“Investigation of Insulin promoter factor 1 gene C18R and R197H variants association with type 2 diabetes mellitus in Bangladeshi Population”

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN BIOTECHNOLOGY.

Submitted by

Tasnin Khan Eusufzai

ID: 14276004

Masters in Biotechnology,

Department of Mathematics and Natural Sciences, BRAC University.

DECLARATION

This to declare that the research work embodying the results reported in this thesis entitled "Investigation of Insulin promoter factor 1 gene C18R and R197H variants association with type 2 diabetes mellitus in Bangladeshi Population" submitted by Tasnin Khan Eusufzai, has been carried out under the supervision and able guidance of Dr. M. Mahboob Hossain, Professor, Microbiology Program, Department of Mathematics and Natural Sciences, BRAC University, Dhaka.

It is further declared that the research work presented here is original, has not been submitted anywhere else for any degree or diploma. Any reference to work done by any other person or institution or any material obtained from other sources has been duly cited and referenced.

_____

(Tasnin Khan Eusufzai)
Candidate

Certified by:

_____

(Dr. M. Mahboob Hossain)
Supervisor
Professor
Microbiology Program
Department of Mathematics and Natural Sciences
BRAC University, Dhaka.

Dedicated

To

Those, who will be benefited from this work!

Acknowledgement

I am extremely grateful to God, who gave me the understanding and perseverance to make this accomplishment possible.

I am gratified and would like to express my sincere thankfulness and honor to my respected supervisor Professor Dr. M. Mahboob Hossain, Microbiology Program, Department of Mathematics and Natural Sciences, BRAC University, for his constant supervision, expert guidance, encouragement to follow new ideas and constant support throughout the entire period of my research work. Without his supervision and help, this dissertation would not have been possible. My special and sincere thanks to Professor Dr. Naiyyum Chaudhury, (Ex-coordinator, Biotechnology Masters Program, BRAC University), Chairman, Bangladesh Atomic Energy Regulatory Authority, for his keen involvement in my work all the time during my thesis period.

I am obliged and thankful to Professor Dr. Md. Zahid Hassan, Ph.D., Head of the Department of Physiology and Biochemistry, Bangladesh University of Health Sciences, who allowed me to work in his laboratory as well as supervise and guide me actively.

I also express my sincere appreciation and gratitude to Professor A F M Yusuf Haider, Ph.D., Chairperson, Department of Mathematics and Natural Sciences, BRAC University for his co-operation and encouragement in this study.

I would also like to express my gratitude to all of the department's senior teachers who helped me solve numerous problems during this project.

My special thanks go to Tridib Paul, Ashrafuzzaman Liton, Mahfuzur Rahman and Aqief Afzal for their continuous assistance in my project.

Last but not the least, I owe my parents, for their prolonged patience and nourishment towards my achievement and will ever be in debt to the people who helped me in carrying out the research work by contributing their precious time and suggestions.

Tasnin Khan Eusufzai

July, 2019

LIST OF ABBRIVIATION

ADA	American Diabetic Association
BADAS	Bangladesh Diabetic Samiti
BIHS	Bangladesh Institute of Health Sciences
BMI	Body Mass Index
BUHS	Bangladesh University of Health Sciences
DM	Diabetes Mellitus
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
EDTA	Ethylene Diamine Tetraacetic Acid
GDM	Gestational Diabetes Mellitus
Ht	Heterozygous
Hz	Homozygous
IDDM	Insulin-dependent Diabetes Mellitus
IPFI	Insulin Promoter Factor
IGT	Impaired Glucose Tolerance
IFG	Impaired Fasting Glucose
IR	Insulin Resistance
IDF	International Diabetic Federation
NIDDM	Non-insulin dependent Diabetes Mellitus
OPD	Out-patient Department
PCR	Polymerase Chain Reaction
RE	Restriction Enzyme
RFLP	Restriction Fragment Length Polymorphism
SPSS	Statistical Package for Social Science
T1DM	Type 1Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
TBE	Tris-boric EDTA
WHR	Waist Hip Ratio
WHO	World Health Organization
NCBI	National Center for Biotechnology Information
MEGA	Molecular Evolutionary Genetics Analysis

PSIPRED	PSI-blast based secondary structure PREDiction
PDB	Protein Data Bank
CDD	Conserved Domain Database
UCSF	University of California, San Francisco
SAMPDI	Single Amino acid Mutation binding free energy change of Protein-DNA Interaction

LIST OF SYMBOLS

%	Percentage
›	Greater than
‹	Less than
˚C	Degree Centigrade
µg	Microgram
µl	Microliter
U	Unit
USD	United States Dollar

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

LIST OF BAR DIAGRAM

ABSTRACT

Diabetes Mellitus (DM) is a common metabolic disorder, resulting from defects in insulin secretion, insulin action, or both. But the molecular mechanism behind the basic defects of DM is not yet fully understood. Defect in insulin secretion is the key factor in Type 2 Diabetes Mellitus (T2DM). The β-cell specific transcription factor, Insulin Promoter Factor -1 (IPF1) gene, is essential to pancreatic development and the maintenance of β-cell mass. Existing scientific evidence have explored the role of IPF1 gene in the development of MODY & T2DM but contradictions also exist in different population. Therefore, in this study, we aimed to investigate whether the polymorphisms of the IPF1 gene's C18R (T/C) and R197H (G/A) variants interplays the pathogenesis of T2DM in the Bangladeshi population.

A total number of 445 subjects were recruited in this study to screen genetic variants in the coding region. Among the subjects, 245 were healthy controls and 200 were T2DM patients. Genomic DNA was extracted from whole blood using the QIAGEN Genomic DNA Extraction Kit. Polymorphism of C18R (T/C) & R197H (G/A) variants of the IPF1 gene, were investigated using the PCR-RFLP method.

Among the study population BMI, WHR and serum glucose level were significantly higher in the diabetic group compared to the non-diabetic group. Of the total, 200 T2DM subjects only 1 C18R (T/C) Single Nucleotide Polymorphism (SNP) was found and in case of R197H (G/A) variant this number was 7, which shows 0.5% mutation found in C18R (T/C) and 3.5% in R197H (G/A) variant of IPF1 gene, in the T2DM subjects.

In the computational study, IPF1 gene shows both of the mutation C18R and R197H belong to the transactivation domain and DNA binding conserved homeodomain, respectively. Several studies had shown, Arginine (Arg) mutation to Cysteine (Cis) residue or vice versa close to the NH_2-terminal and IPF-1's first proline-rich domains has less serious or less functionality impacts. Secondary structure prediction and 3D modeling of IPF1 showed the R197H mutation located within the DNA binding domain but not belong to the residues those have DNA bases contact. Mutating Argine (Arg) to Histidine (His) residue in the 3D model showed there were no clashes/and contacts with the neighboring residues. Since it might accommodate properly within this area and could be the reason behind the negligible binding free energy change due to mutation.

Very little frequency of C18R and R197H mutations were found from the molecular analysis and the computational study also showed, both of these variants are not much influencing, this might be due to they do not contribute to make significant changes in IPF1 protein structure and functionality. From our study, it can be suggested that C18R and R197H mutations of the IPF1 gene, cannot be consider as an influencing factor to develop T2DM in Bangladeshi population.

