and protein tyrosine phosphatases (PTPs) plays role in reverse phosphorylation of insulin receptors. As protein tyrosine phosphatases have importance in insulin regulation, they have part in insulin resistance and diabetes mellitus.

2.2 PTP1B Location and Structure

Protein Tyrosine Phosphatase 1B (PTP1B) is a key member of the protein tyrosine phosphatases (PTPs) which are a superfamily of enzymes mainly involved in inhibiting the function of protein tyrosine kinases (PTKs). PTP1B is an intracellular cytosolic tyrosine-specific phosphatase isolated from the human placenta by Tonks in 1988 (Abdelsalam, et al., 2019). The protein tyrosine phosphatase non-receptor type 1 (PTPN1) gene encodes this enzyme located on chromosome 20q13.13, a region that has been linked with insulin resistance, obesity and diabetes. The genomic coordinates of PTPN1 gene are chr20:50,510,321-50,585,241 having a genome size of 74,921 bases.

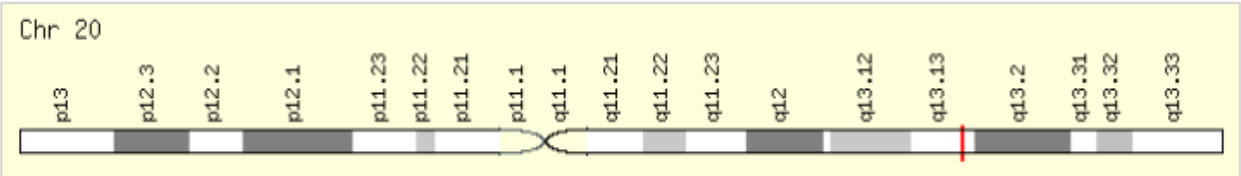


Figure 3: Location of PTPN1 on chromosome 20 (GeneCards, 2022).

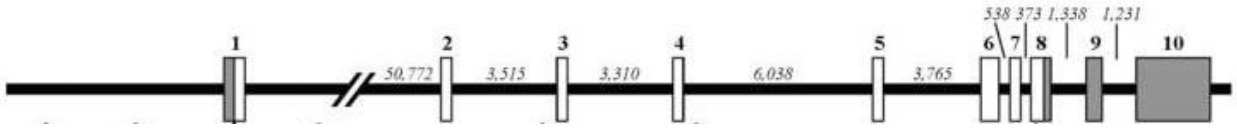


Figure 4: Genomic organization of PTPN1 gene on chromosome 20. Exons are represented by boxes, white region showing translated regions. Introns are represented by regions in between the boxes with size indicated in base pairs (Olivier et al., 2004).

In humans, two transcripts of PTP1B are identified categorized as PTPN-201 comprising of 10 exons and PTPN1-202 comprising of 9 exons.

Transcript Name	Transcript ID	Length (bp)	Protein Length (aa)	UniProt Match
PTPN1-201	ENST00000371621.5	3979	435aa	P18031
PTPN1-202	ENST00000541713.5	3469	362aa	B4DSN5

Table 2: Protein coding transcripts and respective PTP1B isoforms.

PTPN1 gene is translated into 435 amino-acid protein with a molecular weight of 50kDa comprising of a catalytic N-terminal domain crucial for PTP1B catalytic activity, two proline rich domains involved in protein-protein interactions, and a C- terminal hydrophobic region associated with the regulatory function (Zou, 2016). PTP1B is targeted to the cytoplasmic face

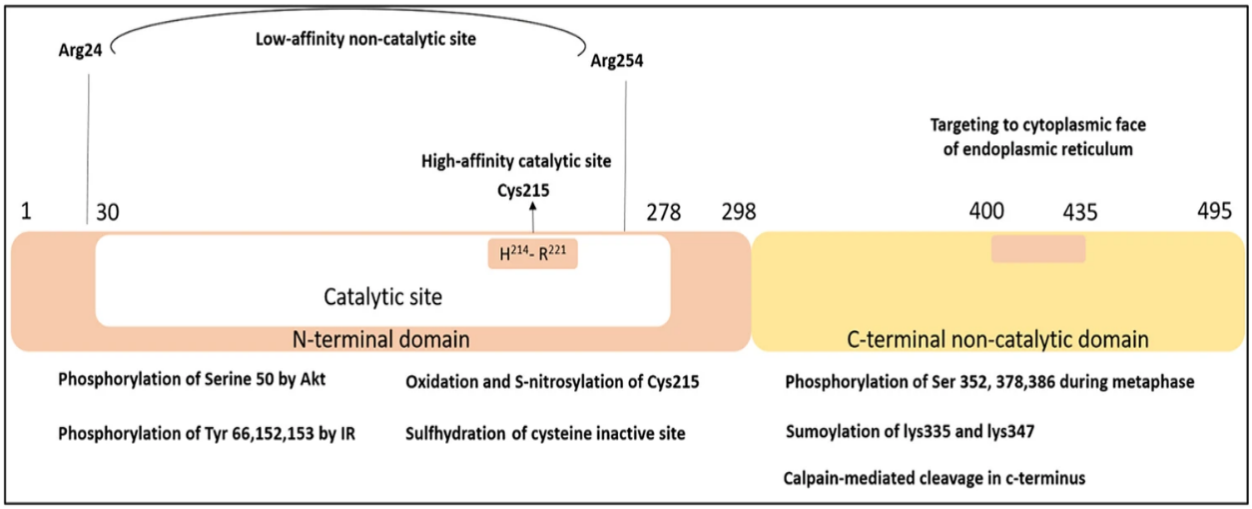


Figure 5: Structural representation of PTPB1 (Teimouri et al., 2022).

of the endoplasmic reticulum by the 35 amino acid C-terminal tail where it interacts with various substrates including nuclear, cytosolic, mitochondrial, adherent junction and plasma

protein bound proteins. The catalytic domain comprising of Cys215 and Arg221 along with the non-catalytic domain regulates the interaction of PTP1B with its substrates (Villamar-Cruz, et al., 2021). PTP1B controls various signaling processes including differentiation, apoptosis, cell growth, and movement by dephosphorylating the target molecules.

2.3 Regulation of PTP1B

PTP1B expression and function is regulated at transcriptional, post-transcriptional, and post-translational levels. At the transcriptional level, various transcription factors like SP1, early growth response protein 1(EGR1), Y box-binding protein 1 (YBX1), nuclear factor Kappa B (NF-KB) and androgen receptor (AR) bind to the promoter region regulating the PTP1B promoter activity (Villamar-Cruz, et al., 2021).

At the post transcriptional level, variety of miRNAs such as, miR193a-3p, miR-146-a, miR122, miR338-3p, miR-206 act as regulators of PTP1B through interaction with its 3'-UTR regions. In T2DM, down regulation of miR146-a causes enhanced PTP1B expression leading to impaired insulin signaling (Abdelsalam et al., 2019).

Post translational modifications such as oxidation, phosphorylation, nitro-sylation, proteolytic cleavage and sumoylation regulate PTP1B activity. Phosphorylation of serine and tyrosine plays a significant and effective role in regulation of PTP1B activity in the insulin signaling pathway. For instance, tyrosine phosphorylation of PTP1B leads to the effective interaction with the activated insulin receptor. Moreover, serine phosphorylation of PTP1B by protein kinase B (Akt) which is a downstream effector in the insulin signaling pathway negatively regulates PTP1B activity and corroborates insulin signaling (Abdelsalam et al., 2019).

2.4 Role of PTP1B in Insulin Signaling Mechanism of Type 2 Diabetes Mellitus

PTP1B acts as a negative regulator in the insulin transduction pathway, which is crucial to glucose and energy metabolism. Increased expression of PTP1B gene has also been observed in liver, skeletal muscles and adipose tissues of diabetic and obese patients. The locus of PTP1B is positioned in the region which is related to obesity, insulin resistance and T2DM and variations in the PTPN1 gene leading to single nucleotide polymorphisms have also been associated with insulin resistance, diabetes and obesity. Studies have confirmed increased insulin sensitivity, low blood glucose and insulin levels in PTP1B knockout mice as compared to wild type mice establishing PTP1B as a significant regulator and a novel therapeutic agent in insulin resistance.

Impaired insulin signaling and defects in insulin secretion by B-cells are the major factors in the pathogenesis of type 2 DM. Insulin resistance can occur in a variety of tissues including liver, muscle, gastrointestinal tract, adipose tissue, pancreatic beta cells and brain and vasculature. Impaired insulin signaling, glycogen synthesis or glucose transport leads to the development of insulin resistance in tissues. In the intricate process of insulin signaling, insulin binds with insulin receptor (IR) resulting in the auto phosphorylation of tyrosine residues on IR causing the activation and binding of insulin receptor substrate proteins. The recruitment of IRS proteins in turn activates phosphatidylinositol 3-kinase (PI3K) and protein kinase B that are intracellular signaling molecules stimulating the GLUT -4 transport to the membrane surface facilitating glucose uptake *(Rocha, et al., 2021).*

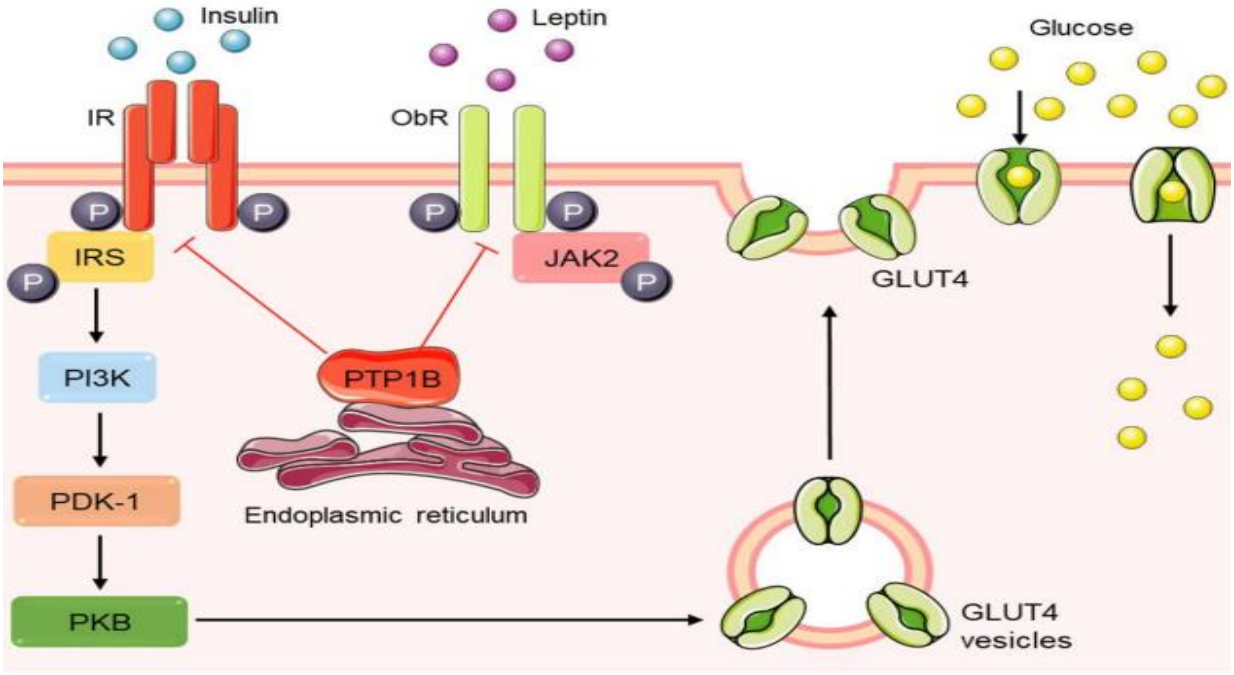


Figure 6: Schematic representation of the down regulation of the insulin signaling pathway by PTPB1(Rocha, et al., 2021).

PTP1B is associated with the down regulation of the insulin signaling cascade by dephosphorylating the tyrosine residues on the insulin receptor (IR) at positions 972, 1162 and 1163. IR/IRS-1 signaling cascade is impaired, resulting in the diminished activity of phosphatidylinositol 3-kinase (PI3K) leading to reduced translocation of glucose transporter type 4 GLUT-4 to the cell surface. Increased expression of PTP1B induces insulin resistance causing Type 2 diabetes. Induction of PTP1B in type 2 diabetes alleviates glucose-stimulated insulin secretion. However, its attenuation results in improved B-cell functioning and pancreatic insulin secretion (Abdelsalam, et al., 2019).

Furthermore, PTP1B negatively regulates the leptin signaling pathway and is considered as a significant target in the treatment of obesity. Both inactivity and obesity are linked to insulin resistance. Genetic predisposition and obesity cause the development of insulin resistance and impaired beta cell functioning leading to decreased insulin secretion progressing to development of type 2 diabetes mellitus (Rocha, et al., 2021).

Leptin hormone is involved in the maintenance of fatty acid homeostasis and promotes fat tolerance mainly in adipose tissue. Leptin functioning is impaired in obesity resulting in lipotoxicity in tissues including muscles and liver. Leptin signaling pathway involves the binding of leptin to ObR (leptin receptor), that activates JAK2 by phosphorylating it resulting in the subsequent activation of the downstream process. Dephosphorylation of ObR associated JAK2 by PTP1B, leads to the deactivation and impairment of leptin signaling pathway. Thus, suppression of leptin can lead to improved leptin secretion, acting as a potential remedial target for type 2 diabetes and obesity (Abdelsalam, et al., 2019).

2.5 Role of PTP1B in Cardiovascular Complications

PTP1B is associated with endothelial functioning as it is involved in regulation of insulin signaling pathway in smooth muscle cells and endothelial cells. Deficiency of PTP1B in endothelium results in restored production of nitric oxide (NO), reversed overproduction of reactive oxygen species, reduced endothelial impairment and cardiac failure. Nitric oxide is a significant mediator in the endothelial functioning as reduction in NO production causes endothelial dysfunction initializing diabetic cardiac complications. In Diabetic patients, phosphorylated levels of Akt in arteries are half as compared to controls, demonstrating the impaired insulin signaling in diabetic patients. Deletion of endothelial PTP1B results in enhanced signaling of vascular endothelial growth factor and angiogenesis, decreased oxidative stress and cardiac hypoxia under hypertrophic conditions leading to ameliorated survival of endothelium and reduction in cardiac heart failure. Cardiac dysfunction and remodeling improved in mice with cardiac heart failure due to reduced cardiac fibrosis and myocytes hypertrophy along with enhanced insulin sensitivity restoring endothelial functioning. PTP1B deficiency is associated with improvement in cardiovascular inflammation and functioning after myocardial infarction and inhibition of cardiomyopathy induced by obesity (Abdelsalam, et al., 2019).

2.6 Role of PTP1B in Cancer

PTP1B demonstrates tissue specific tumor suppressor or promoter role in cancer. In Neu (rat ortholog to human epidermal growth factor receptor 2) driven breast cancer mouse model, PTP1B acts as an oncogene. P62Dok and Src are the two significant mediators involved in the oncogenic signaling of PTP1B. Dephosphorylation of P62Dok by PTP1B results in inhibition of p120RasGAP causing increased levels of GTP-bound Ras.

In pancreatic ductal adenocarcinoma PTP1B is highly expressed and is associated with metastasis and tumor staging, modulating cell growth by regulating PKM2/AMPK/Mtoc1 signaling pathway.

In lymphomas, glioblastomas, melanoma, esophageal cancer and non-small cell lung cancer PTP1B acts as a tumor suppressor. Dephosphorylation by PTP1B of caveolin leads to down regulation of Rab5/Rac1 pathway which is crucial for cell migration. PTP1B inhibition in glioblastomas leads to suppression of adhesion, migration and invasion by IL-13 due to the inhibition of Src dephosphorylation and subsequent inactivation of the pathway (Villamar-Cruz, et al., 2021).

2.7 Role of PTP1B in Diabetic Retinopathy

Diabetic retinopathy (DR) is one of the most significant complications of diabetes and is associated with a variety of cells and disease-causing mechanisms. Mutations in the rhodopsin gene and its photobleaching defects results in elevated levels of PTP1B causing decreased auto phosphorylation of IR in the retina that is associated with retinal degeneration. IR signaling is a crucial pathway in the functioning of retina and defects in this pathway leads to DR. Increased levels of PTP1B impedes IR signaling in the rods in dark therefore, inhibitors of PTP1B can act as significant targets in the treatment of DR. Growth factor receptor bound protein 14 (Grb14) on capturing photon by rhodopsin inhibits the phosphatase activity of PTP1B and act as a therapeutic agent for retinal degeneration (Abdelsalam, et al., 2019).

2.8 PTP1B Inhibitors as Therapeutic Agents

Research efforts are disposed at developing PTP1B inhibitors that interact with both the catalytic as well as the non-catalytic phosphotyrosine binding sites of PTP1B. Efficient inhibitors of PTP1B bind to the high affinity catalytic site of PTP1B that is highly charged. PTP1B inhibitors are under clinical trials including Ertiprotafib, Trodusquemine and ISIS-PTP1BRx as potential therapeutic targets for T2DM, weight loss, breast cancer and obesity (Abdelsalam, et al., 2019).

2.9 Importance of Single Nucleotide Polymorphisms (SNPs) in disease diagnosis

Single nucleotide polymorphisms (SNPs), also pronounced as "snips", are a category of polymorphisms that involve a single base pair variation. Among the people, SNPs are the most frequent type of genetic variation. To be classified as an SNP, the frequency of the variation needs to occur at least in 1% or more of the population (MedlinePlus, What are Single Nucleotide Polymorphisms (SNPs)?, 2022).

The difference between a mutation and an SNP is that a SNP represents a change in a single nucleotide in the DNA genome whereas a mutation has any type of changes that can affect single to many nucleotides in the DNA genomic sequence (Samanthi, 2017).

2.10 Synonymous and Nonsynonymous SNPs

There are two categories of SNPs: Synonymous and Non-synonymous SNPs. Synonymous SNPs are SNPs that do not cause a change in the amino acid sequence. Nonsynonymous SNPs are SNPs that cause an alteration to an amino acid present in the protein coding sequence (Chu & Wei, 2019) (Hunt, et al., 2009).

2.11 Intronic SNPs

Intronic SNPs have the ability to essential functional elements such as the intron splice enhancers and silencers, trans-splicing elements (Nair, et al., 2021), and other controlling elements such as the ability to cause dysregulation of the mRNA splicing so that biological activities can be regulated (Lin, et al., 2019). Intronic SNPs may lead to increase in susceptibility to disease (Nair, et al., 2021).

2.12 Missense SNPs

Missense SNPs have an altered single nucleotide present in the amino acid sequence that may lead to structural and functional disturbances (Lin, et al., 2019) such as the folding, ligand binding, catalysis, allosteric regulation, localization, post-translational modification, or aggregation and half-life (Pal & Moult, 2015). Missense SNPs occur in the coding region of the gene (Marcolino, et al., 2016). These structural and functional disturbances in the coding regions may lead to disease outcome.

Some of these SNPs have been associated with breast cancer, cardiovascular diseases, esophageal squamous cell carcinoma, diabetic retinopathy, obesity, polycystic ovary syndrome and ovarian response, Turner syndrome, and colorectal cancer (Emadi, et al., 2020) (Venkata Subbiah, et al., 2020).

2.12 Selected Intronic SNPs

The selected SNPs for the present study chosen have been found in the French, and Caucasian populations.

- *SNP rs3787345*

  SNP *rs3787345* is located o intron 4. This SNP has been reportedly found in the Caucasian and the French population (Bento, et al., 2004) (Cheyssac, et al., 2006). SNP *rs3787345* has also been associated with breast cancer (Huang, et al., 2019) and cardiovascular diseases (Farbstein Bsc & Levy MD, 2010).

- *SNP rs718049*

  SNP *rs718049* is located between intron 8. This SNP has been reported in the Caucasian population (Bento, et al., 2004*).* SNP *rs718049* has been associated with breast cancer (Huang, et al., 2019) and cardiovascular diseases (Farbstein Bsc & Levy MD, 2010).

3. Materials and Methods

3.1 Selection of Intronic SNPs

The following SNPs were selected for this study and they all are intronic SNPs.

- *SNP rs3787345:* Its position is 50568886 on the Genome Reference Consortium Human Build 38 (GRCh38) and is located between introns 4 and 5.
- *SNP rs718049:* Its position is 50580062 on the GRCh38 and is located between introns 8 and 9.

3.2 In silico Analysis of Missense SNPs

3.2.1 Retrieval of Missense SNPs

For further investigation, the missense SNPs of PTPN1 gene were also analyzed. The missense SNPs data (position number, protein sequence) was collected from NCBI dbSNP (https://ncbi.nlm.nih.gov/snp/). A total of 261 SNPs were retrieved.

3.2.2 Identification of the most damaging SNPs

The missense data were submitted in the following two prediction tools for predicting the functional effects: Ensembl BioMart (https://asia.ensembl.org/info/data/biomart/index.html), and Protein Variation Effect Analyzer (PROVEAN) (http://provean.jcvi.org/index.php). Ensembl BioMart produces results of protein coding regions when filters and attributes are selected. PROVEAN is a software prediction tool that informs if an amino acid substitution or indel is affecting the biological properties of a protein (Sabiha et al., 2021).

The SNPs were filtered out through the following criteria in Ensembl BioMart.