Chapter One:
Introduction

Investigation of Insulin promoter factor 1 gene C18R and R197H variants association with type 2 diabetes mellitus in Bangladeshi Population

1. Introduction

Diabetes mellitus is a group of metabolic diseases that are characterized by hyperglycemia caused by insulin secretion abnormalities, insulin action, or both. The chronic hyperglycemia is correlated with long-term damage, dysfunction and failure of different organs, particularly the eyes, kidneys, nerves, heart, and blood vessels (ADA, 2014) and increases the risk of premature death (WHO, 2016). Several pathogenic processes are involved in the development of diabetes mellitus. These range from autoimmune destruction of the pancreatic β-cells with consequent insulin deficiency to abnormalities that result in resistance to insulin action. The basis of the abnormalities in carbohydrate, fat, and protein metabolism in diabetes mellitus is deficient action of insulin on target tissues. Deficient insulin action results from inadequate insulin secretion and/or diminished tissue responses to insulin at one or more points in the complex pathway. So hormone action impairment of insulin secretion and defects in insulin action frequently coexist in the same patient, and it is often unclear which abnormality, if either alone, is the primary cause of the hyperglycemia. In addition, diabetes mellitus is a factor that contributes to several other morbidity and mortality causes. It also raises the danger of end-stage kidney disease, heart illness, stroke, and other vascular illnesses.

According to World Health Organization (WHO) (2016), "Diabetes mellitus is a serious, chronic disease that occurs either when the pancreas does not produce enough insulin (a hormone that regulates blood sugar, or glucose), or when the body cannot effectively use the insulin it produces." The possible complications associated with this disease are heart attack, stroke, kidney failure, leg amputation, and vision loss and nerve damage. Uncontrolled diabetes in time of pregnancy increases the risk of fetal death and other complications (WHO, 2016).

Diabetes mellitus may present with some distinguishable symptoms like polydipsia (increased thirst), polyuria (frequent urination), polyphagia (increased hunger), and weight loss, blurry eyesight, wound heals slowly as well as an infection. In its most severe forms, ketoacidosis or a non-ketotic hyperosmolar state may develop and lead to daze, coma and in case of ineffective treatment, death may occur. Long term effects of this disease include the progressive development of particular complications of retinopathy with potential blindness, nephropathy which may lead to renal failure as well as the risk of foot ulcers, amputation, Charcot joints, features of autonomic dysfunction (WHO, 1999). Several pathogenic processes are involved in diabetes development including pancreas β-cells destruction that results in insulin deficiency.

1.1. Prevalence

1.1.1. Global prevalence of diabetes mellitus

Globally, diabetes mellitus (DM) is the most common endocrine disorder. Certainly, it is a global health catastrophe of the 21st century. In 2016, diabetes was the direct cause of 1.6 million deaths (WHO, 2018). There are 415 million people with diabetes and 5.0 million deaths due to diabetes estimated in 2015. Among them, 75% live in low and middle-income countries. If this trend continues, the number of people with diabetes will rise to 642 million by 2040 and one in 10 adults will have diabetes (Ogurtsova *et al*, 2017). Moreover, 318 million with impaired glucose tolerance (IGT) has estimated, who have a high risk of developing the disease later on stated in the IDF report in 2015. It has further estimated that 193 million adults having diabetes are yet undiagnosed and so at risk of developing diabetes-oriented complications (IDF, 2015). Recent global feature of DM is summarized in table 1.

In developed countries, up to 91% of adults have type 2 diabetes mellitus (T2DM) (IDF, 2015). The overall prevalence of diabetes varies between 15 – 20 %. The highest prevalence of T2DM (50 %) have been found in Pima Indians (49.4 % in male and 51.1 % in female) in the USA, Nauru (41%) and Australian aboriginal communities (approximately 20%), moderate in UK (1-2 %) and central and eastern Europe (2-4 %) and very low (0 – 1.4 %; 0% in male and 1.4 % in female) was observed among the Mapuches population in Chile and the prevalence was almost nill in rural and peri- rural population of Papua New Guinea (King & Rewers, 1993; WHO, 1994).

More than half a million children develop type 1 diabetes each year which is increasing around 3% every year. One in 7 births is affected in gestational diabetes (IDF, 2015).

It has had shown that about two-third of the rise in total diabetic patients would be constituted from those of the developing countries particularly China, India, Indonesia, Pakistan and Bangladesh (Wild *et al.*, 2004).

Table 1: Estimated global diabetes and impaired glucose tolerance prevalence, mortality for 2015 and 2040. NC = not counted.

	2015	**2040**
General population		
Total world population	7.3 billion	9.0 billion
Adult population (20-79)	4.7 billion	6.2 billion
Diabetes (20-79)		
Global prevalence	8.8 %	10.4 %
Number of people with diabetes	415 million	642 million
Number of deaths due to diabetes	5.0 million	NC
Impaired glucose tolerance (20-79)		
Global prevalence	6.7 %	7.8 %
Number of people with impaired glucose tolerance	318 million	481 million

1.1.2. Prevalence of diabetes mellitus in Bangladesh

Diabetes mellitus has become a global emergency for the world. The Bangladesh scenario is not diverse from this. This country also facing rapid urbanization and so massive changes occur in people's lifestyles. That is why diabetes has becomes an epidemic here. Bangladesh has 7.2 million adults having diabetes and so present in the 10^{th} position of the top ten countries for a number of adults having diabetes (IDF, 2015). In the South Asian region, Bangladesh has a 6.31% prevalence of diabetes among adults (IDF, 2013). In 2011, which had a population of 149.8 million, recent data showed that the prevalence of diabetes among adults is projected to rise from 8.4 million in 2011 to 16.8 million by 2030 (IDF, 2011). In the South Asian region over the past 30 years, the increase in prevalence is rising exponentially. Almost 65% of the registered diabetic subjects of the BIRDEM are from urban areas (Sayeed *et al.,* 2007) whereas the rural population constitutes >85% of the country's total population. It indicates that the number of total diabetic patients will be increased if properly diagnosed in rural areas.

According to the Bangladesh Urban Health Survey 2006 suggests that in an urban area the prevalence is just double (10%) than the rural area (5%). In 2010, the IDF has reported that 5.7 million (6.1%) and 6.7 million (7.1%) of Bangladeshi population are suffering from diabetes and impaired glucose tolerance (IGT) respectively. By 2030, that number of diabetic population is expected to get higher to 11.1 million. This explosion in diabetes prevalence will put Bangladesh among the top seven countries in terms of the number of people living with diabetes in 2030 (IDF, 2012). According to WHO (2011) diabetes mellitus deaths in Bangladesh have reached 2.05% of total deaths. The death rate is 23.80 per 100,000 of population, which ranked Bangladesh in 109th position in the world (World health rankings, 2010).

1.2. Etiology of Diabetes Mellitus (DM)

Type 1 diabetes mellitus results in autoimmune destruction of the β cells of the pancreas. This process occurs in genetically susceptible people and is (presumably) triggered by an environmental factor or factors (Skyler & Ricordi, 2011). Signs and symptoms of type 1 diabetes will be evident when the majorities (approximately 80%-90%) of the β cells have been destroyed and a normal level of serum glucose cannot be maintained. It is also caused by an interaction of genetics and environment.

Type 2 diabetes is characterized by insulin resistance and a progressive decline in pancreatic β cell insulin production. There is no autoimmune-mediated pancreatic β cell damage and most patients with type 2 diabetes do not need insulin during the initial stages of the disease.

Insulin resistance is such a condition in which insulin is produced, but cannot use properly; i.e., a given amount of insulin does not produce the expected result. The chronic inflammation associated with obesity affects the function of the insulin receptors on the cells in the liver, muscles, etc.,

decreases the number of insulin receptors, affects insulin signaling pathways, or inactivates insulin receptors (Allende-Vigo *et al.*, 2010; Olatunbosun & Dagogo-Jack, 2011).

The progressive decline in pancreatic β cell function is due to decreased β cell mass caused by apoptosis (Butler *et al.*, 2003); this may be a consequence of aging, genetic susceptibility, and insulin resistance itself (Unger & Scherer, 2010). The etiology of type 2 diabetes is complex and involved with genetic and lifestyle factors.