<table>
<tr><td>Database</td><td colspan="3">Ensembl Variation 106</td></tr>
<tr><td>Dataset</td><td colspan="3">Human Short Variants (SNPs and indels excluding flagged variants) (GRCh38.p13)</td></tr>
<tr><td rowspan="2">Filters</td><td colspan="2">General Variant Filters</td><td>Variant Source: dbSNP
Clinical Significance</td></tr>
<tr><td colspan="2">Gene Associated Variant Filters</td><td>Gene Stable ID: ENSG00000196396
Variant Consequence: Missense Variant</td></tr>
<tr><td>Attributes</td><td>Variant Associated Information</td><td>Variant Information</td><td>Variant Name</td></tr>
</table>

			Chromosome/scaffold name Chromosome/scaffold position start (bp) Chromosome/scaffold position end (bp) Variant Alleles
	Gene Associated Information	Gene Attributes	Gene Name Transcript Stable ID Transcript Strand Variant Consequence Consequence Specific Allele Protein Allele Variant Start in cDNA (bp) Variant End in cDNA (bp) Variant Start in translation (aa) Variant End in translation (aa) Polyphen Prediction Polyphen Score SIFT Prediction SIFT Score

Table 3: Criteria set for Ensembl BioMart.

3.2.3 Identification of Structural and Functional effects of missense SNPs

For categorizing the SNPs into disease causing SNPs or neutral SNPs, PMut (http://mmb.irbbarcelona.org/PMut/) and MutPred2 (http://mutpred.mutdb.org/software) were utilized. PMut analyzes the SNP effect on the protein's functional abilities, and MutPred2 analyzes and places the amino acid substitution in either the pathogenic or the benign categories. The protein sequence (FASTA format) was submitted to each of these web tools.

3.2.4 Effect of missense SNPs on conserved amino acids

Consurf (https://consurf.tau.ac.il/) is an online software that utilizes protein sequences by making estimations on the evolutionary amino acids. The analysis is based on the phylogenetic relations with homologous sequences. The homologous sequences were submitted so that the estimation of conservation degree of the amino acids could be made. Only those SNPs were selected for further analysis that had highly conserved amino acid residues.

3.2.5 Predicting the effect of missense SNPs on Protein Stability

For the functioning of any protein, its stability is important. Single amino acid mutations can disturb the stability of a protein. I-mutant 3.0 (http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) is a support vector machine (SVM) based tool that automatically predicts the stability of a protein upon single amino acid mutations whether the mutation increases (+) or decreases (-) the stability of a protein. The ΔΔG value and direction are predicted which the mutation is either a stabilizing mutation (increase, >0) or a destabilizing mutation (decrease, <0). The reliability index (RI) of the prediction is also generated along with the results. The criteria were set at 25°C and the pH at 7.0.

3.2.6 Prediction of the effect of missense SNPs on Post Translational Modifications (PTMs) Sites

Post translational modification (PTMs) sites have an important role in protein function regulation. Methylation site at arginine; phosphorylation sites at tyrosine, threonine, and serine; glycosylation sites, ubiquitination sites, SUMOylation sites, palmitoylation sites and more sites were predicted by MusiteDeep (https://www.musite.net/). The amino acid residues that had a score > 0.5 were considered to be phosphorylated, methylated, ubiquitylated, SUMOylated, palmitoylated, and glycosylated.

3.2.7 Protein Modelling

The missense SNPs that were predicted as highly conserved, decreasing the protein stability and containing PTMs are selected so that they undergo 3D modelling. Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) is a homology modelling tool that creates the 3D structural model of both wild and mutant types. Phyre2 works on profile-profile match alignment. The 3D models that are constructed through Phyre2 are then validated by ERRAT (https://www.doe-mbi.ucla.edu/errat/) and VERIFY3D. PROCHECK is also used to confirm the stereo-chemical quality of the protein structures. TM-align tool is also utilized to find any differences between the wild and mutant 3D structures. This algorithm calculates

the root mean square deviation (RMSD); the greater the RMSD value, the greater the deviation between the wild and mutant protein structures The template modelling (TM) score is also generated along with this where 0 and 1 are given.

3.2.8 Retrieval of Intronic SNPs data

The two intronic SNPs (SNPs) that were chosen for this study, were analyzed first through computational methods to check their impact on the PTPN1 gene. The data for each of the intronic SNPs was gathered from the NCBI dbSNP (https://ncbi.nlm.nih.gov/snp/) database.

Next, the PTPN1 gene was searched on Ensembl (http://asia.ensembl.org/index.html) database where it showed the analysis of the exon and intron numbers. The position of the SNPs was crossed checked to observe in which intronic region do each of the two SNPs lie.

3.2.9 Prediction of deleterious effect of SNPs

The SNPs were then placed in the regSNP-itnron (https://regsnps-intron.ccbb.iupui.edu/) software tool. This tool predicts whether intronic SNPs are disease causing or not based on the SNP's genomic and protein attributes. The SNP position, change in allele, and the chromosome number where the SNP's located were inputted into the software and analyzed.

The SNPs were analyzed by the NETGENE2 (https://services.healthtech.dtu.dk/service.php?NetGene2-2.42), SNP2TFBS (https://ccg.epfl.ch/snp2tfbs/), ConSite (http://consite.genereg.net/), and RegulomeDB (https://regulomedb.org/regulome-search/).

NETGENE2 produces results about the splice sites in human DNA. It checks if the SNP lies in the splice site or not. SNP2TFBS is an interface that studies variants that affect the transcription factor binding site occurring on the human genome. ConSite software allows the exploration of the transcription factor binding sites. RegulomeDB allows the identification of DNA features and the regulatory components such as transcription factors, promoter, enhancers, motifs, and binding proteins present in the non-coding regions by submitting the dbSNP IDs.

3.3 Genotypic Analysis

3.3.1 Patient Consent and Sample Collection

To begin the study, blood samples are to be acquired from diabetic patients. The hospital ethical committee and the ASAB review board were informed about the study and garnered approval. Next, the patients were enlightened about the study before blood samples were taken. The blood samples were collected in purple EDTA tubes and were carefully labelled according to

the patient ID, gender, age, sample type and if any special handling is required with any of the blood samples.

3.3.2 Sample Transport and Storage

The blood samples collected were transported carefully in an ice box and were stored at 4°C in a refrigerator.

3.3.3 Genomic DNA Extraction

Before the procedure was started, the following reagents were gathered and utilized for the genome DNA extraction method.

Patient's Blood	Ethanol (100% and 70%)
Proteinase K	Salt Solution (10M Ammonium acetate, 8M LiCl, 5M NaCl/3M Sodium acetate)
TE Buffer (10X, pH 7.8)	Solution A (0.32M sucrose, Tris (pH 7.5), 5mM Magnesium chloride, 1% V/V of Triton X-100)
Solution B (Tris (pH 7.5), 400mM Sodium chloride, and 2mM Ethylene diamine tetra acetic acid (EDTA))	Solution C (Phenol)
Solution D (24 volumes of Chloroform, and 1 volume of Isoamyl-Alcohol)	

Table 4: Reagents required for DNA Extraction.

The function of Solution A is to break the cell's cell membrane so that DNA can be released. Solution B's function is to separate proteins from the DNA and allow DNA to precipitate. The function of Solution C is to separate the aqueous and the organic phases clearly in the solution where the aqueous layer contains the DNA and the organic layer contains cellular components such as lipids. Solution D's function is to stabilize the protein coagulation and to reduce the foaming in the solution so that pure DNA can be isolated successfully.

3.3.4 Procedure of Genomic DNA Extraction

The DNA extraction from the blood samples was done through the phenol-chloroform method and it takes two days.

Day 1 Procedure

1. 750µL of the patient's blood sample is taken and pipetted into an Eppendorf tube.
2. Next, 750µL of Solution A is added into the Eppendorf tube and the tube is inverted about 4-6 times.
3. The tube is kept at room temperature for 10 minutes.
4. The Eppendorf tube is then placed in the centrifuge machine and centrifuged at 13,000rpm for 10 minutes.
5. The supernatant is discarded after centrifugation and resuspended in 400µL of Solution B and pellet is dissolved by repeated tapping.
6. The tube is placed back in the centrifugation machine at centrifuged at 13,000 rpm for 10 minutes.
7. The supernatant is discarded again and 400µL of Solution B is added along with 5µL of proteinase K.
8. The Eppendorf tube is then placed in the incubator overnight at 37°C.

Day 2 Procedure

1. The Eppendorf tube is taken out the next day from the incubator and 250µL of Solution C and 250µL of Solution D are added in a separate Eppendorf tube.
2. The tube containing both Solutions C and D are then poured into the overnight incubated Eppendorf tube.
3. The tube is then placed in the centrifugation machine and centrifuged at 13,000 rpm for 10 minutes.

4. The aqueous layer contains the DNA and is transferred into a separate tube while the remaining layer is discarded.
5. The aqueous layer needs to be submerged in 55µL of 3M sodium acetate and 500µL of chilled isopropanol.
6. The tube is then inverted several times so that DNA can be precipitated.
7. The supernatant is discarded after centrifugation and the DN pellet is resuspended in 100% chilled ethanol.
8. The tube is placed back inside the centrifugation machine and centrifuged at 13,000rpm for 8 minutes and is allowed to be air dried so that ethanol can be evaporated.
9. Once dried, the DNA is submerged in 200µL of TE buffer or PCR water.

3.3.5 Qualitative Analysis

- The extracted DNA has to undergo qualitative analysis and for this, agarose gel electrophoresis procedure was utilized.
- 1% Agarose gel is prepared in 1X TAE buffer (0.5g of Agarose is mixed in 50µL of 1X TAE buffer)
- 5µL of DNA sample is loaded along with 1µL of 6X loading dye.
- The sample was run at 110 volts for 30 minutes and visualization of the gel was done under a UV transilluminator.
- Under the UV light, the presence of DNA was confirmed and the DNA sample was then processed further in PCR.

3.3.6 Detection of PTPN1 Gene SNPs

The goal of this study is to check the presence of single nucleotide polymorphisms (SNPs) in the gene given below.

- PTPN1 gene

Amplification-refractory mutation system (ARMS) was used to detect SNP mutation using sequence-specific SNP primers in PCR, allowing the amplification of sample DNA only if the target allele is present in it.

So, ARMS PCR was done for PTPN1 gene

Sr. No.	SNP Primer Name	Primer Sequence	Annealing Temp. (°C)	Product Size
1	rs718049WF	ATCCTTGGGGAACAGGGC	62.2	336
2	MF	ATCCTTGGGGAACAGGGT	59.9	336
3	R	ATGGTTGTTAACCTGCAAGG	58.4	336
4	rs3787345WF	GAGTGTGGCTACCTCTGC	62.2	234
5	MF	GAGTGTGGCTACCTCTGT	59.9	234
6	R	AGTCAAGAAGGAGAGGTGGC	62.4	234

Table 5: Primer sequences of PTPN1 SNP variants.

3.3.7 ARMS PCR

The ARMS PCR procedure was utilized for PTPN1 gene.

1. PTPN1 Gene preparation:
2. Two tubes were prepared.
3. Tube 1: 1µL of common primer, 1µL of forward primer, 21.5µL of master mix, 1µL of sample, and 0.5µL are added and mixed.
4. Tube 2: 1µL of common primer, 1µL of reverse primer, 21.5µL of master mix, 1µL of sample, and 0.5µL are added and mixed.

1	2	3	4	5	6	7
94 °C	94 °C	57 °C	72 °C	GOTO Step 2	72° C	4 °C
3:00	**00:30**	**00:50**	**00:40**	**30 X**	**3:00**	**∞**

Table 6: PCR program for PTPN gene.

3.3.8 Gel Electrophoresis

Once ARMS PCR was conducted, the product was then analyzed through the gel electrophoresis method.

The following steps were followed by the students performing this experiment.

1. 1.0g of agarose was measured and placed in a conical flask.
2. 100mL of the 1x TAE solution was poured into the same conical flask and then the flask was placed on a hot plate for 3.0 minutes so that all of the agarose was dissolved with the occasional mixing by the stirrer.
3. The agarose solution was allowed to cool down.
4. About 8µL of ethidium bromide was added to both the agarose solution and the running buffer solution.
5. The agarose gel was then gently poured into the casting tray, with the well combs in place. The gel was allowed to settle at room temperature for at least an hour.
6. Next, the loading buffer was added to each of the PCR products.
7. Once the gel was solidified and the well combs removed, it was place in the electrophoresis chamber.
8. The chamber was then filled with the 1x TAE solution until all of the gel was covered.
9. A molecular weight ladder was loaded into the first lane of the gel with the help of a pipette.
10. Next, the PCR products were loaded into the wells with the help of a pipette.
11. Once loading of the samples was done, the gel was run at 120V until the dye line was more than 75% down the gel. The run time of the gel was for 45 minutes.
12. The power supply was turned off and the gel was removed from the electrophoresis chamber carefully.
13. The gel was the visualized under a UV light source so that the fragments of the PCR products were seen as bands.
14. The results of the bands on the gel were analyzed with the DNA ladder in the first lane as a guide.

4. Results

4.1 In Silico Mutational Analysis

4.1.1 Identification of the Most Damaging SNPs.

A total of 261 missense SNPs were retrieved from the dbSNP database. These 261 SNPs were further analyzed by Ensembl Biomart and PROVEAN prediction tools to identify the most damaging SNPs. On the basis of the Polyphen and SIFT scores that were generated through Biomart, 56 SNPs were selected in a range from 0.46 – 1 generated by Polyphen and 0 - 0.04 generated by SIFT. SIFT classified these SNPs as deleterious. These 56 SNPs were inputted in PROVEAN for further analysis. PROVEAN filtered a total of 36 SNPs as the most damaging SNPs with a score ranging from 2.7 to 8.73.

Sr. No.	Variant name	Variant Alleles	PolyPhen score	SIFT score	PROVEAN Score	Prediction
1	rs763558512	G/A	0.789	0	4.08	Deleterious
2	rs1217160644	C/T	0.736	0.01	7.08	Deleterious
3	rs1231034866	C/A	0.611	0.01	3.1	Deleterious
4	rs1246414151	T/A	0.46	0.04	4.43	Deleterious
5	rs1325608761	A/T	0.8	0	3.88	Deleterious
6	rs1363544131	C/T	0.773	0	3.95	Deleterious
7	rs1487620720	G/T	0.864	0	5.76	Deleterious
8	rs372483862	C/T	0.986	0	7.55	Deleterious
9	rs753938375	A/C	0.998	0	8.73	Deleterious
10	rs753938375	A/G	0.994	0	8.73	Deleterious
11	rs759459286	G/C	1	0	5.9	Deleterious
12	rs765072009	G/A	0.998	0	3.47	Deleterious
13	rs762225731	A/G	0.982	0	5.46	Deleterious
14	rs760953373	T/G	0.988	0	2.93	Deleterious
15	rs763558512	G/C	0.927	0.02	4.08	Deleterious
16	rs768861855	G/C	0.999	0	2.97	Deleterious
17	rs779200250	T/C	0.995	0	3.18	Deleterious
18	rs750431969	C/G	0.994	0	2.71	Deleterious
19	rs967086398	G/T	0.984	0.01	6.44	Deleterious
20	rs1000153249	G/C	0.989	0.01	7.63	Deleterious
21	rs1177454007	G/T	0.998	0	4.42	Deleterious
22	rs1181068233	G/A	0.967	0	5.19	Deleterious
23	rs1224554949	G/A	0.994	0.02	7.24	Deleterious
24	rs1234412248	C/G	0.999	0	4.88	Deleterious
25	rs1242137245	T/C	0.992	0	3.6	Deleterious
26	rs1250785943	G/A	0.924	0.02	3.53	Deleterious
27	rs1285079073	A/G	1	0	3.62	Deleterious
28	rs1330974189	T/C	0.972	0.01	3.74	Deleterious
29	rs1364690180	A/G	0.982	0	6.19	Deleterious
30	rs1384590527	G/A	0.999	0.02	3.84	Deleterious
31	rs1435238590	T/G	0.92	0	7.69	Deleterious

32	rs1437231621	C/T	0.938	0.02	7.79	Deleterious
33	rs1452015420	C/T	1	0	7.81	Deleterious
34	rs1467589776	A/G	0.997	0	6.39	Deleterious
35	rs1601415665	A/C	0.999	0	5.29	Deleterious
36	rs1601415670	A/C	0.99	0	5.06	Deleterious

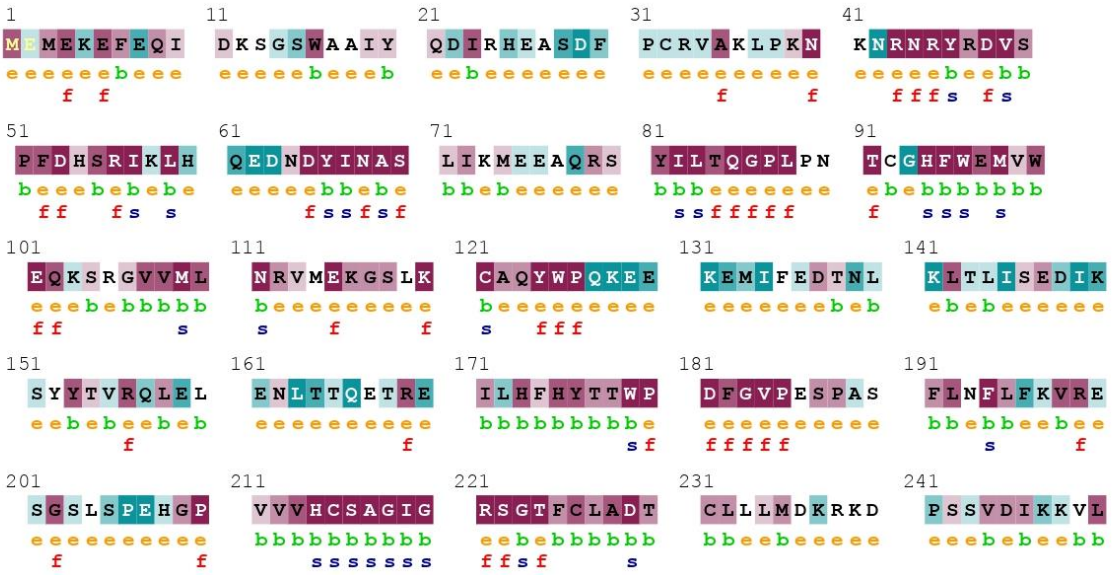
Table 7: Results of PolyPhen, SIFT, and PROVEAN.

4.1.2 Disease Causing SNPs

The 36 SNPs were submitted to PMut and MutPred2 to generate the pathology of these SNPs. PMut generated a score range of 0.5 to 0.9 and informed that all SNPs were disease causing. The SNPs were submitted to MutPred2 for further analysis and the results showed that the snips that had a score >0.5 were classified as pathogenic and the SNPs that had a score <0.5 were classified as benign. 0.5 was kept as the threshold. Provide results in table

4.1.3 Conservation Analysis

The SNPs went under a conservation analysis using ConSurf and the results produced were on the basis of a 1-10 scale with 1 being highly variable region and 9 being the highly conserved region. The results also informed about if the SNP is in a buried or an exposed region and if that SNPs lies in the functional or structural region.