1.3. Types of Diabetes Mellitus

Diabetes mellitus is a heterogeneous gathering of clutters described by persevering hyperglycemia. The two most basic types of diabetes are type-1 diabetes mellitus (T1D, already known as Insulin Dependent Diabetes Mellitus or IDDM) and type-2 diabetes mellitus (T2D, previously known as Non-Insulin Dependent Diabetes Mellitus or NIDDM). And another type is gestational diabetes. All are brought on by a mix of hereditary and ecological danger components. Type-1 diabetes mellitus (insulin-subordinate) obliges insulin to treat, is ordinarily formed as a kid or youthful grown-up, and is an ailment that devastates pancreatic cells meaning no insulin generation is conceivable. Type-2 diabetes mellitus (non-insulin subordinate diabetes) is significantly more normal and ordinarily influences individuals beyond 45 years old, who are likewise overweight. Those distress from type-2 cannot deliver enough insulin, and sugar develops in the blood stream.

Diabetes can happen briefly along with pregnancy, and reports recommend that it happens in 2% to 10% of all pregnancies (WHO, 2016). Noteworthy, hormonal changes along with pregnancy can prompt glucose height in hereditarily inclined people. Glucose rise along with pregnancy is called Gestational Diabetes Mellitus (GDM). It can lead to a serious health risk for both mother and child. Patients with GDM are normally requested that experience an oral glucose tolerance test around six weeks in the wake of bringing forth figure out whether their diabetes has endured past the pregnancy, then again if any confirmation, (for example, weakened glucose resistance) is available that might be a piece of information to the patient's future danger for creating diabetes. It is currently settled that gestational diabetes is associated with an increased risk of type 2 diabetes mellitus later in life for both mother and child. In any case, other uncommon types of diabetes are specifically acquired. These incorporate maturity-onset diabetes in the young (MODY), and diabetes because of transformations in mitochondrial DNA.

Table 2: Etiological classification of diabetes mellitus (ADA, 2014)

i. Type 1 diabetes (β-cell destruction, usually leading to absolute insulin deficiency)
   - Immune mediated
   - Idiopathic

ii. Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with insulin resistance)

iii. Gestational diabetes mellitus

iv. Other specific types
   - Neonatal diabetes
   - Genetic defects of β-cell function
   - Genetic defects in insulin action
   - Diseases of the exocrine pancreas
   - Endrocrinopathies
   - Drug or chemical induced
   - Infection
   - Uncommon forms of immune-mediated diabetes
   - Other genetic syndromes sometimes associated with diabetes

1.3.1. Type 1 Diabetes Mellitus (T1DM)

Type 1 Diabetes Mellitus (earlier known as insulin-dependent, juvenile or childhood-onset diabetes) is characterized by deficient insulin production in the body. T1DM, which accounts for only 5–10% of those with diabetes, results from a cellular mediated autoimmune destruction of the β-cells of the pancreas (IDF, 2015). T1DM is a complex disease involving a combination of factors, such as genetic susceptibility, immunological dysregulation and environmental triggers (Acharjee *et al.*; 2013). It has a risk of 0.4% by age 30 for the general population and about 6% for siblings (Tuomilehto, 2013). T1DM is sub-grouped based on the type of damage of β-cells, namely immune-mediated type1 and non-immune mediated type1 (idiopathic type 1, T1DM) (ADA, 1997; WHO 1999). This autoimmune T1DM is noticed by immune-mediated damage, targeted against self-antigens, destroying the β-cells of the pancreas. Numbers of autoantibodies have been identified against different islet proteins. The rate of β-cell destruction is observed variably wide. For example, it is faster in children and slower in adults (ADA, 2014).

Some T1DM patients are ketosis-prone but lack of classical autoantibodies directed to islets and/or islet cell proteins, which are called idiopathic T1DM. This type of diabetes seems to be most common

in African and Asian countries (Tanaka *et al.*, 2000). Idiopathic T1DM usually happens to obese teens. The exact etiological factor(s) of idiopathic type 1, T1DM are still mysterious; multiple genetic and environmental factors are considered to be involved here (ADA, 1997 & WHO, 1999).

T1DM patients require insulin to maintain their blood glucose level daily. Symptoms of T1DM encompass excessive urination and thirst, constant hunger, weight loss, blurry vision, and fatigue. Some other autoimmune disorders such as Graves' illness, Hashimoto's thyroiditis, Addison's disease, Vitiligo, Celiac Sprue, Autoimmune hepatitis, Myasthenia gravis, and Pernicious anemia are also prone to patients with type-1 diabetes mellitus (ADA, 2014).

1.3.2. Type 2 Diabetes Mellitus (T2DM)

Type 2 diabetes mellitus (T2DM) is more complex than type-1 diabetes mellitus (T1DM). It is estimated that 90–95% of diabetic patients suffering from T2DM previously referred to as non–insulin dependent diabetes or adult-onset diabetes encompasses individuals who have insulin resistance and usually have relative (rather than absolute) insulin deficiency. At the primary stage and often during their lifetime, these individuals do not need insulin treatment to survive. This form of diabetes has many different causes yet the specific etiologies are unknown (ADA, 2014).

T2DM remains undiagnosed for ages due to hyperglycemia is usually not severe enough to provoke noticeable symptoms, unlike T1DM. Significant pathological and functional changes occur for hyperglycemia which can cause organ damage before the diagnosis of T2DM is made (Sanghera & Blackett, 2012). Hyperglycemia in T2DM is an effect of a complex interplay between insulin resistance (sensitivity) and abnormal insulin secretion (DeFronzo, 2004) and initially T2DM is characterized by compensatory insulin secretion associated with insulin resistance (Sanghera & Blackett, 2012).

Long term effects of diabetes along with micro-vascular and macro-vascular complications having genetic susceptibility are at greater risk (Doria, 2010). The progressive development of micro-vascular complications hampers blood supply in capillaries to kidneys and retina cause blindness. On the other hand, macro-vascular complications include hypertension, coronary artery disease, peripheral vascular disease, cerebral vascular disease, and hyperlipidemia. Neuropathic diabetic patients may suffer from foot ulcers, amputations, sexual dysfunction, and non-healing skin wounds. Staphylococcal sepsis infection is more common in diabetics as well as infections of the ear, nose, throat as well as reactivation of tuberculosis associated with the high rate of mortality and morbidity are more likely with poor blood glucose control (Sanghera & Blackett, 2012).

1.3.3. Gestational Diabetes Mellitus (GDM)

Gestational diabetes mellitus is a type of diabetes that arises during pregnancy (usually during the second or third trimester) which is a temporary condition that occurs in pregnancy and carries long term risk of type 2 diabetes. The condition is determined when the blood glucose level is above

normal but still below those diagnostic of diabetes. GDM patients are at increased risk of some complications during pregnancy and delivery, as are their infants (WHO, 2014). In case of some women, it occurs when they cannot produce enough insulin to meet the extra needs of pregnancy. In other women, it may be sensed during the first trimester of pregnancy, and in these women, the condition most likely prevailed before the pregnancy (Claire *et al.*, 2008). It affects up to 7% of all pregnancies, which resulting in more than 200,000 cases annually (ADA, 2014). Overweight or obese women are at a higher risk of gestational diabetes. The lifetime risk of developing T2DM after gestational diabetes is up to 60% (Noctor & Dunne, 2015). About 35% to 60% of women with GDM will inevitably mature type-2 diabetes throughout the following 10 to 20 years. However, the prevalence of GDM is 1% – 4 % of all pregnancies, depending on the population studied and the diagnostic tests employed (ADA, 2003).