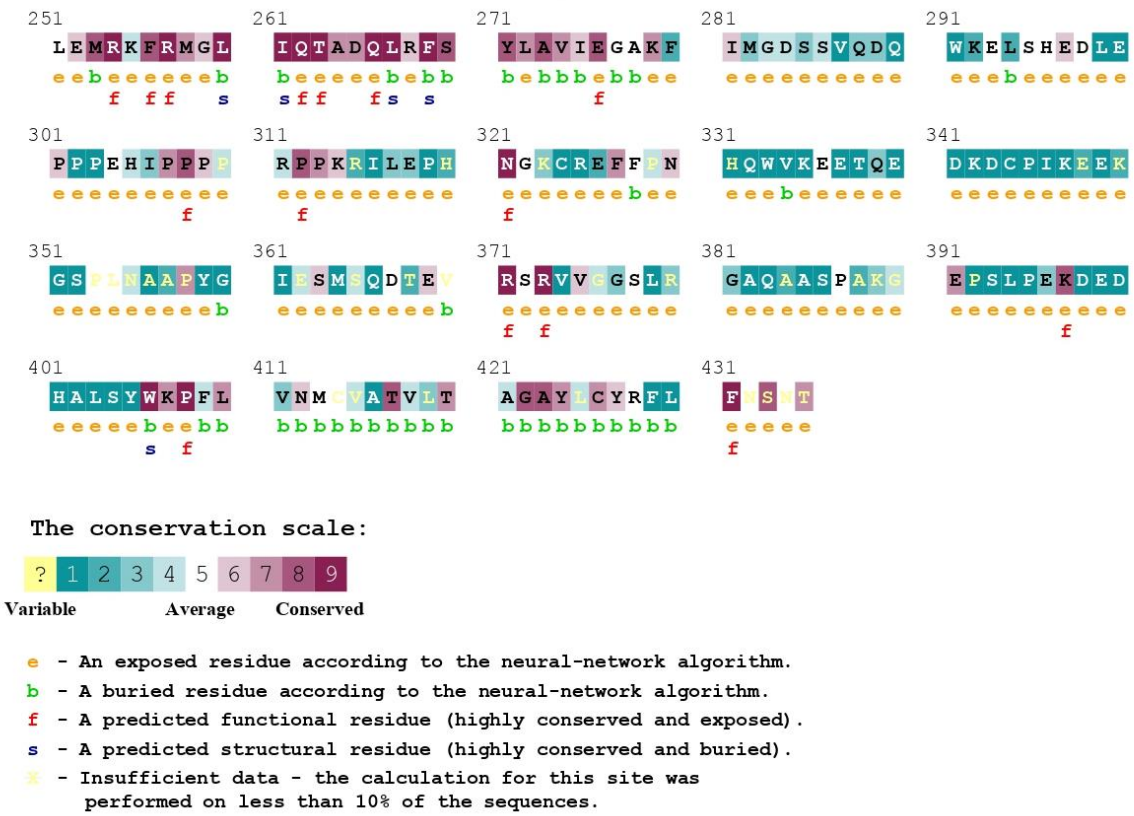


Figure 7: Conservation Analysis Results

Sr. No.	Variant name	Conserved Region Score	Exposed / Buried Site	Functional / Structural Region
1	rs372483862	9	Exposed	Functional
2	rs759459286	9	Exposed	Functional
3	rs765072009	8	Exposed	Functional
4	rs760953373	9	Exposed	Functional
5	rs763558512	7	Exposed	Functional
6	rs768861855	9	Exposed	Functional
7	rs779200250	9	Buried	Structural
8	rs750431969	7	Buried	-
9	rs1181068233	8	Buried	-
10	rs1224554949	7	Exposed	-
11	rs1234412248	9	Exposed	Functional
12	rs1242137245	7	Buried	-
13	rs1250785943	8	Exposed	Functional
14	rs1285079073	9	Exposed	Functional
15	rs1364690180	7	Exposed	-
16	rs1384590527	9	Exposed	Functional

17	rs1452015420	9	Exposed	Functional
18	rs1467589776	8	Exposed	Functional
19	rs1601415665	9	Exposed	Functional
20	rs1601415670	7	Buried	-

Table 8: ConSurf predictions for sites of nSNPs are found.

4.1.4 Predictions the effect of missense SNPs on Protein Stability

The SNPs were submitted to I-mutant 3.0 to check the protein stability. 7 out of 20 SNPs showed that they increased the stability of the protein. Whereas, the rest of the SNPs decreased the stability.

Sr. No.	Variant name	AAS	PhD-SNP	Stability	RI	DDG
1	rs372483862	P180S	Disease	Decrease	7	-1.55
2	rs759459286	R254S	Disease	Decrease	8	-1.16
3	rs765072009	R169Q	Disease	Decrease	1	-0.4
4	rs763558512	D265H	Disease	Decrease	7	-0.68
5	rs768861855	E115Q	Disease	Decrease	7	-0.56
6	rs779200250	I57T	Disease	Decrease	9	-2.52
7	rs750431969	I171M	Disease	Decrease	7	-1.22
8	rs1181068233	C226Y	Disease	Decrease	0	-0.1
9	rs1234412248	S222C	Disease	Decrease	7	-0.64
10	rs1250785943	E276K	Disease	Decrease	8	-0.96
11	rs1364690180	R112G	Disease	Decrease	8	-1.52
12	rs1384590527	R257Q	Disease	Decrease	7	-0.63
13	rs1452015420	R257W	Disease	Decrease	5	-0.14
14	rs1467589776	E276G	Disease	Decrease	8	-1.46
15	rs1601415665	T224P	Disease	Decrease	4	-0.54
16	rs1601415670	T230P	Disease	Decrease	5	-0.72

Table 9: Effect of deleterious nSNPs on Protein Stability.

4.1.5 Protein Modelling

The SNPs were submitted to Phyre2 so that the protein structures can be generated of both wild and mutant types. Once these were generated, they were submitted to ERRAT and VERIFY3D for validation. Followed by checking the stereo-chemical properties of the protein structures through PROCHECK. Lastly, the structures were checked for differences between the wild and mutant types by TM align tool. The RMSD and TM scores were generated as 1 mention your model value here and 0 with 1 representing a perfect match between the wild and mutant types.

Sr. No.	Variant Name	Amino Acid Change	ERRAT Value	VERIFT 3D Value	PROCHECK Score	TM align (RMSD)	TM Score
1	rs779200250	I57	90.6574	94.61	90.3	0	1
		T57	90.6574	94.61	90.3	0	1
2	rs759459286	R254	90.6574	94.61	90.3	0	1
		S254	90.6574	94.61	90.3	0	1
3	rs763558512	D265	90.6574	94.61	90.3	0	1
		H265	90.6574	94.61	90.3	0	1
4	rs768861855	E115	90.6574	94.61	90.3	0	1
		Q115	90.6574	94.61	90.3	0	1
5	rs765072009	R169	90.6574	94.61	90.3	0	1
		Q169	90.6574	94.61	90.3	0	1
6	rs750431969	I171	90.6574	94.61	90.3	0	1
		M171	90.6574	94.61	90.3	0	1
7	rs1181068233	C226	90.6574	94.61	90.3	0	1
		Y226	90.6574	94.61	90.3	0	1
8	rs1234412248	S222	90.6574	94.61	90.3	0	1
		C222	90.6574	94.61	90.3	0	1
9	rs1250785943	E276	90.6574	94.61	90.3	0	1
		K276	90.6574	94.61	90.3	0	1
10	rs372483862	P180	90.6574	94.61	90.3	0	1
		S180	90.6574	94.61	90.3	0	1
11	rs1467589776	E276	90.6574	94.61	90.3	0	1
		G276	91.0035	94.61	90.3	0	1
12	rs1364690180	R112	90.6574	94.61	90.3	0	1
		G112	89.9654	94.95	90.3	0	1
13	rs1384590527	R257	90.6574	94.61	90.3	0	1
		Q257	88.5813	94.61	90.3	0	1
14	rs1601415665	T224	90.6574	94.61	90.3	0	1
		P224	90.6574	94.61	90.3	0	1
15	rs1452015420	R257	90.6574	94.61	90.3	0	1
		W257	90.3114	94.61	90.3	0	1
16	rs1601415670	T230	90.6574	94.61	90.3	0	1
		P230	90.6574	94.61	90.3	0	1

Table 10: Results from ERRAT, VERIFY3D, PROCHECK, and TM Align tools.

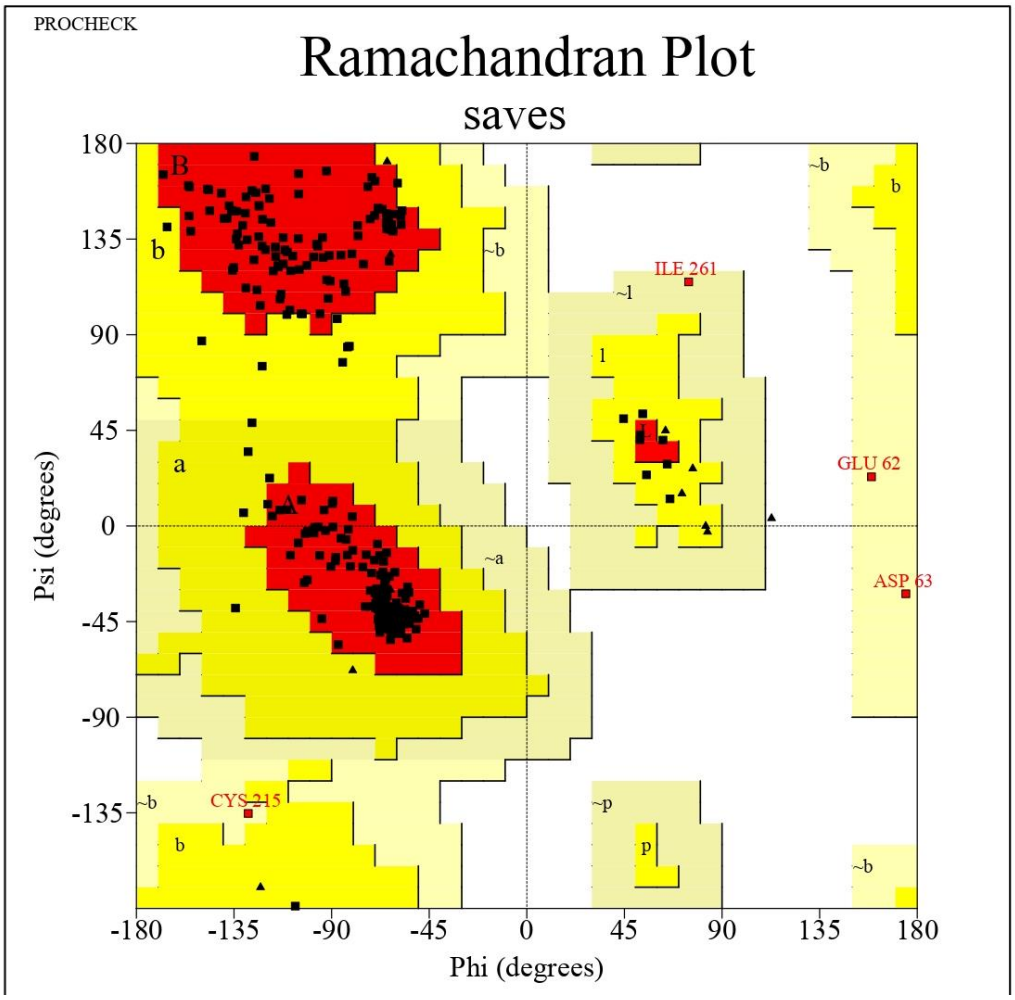


Figure 8: Ramachandran plot graph of PTPN1mutant (G276) values showing number of residues favored, allowed, and outlier regions.

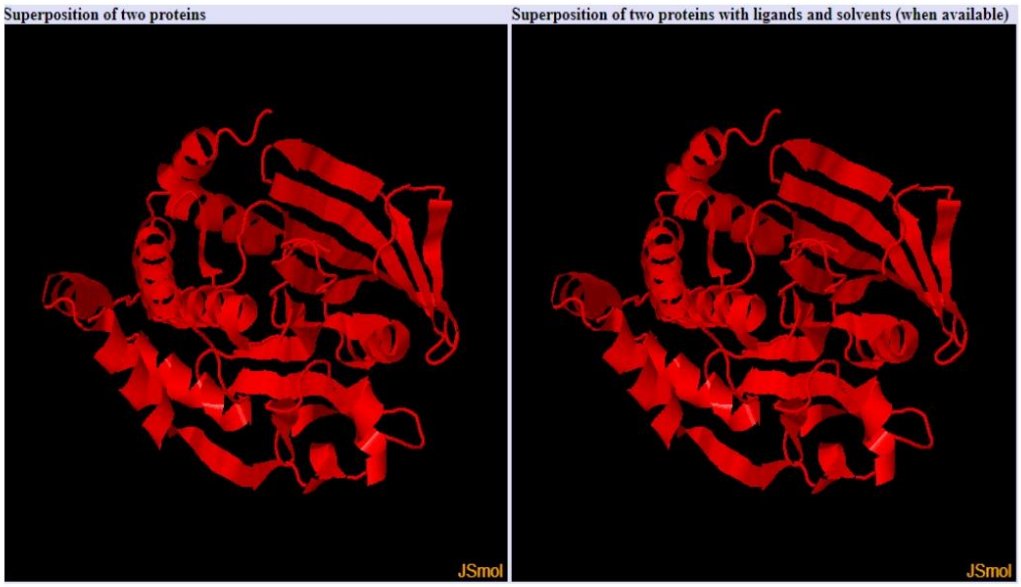


Figure 9: Superimposition of PTPN1 wild (E276) and mutant (G276) using TM-align tool.

4.1.6 Results of deleterious effects of intronic SNPs

Intronic SNPs were analyzed in regSNP-intron software. The results showed that both the SNPs were benign with a lower probability ranging from 0.1 to 0.37.

Sr.no.	SNP variants	position	ref	alt	disease	probability	splicing site
1	rs3787345	49185423	C	T	B	0.102252	off
2	rs718049	49196599	C	G	B	0.237641	off
3	rs718049	49196599	C	T	B	0.373311	off

Table 11: regSNP-intron results for intronic SNPs.

4.1.7 Results for the presence of splice sites.

The SNPs were further analyzed in the NetGene2 tool to evaluate whether the SNPs are located in the splice site region or not. The results showed that no donor and acceptor splice sites are found the or mention the threshold value for both of these intronic SNPs.

4.1.8 Results for the effect on Transcription Factor Binding Sites (TFSB)

In order to determine the effect of variants on the transcription binding sites (TFBS), the SNPs were analyzed in the SNP2TFBS software. The results demonstrated that no TFBS were affected by the SNP *rs718049* and *rs3787345*. The transcription binding sites were explored using the ConSite software and the results revealed that no transcription binding sites were detected in the region where intronic variants are located.

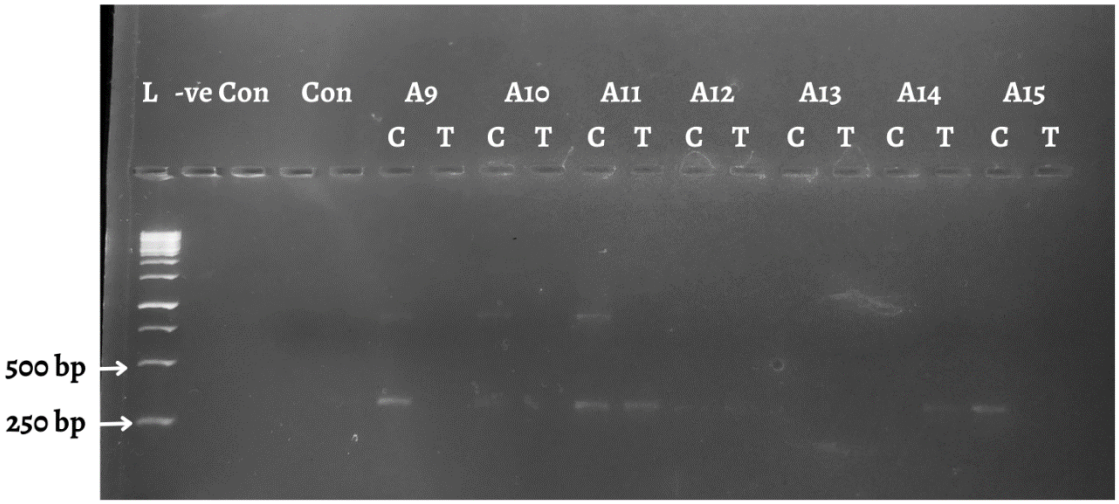
4.1.9 Results for the regulatory potential

The SNPs were investigated in the RegulomeDB. Results revealed a score of 3a for the *rs3787345* SNP and a score of 4 for the *rs718049* SNP demonstrating the regulatory potential evidence.

4.2 Genotypic Analysis

The results of the ARMS PCR were validated by gel electrophoresis for the two intronic SNPs.

- SNP *rs718049*
  SNP *rs3787345*



L = Ladder of 1k bp C allele amplicon size = 336bp
A = Diabetic samples T allele amplicon size = 336bp
Con = Control

Figure 10: Gel results of *PTPN1 rs718049* variant.

4.3 Statistical Analysis

	Total Control	**Total patients**				
	N=4 n (%)	N=31n (%)	OR	P	RR	P
CC	0(0)	12(38.7)	5.7692	0.2533	3.9063	0.3163
CT	0(0)	19(61.2)	14.0400	0.0850	6.0938	0.1804
TT	4 (100)	0(0)	0.0018	0.0021	0.0174	0.0041

Table 12: Genotypic frequencies of *PTPN1 rs718049* polymorphism.

Genotypic distribution of *rs718049* (C-T) shows that the heterozygous genotype (CT) was prevalent in the Type 2 diabetic patients (61.2%) whereas the homozygous (TT) as predominant in the control group (100%). Among all the genotypes CT was prominent in patients, and TT was prominent in control group. However, homozygous (CC) genotype was also found in diabetic patients (38.7) having a lower prevalence than heterozygous (CT)

genotype. The OR value for the association of *rs718049* polymorphism depicted a strong association of genotype CT with the incidence of T2DM (OR=14.0400). However, a larger sample size is required to further validate the role of this genotype (CT) in the incidence of T2DM.

5. Discussions

Type 2 Diabetes mellitus is one of the most common metabolic diseases worldwide. Diabetes is associated with a wide range of life-threatening complications. Diabetes is characterized by insulin resistance, hyperglycemia and dysfunction of beta cells. People diagnosed with T2DM diabetes are at an elevated risks of developing a scope of complications, such as nephropathy, cardiovascular diseases, and visual impairment. Genetic disposition and ecological impacts have been observed as the main reasons for the occurrence of the disease.

5.1 In Silico Analysis

5.1.1 Missense SNPs

Single nucleotide polymorphisms (SNPs) are the most frequent type of genetic variations that occur in the world and have been linked to many genetic diseases. About 2% SNPs that are linked with genetic diseases are SNPs and they occur in the protein coding regions and have led to a single amino acid substitution. The SNPs identified for PTPN1 gene have been corelated with the PTP1B protein that is encoded by the gene. Disturbances in the PTP1B protein have been associated with diseases such as Type 2 Diabetes Mellitus. Through in silico analysis, the SNPs were identified along with their effect on the structure and function of the PTP1B protein.

The 261 SNPs were submitted to BioMart that produced Polyphen and SIFT results. PolyPhen is a software which helps in predicting the effect of an amino acid substitution on the structural and functional properties of a protein. PolyPhen scores are generated on the basis of 0 being the least damaging, and 1 being the most damaging. The SNPs that had a score closest to 1 were selected that were further filtered out by SIFT. Sorting intolerant from tolerant (SIFT) is a software that predicts the impact of an amino acid substitution has on the functional properties of a protein that is based on the sequence homology and the physical properties. SIFT scores are generated on the basis of 0, being the most damaging, and 1, being the least damaging. The SNPs that had a score closest to 0 were selected. In total, 56 SNPs were selected which were further analyzed by PROVEAN and a total of 36 SNPs were classified as the most damaging SNPs. These 36 SNPs were common among these two softwares.

When the 36 SNPs were submitted to PMut to know whether they are disease causing or not, the results showed that they were all disease-causing snips. For further analysis, the snips were submitted to MutPred2 where the results showed majority of the SNPs were pathogenic. Mutpred2 informed that 6 of the snips were had a score in the range of 0-0.5 which meant that these snips were benign while the rest of the thirty snips were pathogenic as they were in 0.5-

1.0 range. The probability of the snips to be benign or pathogenic is corelated with the gain and loss of function of the protein. The snips that showed a probability of 0.5 and lower were classified as benign and the snips that had a probability of 0.5 and higher were classified as pathogenic. The pathogenic snips can cause significant functional damage to the protein. 30 out of 36 SNPs had a score >0.5 and were selected for further analysis.

The conservation analysis of these 30 deleterious SNPs was determined by ConSurf. The alignments of amino acid residues in protein have marked that may lead to disease if mutations are caused. The conserved regions are important for a protein both functionally and structurally. The SNPs that can replace the conserved amino acids can cause a disruption in the function and structure of a protein. SNPs present at the conserved regions have known to be highly damaging for the protein and may lead to disease. 20 out of 30 SNPs (grade 7 and above) were considered to be detrimental to the protein. From those 20 SNPs, 14 (grade 8 and 9) of the SNPs were replacing amino acids in highly conserved regions and were predicted by ConSurf as functional residues. Therefore, these 14 SNPs can cause a significant functional damage to the protein. Whereas, only one SNP was substituting a highly conserved amino acid and can cause significant structural damage to the protein (Table *8*).

Missense SNPs can affect the protein stability and the results were in correspondent with this. The 16 of the SNPs were predicted to decrease the stability of the protein while 4 were increasing the protein stability. Two of the SNPs that were increasing the protein stability were contradicting with their ΔΔG values as they below 0. According to the I-mutant3.0 website, the predictions of the ΔΔG values and signs are carried by two different SVMs which can contradict the sign and value. This occurs when the reliability index is low, which is true for one of the contradictory cases only. When the protein is destabilized, it can lead to increase in degradation, aggregation, and misfolding of the protein. The 16 SNPs that were decreasing the protein stability were submitted forward for additional analysis (Table *9*).

The 16 SNPs were submitted to Phyre2 software so that the protein structures of wild and mutant types for each of the snips could be generated. The accuracy of protein structure models was then verified through ERRAT, and VERIFY3D to evaluate the information of the non-bonded connections between the atoms and analyze the compatibility of the 3D models with the amino acid sequences. The stereo-chemical properties of the protein models were verified by PROCHECK where Ramachandran plot graphs were created. The Ramachandran plot graphs revealed that the key residues of the PTPN1 models were portioned in acceptable

segments of the map (**Error! Reference source not found.**). In all the models, the largest number of residues was present in the mutant G276 in favorable regions with only few residues seen in unfavorable regions, which signifies the quality of the models. G276 was chosen on the basis of attaining the highest ERRAT score (91%). TM align tool superimposed these 16 SNPs, both wild and mutant structure as to check any differences between the structures. All of the SNPs scored a 1 which means that the structures are a perfect match. The RMSD scores of this superimposed structure was 0 which meant that there was no change in the structures. The superimposition of the PTPN1 wild (E276) and mutant (G276) types were also on the basis of the ERRAT score.