1.3.4. Neonatal Diabetes

Diabetes diagnosis in the first six months of life is not always typical autoimmune type1diabetes perhaps called neonatal diabetes. It can be transient or permanent. Defect on ZAC/HYAMI imprinting resulting transient disease, whereas permanent neonatal diabetes is most commonly defect on a gene coading Kir6.2 subunit of the β-cell KATP channel (ADA, 2014).

1.3.5. Monogenic β-cell diabetes

Several forms of diabetes are associated with monogenetic defects in β-cell function or insulin action. This form of diabetes is characterized by onset hyperglycemia at an early age before 25. This form of hyperglycemia/ diabetes is referred to as Maturity–onset Diabetes of young (MODY) (ADA, 2014). MODY is characterized by a slow onset of symptoms, in the absence of obesity, no ketosis, and no evidence of autoimmunity of β-cell. It is most often managed without the need for exogenous insulin (Stride & Hattersley, 2002). According to estimation 1-2% of total diabetic patients in Europe having, MODY but often misdiagnosed as type 1 or type 2 diabetes (Frayling *et al.*, 2001).

Monogenic β-cell diabetes, a rare form of diabetes mellitus (DM), is caused by defects in a group of genes controlling pancreatic β-cell development and function (Sujjitjoon *et al.*, 2008). In the earlier classification of diabetes, MODY relays into juvenile-onset and maturity-onset diabetes. Recently, the group of "Genetic defect in β-cell function" has included MODY as its part with its subclassification according to the gene involved (Vaxillaire & Froguel, 2008). MODY is an ancient form of early-onset type 2 diabetes. It is a monogenic form of diabetes mellitus inherited in autosomal dominant mode, primarily an outcome of impaired β-Cells of the pancreas (Frayling *et al.*, 2001). Again, MODY represents genetic, metabolic, and clinical heterogeneity. Generally, it develops in middle age, and mainly coupled with the primary scantiness of insulin secretion (Vaxillaire & Froguel, 2008).

MODY is characterized by impaired insulin secretion with minimal or no defects in insulin action. Responsible genes for MODY are inherited in an autosomal dominant pattern. Up to now, anomalies

have been recognized in six genetic loci on different chromosomes. Mutations on chromosome 12 in a hepatic transcription factor called the hepatocyte nuclear factor (HNF)-Iα, Glucokinase (GCK) gene on chromosome 7p, are connected with the most prevalent type. Less common forms result from mutations in other transcription factors including HNF-Iα, HNF-Iβ, Insulin Promoting Factor (IPF)-1 and NeuroD1. Recently discrete subtypes of MODY are classified as MODY 1 to MODY 6 according to the defect of genes.

Table 3: Genes in which mutations cause different types of MODY (Ellard *et al.*, 2008).

	Gene Symbol (other Symbol)					
	HNF-4α	**GCK**	**HNF-1α (TCF1)**	**PDX1(IPF 1)**	**HNF-1β (TNF2)**	**NeuroD1**
Types of MODY	MODY 1	MODY 2	MODY 3	MODY 4	MODY 5	MODY 6
Protein	HNF-4 alpha	Glucokinase	HNF-1 alpha	IPF- 1	HNF-1 beta	NEUROD -1
Chromosome locus	20q13.12	7p13	12q24.31	13q12.2	17q12	2q31.3
Gene Accession no.	NM_000457 .3[a]	NM_000162. 2	NM_00054 5.4	NM_00020 9.2	NM_0004 58.1	NM_0025 00.2
Mutation frequency (%) (not known in ~ 20% of cases)	~5	20-50	20-50	<1	~5	<1

1.4. Prevalence of type 2 diabetes mellitus

Type 2 diabetes mellitus (T2DM) has become progressively more frequent in obese young people in the last two decades (Pinhas-Hamiel & Zeitler, 2005). In the SEARCH study (Dabelea *et al.*, 2007) the incidence rate (per 100,000 person-year) of T2DM among children and youth varies greatly by ethnicity, with the highest rates observed among youths aged 15–19 years in minority populations. In particular, the reported incidence rate is 49.4% for Native Americans, 22.7% for Asian/Pacific Islanders, 19.4% for African Americans, 17% for Hispanics, and 5.6% for non-Hispanic whites.

Type-2 diabetes in youth is not just an American phenomenon rather it is worldwide. For instance, in Japan (Pinhas-Hamiel & Zeitler, 2005) 80% of all new cases of diabetes in children and adolescents are diagnosed as T2DM patients. Similarly, in Taiwan, 54.2% of new cases are diagnosed with

T2DM, with an incidence of 6.5 per 100,000. In contrast, in the U.K. the minimum incidence of T2DM in children (<17 years of age) is 0.53 per 100,000 in 1 year (Haines *et al.*, 2007). In Austria, the calculated incidence in children and young people (<15 years of age) is 0.25/100,000 (Pinhas-Hamiel & Zeitler, 2005). Certainly, many studies from Europe (Dabelea *et al.*, 2007) indicate that type-2 diabetes mellitus is not as common as in the U.S. in these populations, accounting for only 1–2% of all diabetes mellitus cases.

Also, some studies support the notion that T2DM has a greater prevalence in the high-risk ethnic groups. It accounts for 14.9% of all diabetes cases among non-Hispanic white adolescents (Dabelea *et al.*, 2007).

1.5. Pathogenesis of type 2 diabetes mellitus

The maintenance of normal glucose level relies on a precisely balanced and dynamic interaction between tissue sensitivity to insulin and insulin secretion (Röder *et al.*, 2016). It is recommended T2DM results from defects in secretion or/and action of insulin. However, both of these defects can have a genetic as well as an acquired component (Röder *et al.*, 2016; Natali *et al*., 1998).

1.5.1. Impaired insulin secretion and insulin resistance

Insulin is synthesized and produced by pancreatic β-cells. It is a peptide hormone that regulates glucose metabolism. Hyperglycemia observed when an obstruction occurs in insulin biosynthesis and/or functioning process. And so higher blood glucose concentration find than in normal subjects in T2DM patients.

Pancreatic β-cells secrete insulin in a biphasic manner. Type 2 diabetes mellitus (T2DM) is characterized by defects in insulin secretion and insulin action (Tripathy & Chavez, 2010). Defect in insulin secretion assumes to mean lower insulin secretion. The first phase of insulin secretion is determined by complex multiple signaling pathways including hypoxia-inducible factor 1α, von Hippel-Lindau, factor inhibiting HIF, nicotinamide phospho-ribosyl-transferase and many other crucial regulatory factors (Cheng *et al.*, 2013). An independent predictor of T2DM onset develops due to the loss of the first phase in insulin secretion. In T2DM patients, the restoration of this phase suppresses the hepatic glucose output and improves the blood glucose concentration (Almgren *et al.*, 2011). Furthermore, the earliest detectable defect and exhaustion of β- cells function appeared in a reduction of the first-phase insulin secretion. The second phase of insulin secretion involves the release of newly synthesized insulin in response to glucose. The reductions in both phases occur equally early and precede the development of insulin resistance (IR) (Szoke & Gerich, 2005).

Impaired insulin secretion is associated with β-cell dysfunction that results in a reduced insulin secretion response to rises in blood glucose after eating (Kahn, 2003). The insulin sensitivity (IS) is

often marked by the fasting plasma insulin concentration and the changing in acute insulin secretion following glucose test is an early observation of progressive pancreatic defect.

The contributing factors of continuous decline in pancreatic function are Glucotoxicity and lipotoxicity (DeFronzo, 2009; Del Prato *et al.*, 2002). Thus insulin secretion is already defective at the time of diagnosis and declines steadily thereafter. Other β- cell dysfunction feature is defective cleavage of the insulin precursor, leading to variably increased concentrations of proinsulin and its early 'split product' derivatives (Kakehi *et al.*, 1988). IR and progressive β-cells dysfunction are the two important defects in the development of the disease (Nijpels, 1998). IR believed as the primary features in the pathogenesis of T2DM and conceivably results from a genetically determined reduction in insulin sensitivity, compounded by exposure to the environmental factors. T2DM patients inherit genes from parents that make their tissues resistant to insulin (DeFronzo, 2009). Insulin resistance is defined as the increased demand for insulin to maintain normoglycemia because of decreased action of insulin in specific tissues (Pillay *et al.*, 1995). Defects in the following three main steps are involved in the generation of insulin resistance: i) insulin binding to the cell membrane receptor; ii) Insulin receptor phosphorylation and iii) intracellular insulin signaling. Insulin resistance (IR) in muscle, liver, peripheral tissue, skeletal muscle, fat as well as β-cell failure, represent the core pathophysiologic defects in the development of T2DM (DeFronzo, 2009). In muscle, IR is due to decreased glucose uptake and impaired utilization of glucose by non-oxidative pathway, primarily glycogen formation as well as a slight decrease in glucose oxidation. In the liver, IR leads to the failure of insulin to suppress hepatic glucose production, which is fuelled by glycogen breakdown and particularly by gluconeogenesis (Röder *et al.*, 2016). The majority of patients with insulin resistance are obese and obesity itself causes or aggravates insulin resistance (Campbell *et al.*, 1993; Bogardus *et al.*, 1985). Patients with this form of diabetes may have an insulin level that appears normal or elevated, the high blood glucose levels in these diabetic patients would be expected to result in even higher insulin values had their β- cell function been normal (Bell & Polonsky, 2001).

The pathophysiology of T2DM varies with tissues or organs as detailed below. The contribution of each tissue or organ to T2DM is summarized in figure 1.

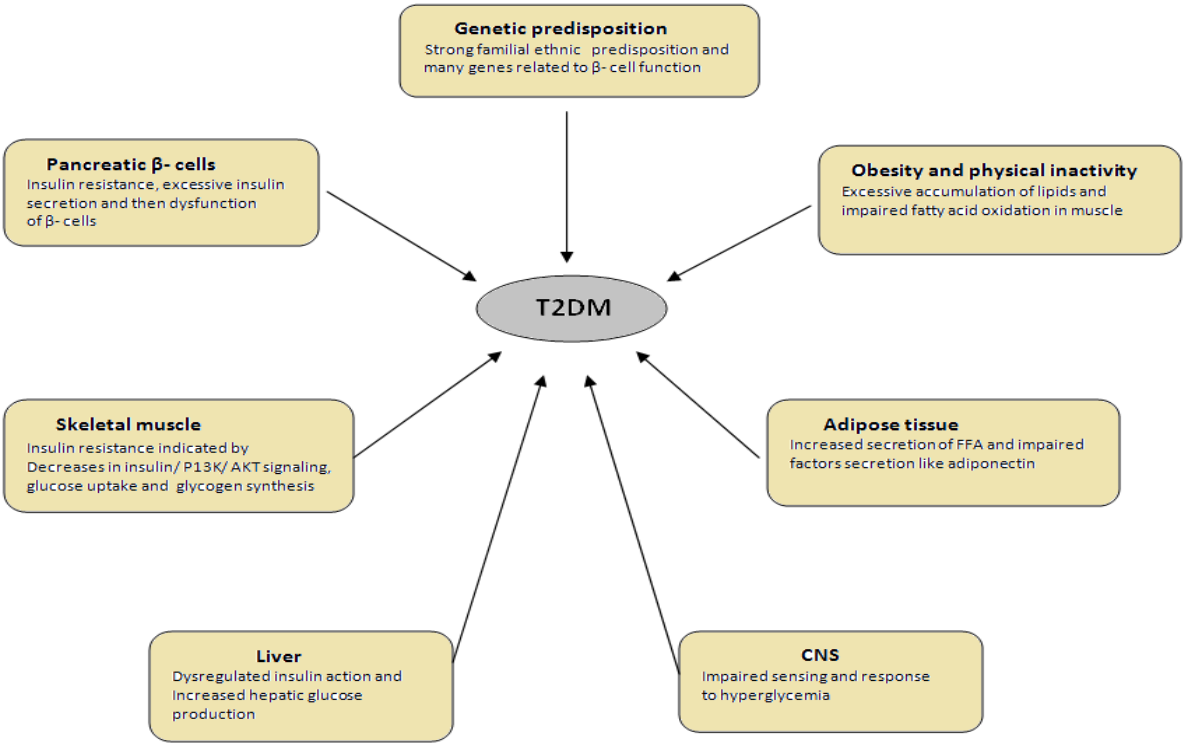


Figure 1: Role of genetic predisposition & organs in pathogenesis of T2DM (Lin & Sun, 2010).

The glycemic progression from normal to impaired glucose tolerance (IGT) mainly due to IR and from this state to DM is obvious. Extrapolation of the observed rate of β-cell decline observed in the UKIPDS study suggests that loss of β-cell function begins some 10 – 12 years before diabetes is diagnosed (Turner *et al.*, 1998). San Antonio and Hoorn Study projected the high fasting pro-insulin reflects the β-cell dysfunction and the progression to DM. The most effective calculation for the development of T2DM is two hours blood glucose level test followed by oral glucose load (Nijpels, 1998).

In glucose metabolism impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) is the intermediate states that exist between normal glucose tolerance (GT) and obvious diabetes which is also associated with the reduction in early-phase insulin secretion. However, patients with IGT also have impaired late-phase insulin secretion. The latter group has demonstrated a significantly higher muscle IR and low hepatic IR while patients with IFG have severe hepatic IR with normal or near-normal muscle IS (Abdul-Ghani *et al.*, 2006). A prospective study among Pima Indians suggests that passage from normal to impaired glucose tolerance (IGT) and finally to T2DM that accompanied by a progressive decline in β-cell secretory capacity (Weyer *et al.*, 1999).