5.1.2 Intronic SNPs

In silico analysis of the selected intronic SNPs *rs3787345* and *rs718049* was carried out for determining the regulatory effect of these SNPs. The RegSNP-intron software was used to determine the disease-causing nature of these SNPs. Chromosome number, coordinate position, and the reference and variant alleles were entered. Both of these SNPs were revealed to be benign in nature with a lower probability of causing the disease ranging from 0.1 to 0.37.

Further analysis was conducted by NetGene2 software to evaluate whether the SNPs are located in the splice site region or not. For both these intronic SNPs, no splice sites in the donor and acceptor region were found above the threshold. The cut off value used for confidence of nearly all true donor sites is 50% and for nearly all true acceptor sites is 20%.

To determine the role of these variants on TFBS, SNP2TFBS tool is used which maps the variants to the TFBS. No effect was manifested on the transcription binding sites by these variants. Further analysis for the presence of transcription binding sites in the region of SNP was checked using ConSite software. Transcription binding sites were not detected in that region showing that these SNPs are not involved in the regulation of transcription.

The regulatory functions of these intronic SNPs were studied using RegulomeDB which is a database used for the identification of the DNA features and regulatory potential of the variants in the non-coding regions. Variants in this database are described into four categories that represent their regulatory potentials. Intronic variant *rs3787345* has a RegulomeDB score of 3a predicting a less likely effect of this variant on binding. Similarly, the variant *rs718049* has a RegulomeDB score of 4 predicting minimal binding evidence of this variant.

5.2 Genotypic Analysis

The Amplification Refractory Mutation System PCR (ARMS PCR) is a method utilize for genetic diseases diagnosis. ARMS PCR was used to check for the presence of the wild and/or mutant alleles of intronic SNPs. The results of the ARMS PCR of the following two intronic SNPs were verified by agarose gel electrophoresis. The *rs718049* has **"C"** as the wild allele and the **"T"** as the mutant allele in the Caucasian population. For the South Asian population, the **"T"** is the mutant allele and the **"C"** is the mutant allele for the *rs718049*. The *rs718049* results were shown in the gel that were visualized by the UV illuminator. Both the "C" and the "T" alleles were visible, which means that rs718049 is heterozygous for the diabetic patients.

6. Conclusion

6.1 Conclusion

In silico analysis of 16 damaging missense SNPs revealed their involvement in dysregulating the function of *PTPN1* representing they may be involved in disease pathogenesis. Next, in in silico analysis of intronic SNPs *rs3787345* and *rs718049* showed no significant role in regulation of *PTPN1* gene. Further validation of the selected intronic SNPs through genotypic analysis represented a role of *rs718049* in the pathogenicity of diabetic patients. The genotypic distribution depicts higher prevalence of heterozygous trend in *rs718049*. In case of *rs3787345*, resulted nonspecific bands, hence needs additional optimization for detecting its role in diabetic patients. For further verification of results, larger sample size is required in order to replicate these findings.

6.2 Future Prospects

Study of *rs3787345* and *rs718049* polymorphisms in *PTPN1* gene of a large sample size in Type 2 diabetic patients can be beneficial in providing significant details for the development of new therapeutic drug target. Also, genome wide association sequencing (GWAS) is required for this gene to help in diagnosing the disease. Moreover, further validation through wet-lab analysis is needed.

.

7. References

7.1 References

Abdelsalam, S. S., Korashy, H. M., Zeidan, A., & Agouni, A. (2019). The role of protein tyrosine phosphatase (PTP)-1B in cardiovascular disease and its interplay with insulin resistance. In Biomolecules (Vol. 9, Issue 7). MDPI AG. https://doi.org/10.3390/biom9070286

Ali, O. (2013). Genetics of type 2 diabetes. World Journal of Diabetes, 4(4), 114. https://doi.org/10.4239/wjd.v4.i4.114

Amiel, S. A. (2011). Pathophysiology of Type2 Diabetes Mellitus. In S. A. Amiel, Oxford Textbook of Endocrinology and Diabetes. Oxford Medicine. doi:https://oxfordmedicine.com/view/10.1093/med/9780199235292.001.1/med-9780199235292-chapter-130302

Anhalt, D. H., Diaz-Thomas, M. M., Umpierrez, M. G., Kitabchi, M. P., Arora, M. F., Mestman, M. J., Mifrakhraee, M. S. (2022, January 23). Diabetes and Endocrine Function. (Endocrine Society) Retrieved from Endocrine : https://www.endocrine.org/patient-engagement/endocrine-library/diabetes-and-endocrine-function#:~:text=The%20pancreas%20produces%20the%20hormone,of%20glucose%20in%20the%20blood

Asif, M. (2014). The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern. Journal of Education and Health Promotion, 3(1), 1. https://doi.org/10.4103/2277-9531.127541

Bento, J. L., Palmer, N. D., Mychaleckyj, J. C., Lange, L. A., Langefeld, C. D., Rich, S. S., Bowden, D. W. (2004, November 01). Association of Protein Tyrosine Phosphatase 1B Gene Polymorphisms With Type 2 Diabetes. Brief Genetics Report, 6. doi:https://doi.org/10.2337/diabetes.53.11.3007

CDC. (2021, August 10). Diabetes. Retrieved from CDC: https://www.cdc.gov/diabetes/basics/getting-tested.html

CDC. (2021, August 10). Diabetes Tests. Retrieved from CDC: https://www.cdc.gov/diabetes/basics/getting-tested.html

Cerf, M. E. (2013). Beta cell dysfunction and insulin resistance. In Frontiers in Endocrinology (Vol. 4, Issue MAR). https://doi.org/10.3389/fendo.2013.00037

Cheyssac, C., Lecoeur, C., Dechaume, A., Bibi, A., Charpentier, G., Balkau, B., Vaxillaire, M. (2006, May 05). Analysis of common PTPN1 gene variants in type 2 diabetes, obesity and associated phenotypes in the French population. BMC Medical Genetics, 10. doi:10.1186/1471-2350-7-44

Chu, D., & Wei, L. (2019). Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. BMC Cancer, 19(1). https://doi.org/10.1186/s12885-019-5572-x

Emadi, E., Akhoundi, F., Kalantar, S. M., & Emadi-Baygi, M. (2020). Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences. BMC Genetics, 21(1). https://doi.org/10.1186/s12863-020-00890-y

Farbstein Bsc, D., & Levy MD, P. A. (2010, August 28). The Genetics of Vascular Complications in Diabetes Mellitus. ScienceDirect, 19. doi:https://doi.org/10.1016/j.ccl.2010.04.005

Galicia-Garcia, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K. B., Ostolaza, H., & Martín, C. (2020). Pathophysiology of type 2 diabetes mellitus. In International Journal of Molecular Sciences (Vol. 21, Issue 17, pp. 1–34). MDPI AG. https://doi.org/10.3390/ijms21176275

GeneCards, T. H. (2022, February 21). PTPN1 Gene - Protein Tyrosine Phosphatase Non-Receptor Type 1. Retrieved from GeneCardsSuite: https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTPN1

Goyal, R., & Jialal, I. (2021, September 28). NCBI. Retrieved from NCBI: https://www.ncbi.nlm.nih.gov/books/NBK513253/

Huang, S., Liu, L., Xiang, Y., Wang, F., Yu, L., Zhou, F., Yu, Z. (2019, February 25). Association of PTPN1 polymorphisms with breast cancer risk: A case-control study in Chinese females. Journal of Cellular Biochemistry, 12. doi:https://doi.org/10.1002/jcb.28490

Hunt, R., Sauna, Z. E., Ambudkar, S. v., Gottesman, M. M., & Kimchi-Sarfaty, C. (2009). Silent (synonymous) SNPs: should we care about them? In Methods in molecular biology (Clifton, N.J.) (Vol. 578, pp. 23–39). https://doi.org/10.1007/978-1-60327-411-1_2

IDF. (2021). IDF Diabetes Atlas. Retrieved from Diabetes Atlas: https://diabetesatlas.org/

IDF. (2021, December 06). IDF Diabetes Atlas. Retrieved from Diabetes Atlas: https://www.idf.org/index.php?option=com_attachments&task=download&id=2644:WDD2021_PAKISTAN_PR_Final

Kolb, H., & Martin, S. (2017). Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. In BMC Medicine (Vol. 15, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12916-017-0901-x

Lin, H., Hargreaves, K. A., Li, R., Reiter, J. L., Wang, Y., Mort, M., Cooper, D. N., Zhou, Y., Zhang, C., Eadon, M. T., Dolan, M. E., Ipe, J., Skaar, T. C., & Liu, Y. (2019). RegSNPs-intron: A computational framework for predicting pathogenic impact of intronic single nucleotide variants. Genome Biology, 20(1). https://doi.org/10.1186/s13059-019-1847-4

Mahler, R. J., & Adler, M. L. (1999). CLINICAL REVIEW 102 Type 2 Diabetes Mellitus: Update on Diagnosis, Pathophysiology, and Treatment. https://academic.oup.com/jcem/article/84/4/1165/2864079

MayoClinic. (2021, January 21). Type 2 Diabetes. Retrieved from MayoClinic: https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/diagnosis-treatment/drc-20351199#:~:text=Type%202%20diabetes%20is%20usually,past%20two%20to%20three%20months.

MedlinePlus. (2017, November 01). Type 2 Diabetes. (NIH) Retrieved from National Library of Medicine: https://medlineplus.gov/genetics/condition/type-2-diabetes/#inheritance

MedlinePlus. (2022, March 22). What are Single Nucleotide Polymorphisms (SNPs)? (NIH) Retrieved from MedlinePlus - Trusted Health Information for You: https://medlineplus.gov/genetics/understanding/genomicresearch/snp/

NIDDK. (2016, December). Diabetes Tests and Diagnosis. (NIH) Retrieved from National Institute of Diabetes and Digestive and Kidney Diseases: https://www.niddk.nih.gov/health-information/diabetes/overview/tests-diagnosis#diagnosediabetes

NIDDK. (2022, March). Insulin, Medicines, & Other Diabetes Treatments. (NIH) Retrieved from National Institute of Diabetes and Digestive and Kidney Diseases: https://www.niddk.nih.gov/health-information/diabetes/overview/insulin-medicines-treatments

Reed, J., Bain, S., & Kanamarlapudi, V. (2021). A review of current trends with type 2 diabetes epidemiology, aetiology, pathogenesis, treatments and future perspectives. In Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy (Vol. 14, pp. 3567–3602). Dove Medical Press Ltd. https://doi.org/10.2147/DMSO.S319895

Rocha, S., Luísa Corvo, M., Fernandes, E., & Freitas, M. (2021). Journal of Diabetes and Clinical Research Commentary https://www.scientificarchives.com/journal/journal-of-diabetes-and-clinical-research The Emerging Target Protein Tyrosine Phosphatase 1B (PTP1B) for Type 2 Diabetes Mellitus Management. In J Diabetes Clin Res (Vol. 3, Issue 4). https://www.scientificarchives.com/journal/journal-of-diabetes-and-clinical-research

Sabiha, B., Bhatti, A., Roomi, S., John, P., & Ali, J. (2021). In silico analysis of non-synonymous missense SNPs (nsSNPs) in CPE, GNAS genes and experimental validation in type II diabetes mellitus through Next Generation Sequencing. Genomics, 113(4), 2426–2440. https://doi.org/10.1016/j.ygeno.2021.05.022

Samanthi, D. (2017, February 27). Difference Between SNP and Mutation. Retrieved from DifferenceBetween.com: https://www.differencebetween.com/difference-between-snp-and-vs-mutation/

Teimouri, M., Hosseini, H., ArabSadeghabadi, Z., Babaei-Khorzoughi, R., Gorgani-Firuzjaee, S., & Meshkani, R. (2022). The role of protein tyrosine phosphatase 1B (PTP1B) in the pathogenesis of type 2 diabetes mellitus and its complications. In Journal of Physiology and Biochemistry. Springer Science and Business Media B.V. https://doi.org/10.1007/s13105-021-00860-7

Venkata Subbiah, H., Ramesh Babu, P., & Subbiah, U. (2020). In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene. Egyptian Journal of Medical Human Genetics, 21(1). https://doi.org/10.1186/s43042-020-00110-3

Villamar-Cruz, O., Loza-Mejia, M. A., Arias-Romero, L. E., & Camacho-Arroyo, I. (2021). Recent advances in PTP1B signaling in metabolism and cancer. In Bioscience Reports (Vol. 41, Issue 11). Portland Press Ltd. https://doi.org/10.1042/BSR20211994

WHO. (2021, November 10). WHO Newsroom. Retrieved from WHO: https://www.who.int/news-room/fact-sheets/detail/diabetes

Wondmkun, Y. T. (2020). Obesity, insulin resistance, and type 2 diabetes: Associations and therapeutic implications. In Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy (Vol. 13, pp. 3611–3616). Dove Medical Press Ltd. https://doi.org/10.2147/DMSO.S275898

Zheng, Y., Ley, S. H., & Hu, F. B. (2018a). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. In Nature Reviews Endocrinology (Vol. 14, Issue 2, pp. 88–98). Nature Publishing Group. https://doi.org/10.1038/nrendo.2017.151

Zou, W. (2016). Type 2 Diabetes Mellitus and Protein-Tyrosine Phosphatase 1B1. Journal of Diabetes, Metabolic Disorders & Control, 3(8). https://doi.org/10.15406/jdmdc.2016.03.00096

BS Thesis-2022

ORIGINALITY REPORT

11%	7%	6%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Bibi Sabiha, Attya Bhatti, Sohaib Roomi, Peter John, Johar Ali. "In silico analysis of non-synonymous missense SNPs (nsSNPs) in CPE, GNAS genes and experimental validation in type II diabetes mellitus through Next Generation Sequencing", Genomics, 2021 Publication	1%
2	Submitted to Higher Education Commission Pakistan Student Paper	1%
3	hdl.handle.net Internet Source	<1%
4	Maryam Teimouri, Hossein Hosseini, Zahra ArabSadeghabadi, Reyhaneh Babaei-Khorzoughi, Sattar Gorgani-Firuzjaee, Reza Meshkani. "The role of protein tyrosine phosphatase 1B (PTP1B) in the pathogenesis of type 2 diabetes mellitus and its complications", Journal of Physiology and Biochemistry, 2022 Publication	<1%

5	Wei Zou. "Type 2 Diabetes Mellitus and Protein-Tyrosine Phosphatase 1B1", Journal of Diabetes, Metabolic Disorders & Control, 2016 Publication	<1%
6	www.slideshare.net Internet Source	<1%
7	www.ncbi.nlm.nih.gov Internet Source	<1%
8	Submitted to Adventist University of Health Sciences Student Paper	<1%
9	plus.si.cobiss.net Internet Source	<1%
10	Submitted to University of Cambridge Student Paper	<1%
11	link.springer.com Internet Source	<1%
12	www.science.gov Internet Source	<1%
13	oamjms.eu Internet Source	<1%
14	www.nature.com Internet Source	<1%
15	Submitted to UC, San Diego Student Paper	<1%

16	www.frontiersin.org Internet Source	<1%
17	www.freepatentsonline.com Internet Source	<1%
18	www.mdpi.com Internet Source	<1%
19	Submitted to National University of Singapore Student Paper	<1%
20	Submitted to Universitas Negeri Semarang Student Paper	<1%
21	Howeida Abdullah Mustafa, Afraa Mohamed Suliman Albkrye, Buthiena Mohamed AbdAlla, Mona AbdelRahman Mohammed Khair et al. " Computational determination of human gene: SNPs and prediction of their effect on protein functions of diabetic patients ", Clinical and Translational Medicine, 2020 Publication	<1%
22	Ki Bum Park, Kyong-Hwa Jun, Kyo Young Song, Hyungmin Chin, Han Hong Lee. "Development of a staging system and survival prediction model for advanced gastric cancer patients without adjuvant treatment after curative gastrectomy: A retrospective multicenter cohort study", International Journal of Surgery, 2022 Publication	<1%

23	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
24	portlandpress.com Internet Source	<1%
25	prr.hec.gov.pk Internet Source	<1%
26	R. Magesh, C. George Priya Doss. "Computational Methods to Work as First-Pass Filter in Deleterious SNP Analysis of Alkaptonuria", The Scientific World Journal, 2012 Publication	<1%
27	Submitted to VIT University Student Paper	<1%
28	core.ac.uk Internet Source	<1%
29	Submitted to iGroup Student Paper	<1%
30	Lipika R. Pal, John Moult. "Genetic Basis of Common Human Disease: Insight into the Role of Missense SNPs from Genome-Wide Association Studies", Journal of Molecular Biology, 2015 Publication	<1%
	docksci.com	

31	Internet Source	<1%
32	Samia Sultana Lira, Ishtiaque Ahammad. " A comprehensive investigation into the nsSNPs of gene predicts significant functional consequences in dopamine signaling and pharmacotherapy ", Cold Spring Harbor Laboratory, 2021 Publication	<1%
33	Submitted to Republic of the Maldives Student Paper	<1%
34	adamcertificationdemo.adam.com Internet Source	<1%
35	Submitted to King's College Student Paper	<1%
36	Submitted to University of East London Student Paper	<1%
37	Submitted to University of Warwick Student Paper	<1%
38	es.scribd.com Internet Source	<1%
39	journals.lww.com Internet Source	<1%
40	seabreezed.com Internet Source	<1%

41	Abdelrahman H. Abdelmoneim, Mujahed I. Mustafa, Thwayba A. Mahmoud, Naseem S. Murshed, Mohamed A. Hassan. " In Silico Analysis and Modeling of Novel Pathogenic Single Nucleotide Polymorphisms (SNPs) in Human Gene ", Cold Spring Harbor Laboratory, 2019 Publication	<1%
42	"JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice", Heart, 2005 Publication	<1%
43	Ingrid Rebouças de Moura, Franco Jefferds dos Santos Silva, Luis Henrique Gonçalves Costa, Edmon Darwich Neto et al. "Airport pavement evaluation systems for maintenance strategies development: a systematic literature review", International Journal of Pavement Research and Technology, 2020 Publication	<1%
44	Submitted to Oxford Brookes University Student Paper	<1%
45	Rania Jbir Koubaa, Mariem Ayadi, Mohamed Najib Saidi, Safa Charfeddine, Radhia Gargouri Bouzid, Oumèma Nouri-Ellouz. "Comprehensive Genome-Wide Analysis of The Catalase Enzyme Toolbox In Potato	<1%

(Solanum Tuberosum L.)", Research Square Platform LLC, 2021
Publication

46	docplayer.nl Internet Source	<1%
47	tampub.uta.fi Internet Source	<1%
48	www.4healtheducation.com Internet Source	<1%
49	Corton, Marta, Sorina D Tatu, Almudena Avila-Fernandez, Elena Vallespín, Ignacio Tapias, Diego Cantalapiedra, Fiona Blanco-Kelly, Rosa Riveiro-Alvarez, Sara Bernal, Blanca García-Sandoval, Montserrat Baiget, and Carmen Ayuso. "High frequency of CRB1 mutations as cause of Early-Onset Retinal Dystrophies in the Spanish population", Orphanet Journal of Rare Diseases, 2013. Publication	<1%
50	D. Thirumal Kumar, C. George Priya Doss, P. Sneha, Iftikhar Aslam Tayubi, R. Siva, Chiranjib Chakraborty, R. Magesh. "Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular dynamics approach", Journal of Biomolecular Structure and Dynamics, 2016 Publication	<1%

51	Macsphere.mcmaster.ca Internet Source	<1%
52	Submitted to University of Nevada Reno Student Paper	<1%
53	doaj.org Internet Source	<1%
54	journals.plos.org Internet Source	<1%
55	onlinelibrary.wiley.com Internet Source	<1%
56	pericles.pericles-prod.literatumonline.com Internet Source	<1%
57	George Priya Doss, C., B. Rajith, Chiranjib Chakraboty, V. Balaji, R. Magesh, B. Gowthami, Sneha Menon, M. Swati, Manjari Trivedi, Jasmine Paul, Richa Vasan, and Maitreya Das. "In silico profiling and structural insights of missense mutations in RET protein kinase domain by molecular dynamics and docking approach", Molecular BioSystems, 2014. Publication	<1%
58	Yoko Koh, Kosuke Nakano, Kotoe Katayama, Gaku Yamamichi et al. "Early dynamics of circulating tumor DNA predict clinical response to immune checkpoint inhibitors in	<1%

metastatic renal cell carcinoma", International Journal of Urology, 2022
Publication

59	scholarbank.nus.edu.sg Internet Source	<1%
60	worldwidescience.org Internet Source	<1%
61	www.ajas.info Internet Source	<1%
62	www.jaypeedigital.com Internet Source	<1%
63	www.minervamedica.it Internet Source	<1%
64	www.nutsocwyb.org Internet Source	<1%
65	Bhoomi Tarapara, Franky Shah. " An analysis to identify structural, functional and regulatory role of SNPs in ", Journal of Biomolecular Structure and Dynamics, 2022 Publication	<1%
66	Christina N. Lessov, Gary E. Swan, Huijun Z. Ring, Taline V. Khroyan, Caryn Lerman. "Genetics and Drug Use as a Complex Phenotype", Substance Use & Misuse, 2004 Publication	<1%

67 Momand, Jamil. "Concepts in Bioinformatics and Genomics", Oxford University Press <1%

Publication

Exclude quotes	On	Exclude matches	< 4 words
Exclude bibliography	On		