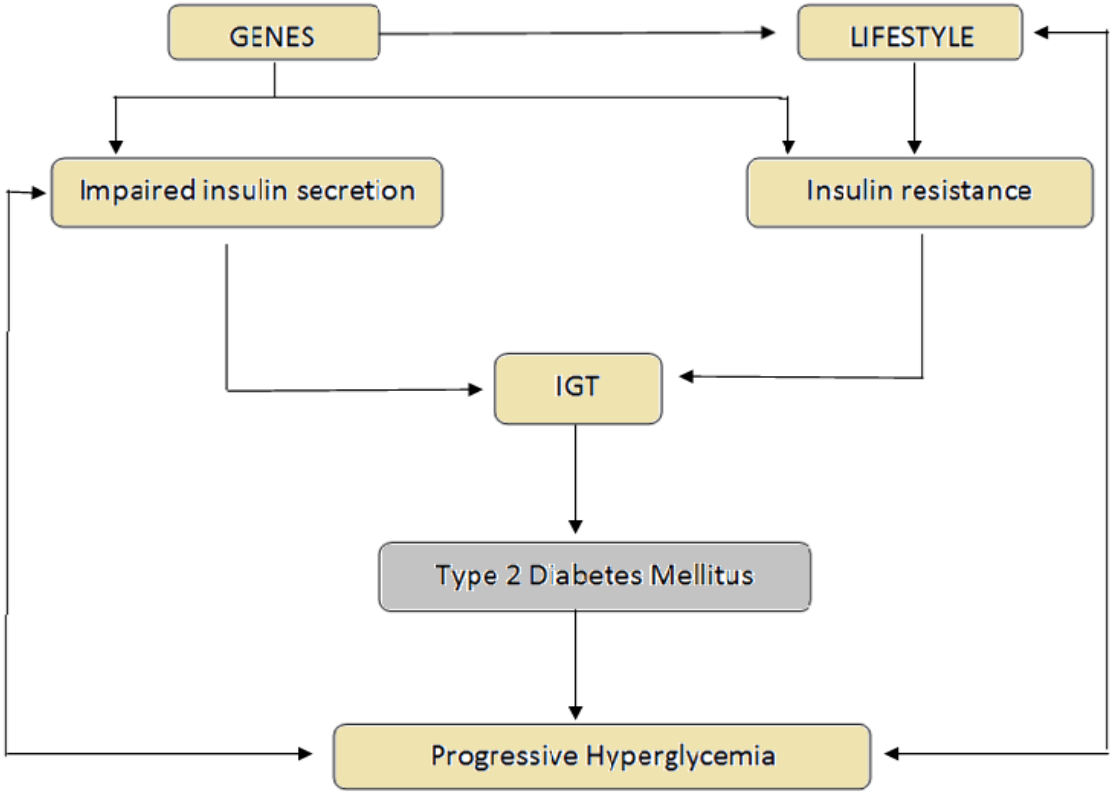


Figure 2: Pathogenesis of T2DM (Ozougwu *et al.*, 2013).

1.6. Prevalence of Monogenic β-cell diabetes or MODY

It is difficult to determine the actual prevalence of monogenic β-cell diabetes or MODY in the general population because its clinical phenotype is highly heterogeneous and sometimes misdiagnosed as other types of DM (Sujjitjoon *et al.*, 2008). According to estimation 1-2% of a total diabetic patient in Europe having, MODY but often misdiagnosed as type 1 or type 2 diabetes (Bulman *et al.*, 2001; Lenderman *et al.*, 1995). However, the prevalence of each MODY subtype also varies among ethnic groups. Nevertheless, there is some evidence that showed that 1-5% of total DM cases are MODY. MODY2 is the most common cause of MODY in France, responsible for more than 60% of studied families, whereas its prevalence in the United Kingdom and Germany are 11% and 8%, respectively. In general, MODY3 is most common in Caucasians but its prevalence varies from 21% to 64%. The other four types of MODY are rare. MODY with unknown genetic etiology (MODY-X) represents 16-45% of MODY cases in Caucasians and more than 90% of the cases in Asian populations (Sujjitjoon *et al.*, 2008). Then again, Johansen *et al.* (2005) findings suggest a relative prevalence of 3% of MODY1 (2 different mutations in 2 families), 10% of MODY2 (7 in 8), and 36% of MODY3 (21 in 28) among Danish kindred clinically diagnosed as MODY. Pinterove *et al.* (2007) screens the GCK gene that is known to cause MODY 2, 29% of patients are found in the Czech population (Attiya & Sahar, 2012).

1.6.1. Pathogenesis of MODY

Several forms of diabetes are associated with monogenetic defects in β-cell function, which has been misdiagnosed as type 1 or type 2 DM, mostly (ADA, 2014). The diagnostic criteria of MODY are as follows (Hattersley, 2003): (i) early onset of diabetes (usually less than 25 years), (ii) autosomal dominant inheritance, (iii) rarely obese and non-ketotic diabetes, and (iv) diabetes results from β-cell dysfunction. MODY is usually of insidious origin and typically diagnosed on routine examinations. MODY patients show reduced secretion of insulin in response to glucose load (O'Rahilly *et al.*, 1988). However, hyperinsulinemia has been described in a few families, although there was very little pathophysiological evidence of insulin resistance. The longitudinal studies suggested that, in general, insulin response declines with increasing duration of the disease (Fajans & Vinik, 1989).

Molecular genetics of MODY has been more progressively studied than other forms of DM. Advanced genetic linkage analysis through the autosomal dominant model of inheritance displays the ability to collect large multi-generation pedigrees of MODY causing genes (Sujjitjoon *et al.*, 2008). These genetic studies have defined several subtypes of MODY. Eleven genes are discovered contributing to MODY in the European population (Attiya & Sahar, 2012). Mutations in the genes encoding hepatocyte nuclear factor 4 (HNF4), glucokinase (GCK), hepatocyte nuclear factor 1 alpha (HNF1α) and hepatocyte nuclear factor 1 beta (HNF1β), insulin promoter factor 1 (IPF1) and NEUROD 1 are the causes of six known forms of MODY. Mutations in these genes may disrupt the development of β-cell in the embryo and follow-on dysfunctional β-cells in adult age, resulting in hyperglycemia.

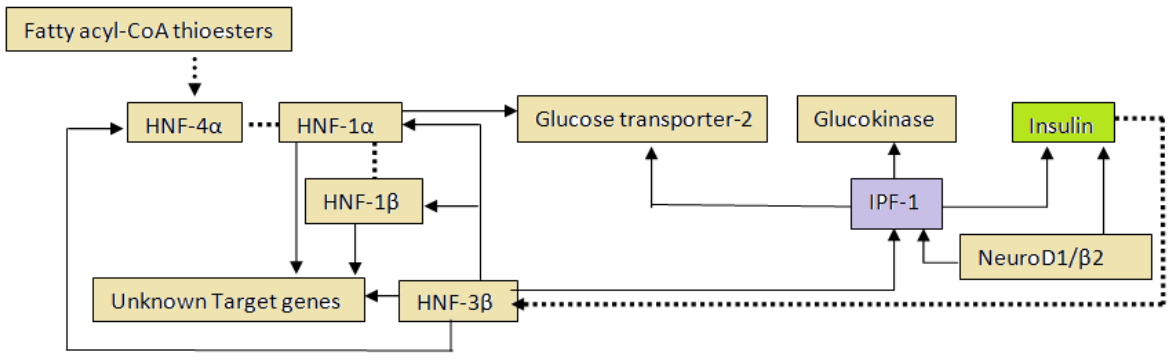


Figure 3: Network of DNA- binding factors involved in transcription in the pancreatic genes. Where IPF-1 is the insulin promoter factor in insulin transcription process. HNF1A and NEUROD1 is a β- cell transcription factor (So *et al.*, 2000)**.**

1.7. Risk factors for T2DM

Risk factors for diabetes rely on the type of diabetes. The pathogenesis of T2DM suggested that- (i) lifestyle, birth weight, obesity, physical inactivity, pregnancy, age, sex, etc. and (ii) genetics/

heritability; resulting in alterations in DNA sequencing. Both these genetic and environmental factors interact and play important roles in the development of T2DM. Gene-environment interactions can be defined as genetic effects on a disease that differ in level and sometimes direction across environmental contexts. By changing lifestyle it is possible to reduce the risk of diabetes at any level of genetic risk (Hivert *et al.*, 2011).

1.7.1. Environmental factors

Numbers of environmental factors involved with the pathogenesis of type 2 varieties, which include ethnicity, age, diet, and viral infection and stress (Knip & Akerblom, 1999; Dahlquist, 1998). Genetic factors of T2DM are mostly influenced by these environmental factors. Presently, developed countries' population are facing less T2DM than developing by controlling the responsible environmental facts like obesity, diet, etc.

1.7.2. Ethnicity

Records in ethnic differences in the prevalence of type 2 diabetes are well mentioned. In 2004 in the U.S. it is demonstrated the prevalence of diagnosed diabetes is higher for blacks and Hispanics than for whites across all age-groups (Cossrow & Falkner, 2004). T2DM is up to six times more common in people of South Asian origin and up to three times among people of Africa and Africa-Caribbean origin (Oldroyd *et al.*, 2005). Diabetes is almost four times higher as prevalent in Bangladeshi men, and almost three times in Pakistani and Indian men compared with British men mentioned in the Health Survey for England 2004. On the other hand, among women, diabetes is more than five times higher among Pakistani women, at least three times as likely in Bangladeshi and Black Caribbean women as well as two and a half times as likely in Indian women, compared with women in the general population. Also, diabetes seems to be rare among the people aged 16 to 34 years but is the highest in Indian men (2%), Black African men (1.7%) and Irish women (1.7%).

1.7.3. Obesity

Substantial epidemiological studies have shown that obesity is the most important risk factor for T2DM, which may influence the development of insulin resistance and disease progression (Belkina *et al.*, 2013). As a result, the T2DM rate is increasing dramatically worldwide (Zimmet *et al.*, 2001). Nearly 90% of diabetic patients develop T2DM mostly relating to excess body (WHO, 2011). Several studies have indicated that Obstructive Sleep Apnea (OSA) in T2DM patients is much more prevalent (36%–60%) than in the general population (Schober *et al.*, 2011). Obstructive Sleep Apnea (OSA) is the most common treatable reason for overweight and obese adults which help to develop prediabetes condition (20%-67%) and T2DM (15%-30%) (Pamidi *et al.*, 2012; Ioja *et al.*, 2017). It has been estimated that 85% of the T2DM children are either overweight or obese among Pima Indians (Fagot-Campagna *et al.*, 2000). While in men 64% cases of diabetes and 74% in women are reported due to overweight and obesity (BMI>25 kg/m2) in lean subjects (Chan *et al.*, 1994).

1.7.4. Physical inactivity

Physical inactivity is another major T2DM risk factor that influences obesity and increases the risk of T2DM.Whereas physical activity helps to control weight, exercise improves glucose and lipid metabolism that decreases the T2DM risk. It also inversely related to body mass index and imparted glucose tolerant (IGT). Several intervention studies in China (Pan *et al.*, 1997), Finland (Tuomilehto *et al.*, 2001) and the US (Diabetes Prevention Program Study Group, 2002) have shown that only healthy diet and exercise reduce the risk of progression of T2DM from IGT about 60%. On the other hand, the oral hypoglycemic medication reduces only 30% progression risks.

1.7.5. Diet

High-fat, low-carbohydrate diets impair insulin action (Bisschop *et al.*, 2001). Consuming excessive amounts of saturated fats has lead to obesity and thus increases the risk of developing insulin resistance condition in individuals (Lovejoy *et al.*, 2002).

1.7.6. Birth weight

Nutrition in uteri may be an important component of the susceptibility to T2DM. A hypothesized have suggested by Hales *et al.* in 1991 that "fetal and childhood malnutrition, by programming metabolism, predispose to certain diseases in adult life" including T2DM.

1.7.7. Age

Risk increases in older age. Because in older age one tends to exercise less, lose muscle mass and gain weight (http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/dxc-20169861). And so glucose tolerance or insulin sensitivity has been shown to decrease with age (Shimokata *et al.*, 1991).

1.7.8. Sex

The prevalence of type 2 diabetes in the first half of the last century was higher among women than men, but this consequence has shifted, so more men than women are now diagnosed with type 2 diabetes. This change is mainly caused by a more sedentary lifestyle particularly among men, resulting in increased obesity. That is how men develop diabetes at a lower degree of obesity than women. However, differences in sex in condition to body fat distribution, insulin resistance, sex hormones, and blood glucose levels support this concept (DIAPEDIA).

1.7.9. Polycystic ovary syndrome

Polycystic ovary syndrome (POSY) a common condition characterized by irregular menstrual periods and obesity increases the risk of T2DM (Gambineri *et al.*, 2012).

1.7.10. History of Gestational diabetes

In pregnancy period some women develop GDM they are at high risk of developing prediabetes and T2DM later. Additionally, if any women gave birth an overweight baby (> 9 pounds /4 kilograms), also at risk of T2DM (http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/dxc-20169861).

1.7.11. Stress

Stress is another important factor for developing T2DM. Slawson et al. (1963) have demonstrated 80% of a group of 25 adult diabetes patients gave a history of stress, 1-48 months before the onset of diabetes. These were mostly patients of depression. This metabolic stress may develop due to an unbalanced amount of free radical productions in diabetes by glucose autoxidation, polyol pathway and non-enzymatic glycation of proteins.

1.7.12. High blood pressure

Blood pressure over 140/90 millimeters of mercury (mm Hg) increase the risk of T2DM (http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/dxc-20169861).

1.7.13. Abnormal cholesterol and triglyceride levels

A low level of high-density lipoprotein (HDL) or good cholesterol increases the risk of T2DM. While a high level of triglycerides increases the risk of T2DM (http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/dxc-20169861).

1.7.14. Genetic factors

The inheritance of T2DM is polygenic which means the simultaneous presence of several abnormal genes or polymorphisms is necessary for the development of the disease (McCarthy *et al.*, 1994; Aitman & Todd, 1995). Several lines of evidence support the principle of inherited genetic susceptibility as an important risk factor for common T2DM. The offspring of a parent with T2DM face a 40% lifetime risk of developing T2DM, increasing to 70% when both parents have T2DM (Meigs *et al.*, 2000). Interestingly, the risk is higher if the mother, rather than father, is affected (Mahtani *et al.*, 1996). First degree relatives of individuals with T2DM are more about 3 times to develop the disease than individuals without a positive family history (Gloyn *et al.*, 2003). Concordance rates for monozygotic twins ranged from 60-90%, which are significantly higher than those for dizygotic twins. Therefore, it is clear that T2DM has a strong genetic component. T2DM risk is higher in certain ethnic groups. Recent genetic studies, especially genome-wide association studies (GWAS), have identified a huge amount of variants associated with T2DM (Dorajoo *et al.*, 2015) as well as associated metabolic traits, mostly in Europeans, and some in African, and South Asian populations (Sanghera & Blackett, 2012). Approximately 80 different genetic loci have been identified as predisposing to risks of T2DM in the European population. Twin and ancestral studies

point out a substantial heritable component to T2DM, estimated to be 40% to 80% (Kaprio *et al.*, 1992; Almgren *et al.*, 2011). However, T2DM is a multi-factorial disease that relates to a huge number of genetic variants to interact with lifestyle and environmental factors (Gene × Environment) (Hetherington & Cecil, 2010; Franks *et al.*, 2013). Linkage and candidate-gene studies have become successful to identify some rare inherited forms of T2DM presenting at young ages called maturity-onset diabetes of young (MODY), mitochondrial diabetes and neonatal diabetes. But this linkage and association studies on the common form of T2DM provided inconsistent results and failed replication in multiple populations (Barroso *et al.*, 2005). Only PPARG, KCNJ11, and TCF7L2 were identified as established genes associated with common forms of T2DM (Zeggini, 2007).

1.7.15. Familial aggregation

Families share environments, culture, and habits as well as genes. So, familial aggregation of the disease is another source of evidence for a genetic contribution. Evidence showed, nearly 4-fold increased risk for T2DM in siblings of a diabetic proband compared with the general population, the odds ratio (OR) of 3.4–3.5 with only a single affected parent, and the increase in the OR to 6.1 if both parents are affected (Meigs *et al.*, 2000). Several studies showed an increased prevalence of T2DM in the offsprings of parenting diabetes. It is reported that if neither parent has the disease the prevalence of T2DM is 10.4% in the siblings, but if one is affected the prevalence was 17.8% and it increased 25.2% when both parents are diabetic. Another study demonstrated the risk of diabetes is strongly connected with family history: 14.8% for those who have no family history of diabetes, 22% for those with one parent having diabetes and 41% in those with both parents having diabetes (Weijnen *et al.*, 2002). The lifetime risk of T2DM is about 7% in the general population whereas 40% in offspring of one parent having T2DM and about 70% if both parents are affected (Köbberling, 1982; Majithia & Florez, 2009).

1.7.16. Twin studies

Multiple studies of twin concordance rates have been undertaken in T2DM. Estimates for concordance rates have ranged from 0.29 to 1.00 in monozygotic (MZ) twins, while in dizygotic (DZ) twins the range was 0.10–0.43 (Poulsen *et al.*, 1999; Medici *et al.*, 1999). The high concordance in MZ twins and the 50% fall in DZ twins provide compelling evidence for a genetic component of T2DM (Poulsen et al., 1999). The risk of developing T2DM in siblings is 4-6 folds higher than those of unrelated individuals (Redondo *et al.*, 2001). In another study, the concordance rate in MZ twins is about 70% whereas the concordance in DZ twins has been observed to be only 20%-30% (Kaprio *et al.*, 1992). Comparatively, the concordance rate for T1DM among MZ twins found to be higher between 20-71% compared to that of among DZ twins (Kyvik *et al.*, 1995) suggesting a strong genetic role in its pathogenesis. Poulsen in 1999 also suggested the concordance rate in MZ twins was ranged from 34%-83% whereas it was 16%-40% in DZ twins.

1.7.17. Heritability of T2DM

Insulin sensitivity and insulin secretion deteriorate in parallel in most human T2DM. Both defects predicted subsequent T2DM in several studies and both defects are shown to be present in nondiabetic but genetically identical co-twins of a diabetic proband (Vaag *et al.*, 1995).

It is estimated that the heritability of the T2DM range is from 20%-80% and evidence comes from a variety of population, family and twin-based studies (Meigs *et al.*, 2000; Poulsen *et al.*, 1999). First degree relatives of individuals with T2DM are about 3 times more likely to develop the disease than individuals without a positive family history of the disease (Florez *et al.*, 2003).

1.8. Genome-Wide Association Studies (GWAS)

Completing the Human Genome Project in 2003 (Collins *et al.*, 2003) led to subsequent advancement in biomedical research. After then, a new technology 'genome-wide chips' in 2007 has facilitated remarkable progress in T2DM genetic research with the first publication of five large GWA scans showed more than 500,000 single nucleotide polymorphism (SNP) markers distributed across the genome (Zeggini, 2007; Sexena R et al., 2007; Sladek *et al.*, 2007). GWAS has revealed more than 70 completely new T2DM loci which suggest that the associations are not limited to candidate genes only (Petretto *et al.*, 2007) and exposed the best chance of detecting and identifying additional genes for T2DM. These studies are being pursued in different population groups, to provide a more realistic 'genetic risk landscape' of the disease (Genomics of Common Disease, 2007), and account for variation in population-specific environmental interactions. In connection with these studies, researchers have confirmed the polygenic nature of the disease and implicate a massive hit to β-cell function (insulin secretion) as opposed to those associated with insulin resistance (McCarthy & Zeggini, 2009; Bonnefond *et al.*, 2015).

GWAS identified most of the T2DM loci which are common variants with small effects. A French group of investigators published the first GWAS in T2DMthat has identified a zinc transporter and member of solute carrier family SLC30A8 and HHEX along with confirming the connection of TCF7L2 and KCNJ11 with T2DM (Sladek *et al.*, 2007). Other three further GWAS publications have confirmed TCF7L2, KCNJ11, PPARG, SLC30A8, and HHEX. This group of researchers also discovered novel T2DM loci called CDKAL1and a variant near CDKN2A-B (Saxena *et al.*, 2007; Welcome trust case-control consortium, 2007). Icelandic researchers group the decode also confirmed these findings of GWAS (Steinthorsdottir *et al.*, 2007). FTO gene's association with T2DMhas also first discovered by GWAS for obesity [OR for T2DM being 1.27, $p < 10^{-8}$] (Frayling *et al.*, 2007) and later confirmed in replication studies (Zeggini, 2007).

In addition, six novel loci, JAZF1, CDC123, TSPAN8, THADA, ADAMTS9, and NOTCH2, with strong evidence of association have detected through a joint meta-analysis of three previously conducted T2DM GWA scans by the Diabetes Genetics Replication And Meta-analysis (DIAGRAM)

consortium (Zeggini *et al.*, 2008). Another European study has reported a strong association of an intronic SNP (rs560887) in G6PC2 with fasting plasma glucose levels by screening 392,935 SNPs in 654 non-T2DM participants using GWAS (Bouatia-Naji *et al.*, 2008). The FUSION study replicates the association of G6PC2 with fasting plasma glucose levels too (Chen *et al.*, 2008). Further, two separate GWA studies have detected a variant near MTNR1B that increase FBG levels (Bouatia-Naji *et al.*, 2008; Lyssenko *et al.*, 2009).

To identify loci influencing glycemic traits, a meta-analysis was performed on 21 GWAS comprising 46,186 participants. The follow-up studies on 25 loci have performed in up to 76,558 additional participants in European populations and detected nine loci associated with fasting glucose and HOMA-B (ADCY5, MADD, ADRA2A, CRY2, FADS1, GLIS3, SLC2A2, PROX1 and C2CD4B) and one influencing fasting insulin and HOMA-IR (IGF1) (Dupuis *et al.*, 2010). Analysis of another study has combined GWAS data from 8,130 individuals with T2DM and 38,987 controls approach, their mutual effect may be significant (Petretto *et al.*, 2007).

GWAS has identified numbers of genes those are associated with β-cell failure in T2DM and only <5% are inherited. More than 60% of heritability appears clinically evident, environmental interactions, biochemical DNA modifications (epigenetic) and the combined effect of rare variants. On the other hand, other genes present potential challenges in determining risk association. The beginning of the GWAS era and the recognition of multiple risk loci have illuminated the pathophysiology of T2DM undoubtedly (Dorajoo *et al.*, 2015).

1.9. Commonly identified T2DM genes

These genes are identified by linkage study. Candidate genes are chosen due to their involvement in pancreatic β cell function, insulin action/glucose metabolism, or other metabolic conditions that increase T2DM risk (e.g., energy intake/expenditure, lipid metabolism). So far, more than 50 candidate genes for T2DM have been studied in different populations all over the world. However, results for essentially all candidate genes have been conflicting possibly for the different findings include small sample sizes, differences in T2DM susceptibility across ethnic groups, variation in environmental exposures, and gene-environmental interactions. For this controversy, only four (PPARγ, ABCC8, KCNJ11, and CALPN10) candidate genes are thought to be most promising.

Table 4: Commonly estimated genes for Type 2 Diabetic Mellitus Susceptibility Genes (Parra e*t al.*, 2004).

Gene	Locus	Variant	Estimated Related Risk (RR)
PPARy	3p25	Pro12Ala	1-3
ABCC8	11p15.1	Ser1369Ala	2-4
KCNJ11	11p15.1	Glu23Lys	1-2
CALPAIN10	2q37.3	A43G	1-4

PPARy gene is located in chromosome number 3 and currently is an important regulator of adiposite development and lipid metabolism (Tontonoz *et al.*, 1994) has become a potential therapeutic target for the treatment of a diverse array of disorders, including T2DM, dyslipidacmia, inflammation and malignancy. One form of the *PPARy* gene (Pro) decreases insulin sensitivity and increases T2DM risk several fold. Approximately 98% of Europeans carry at least one copy of the Pro allele. Thus, its considerable proportion (~25%) contributes to T2DM, particularly among Caucasians.

ABCC8 (ATP binding cassette, subfamily C, member 8) gene encodes the high-affinity sulfonylurea receptor (SUR1) subunit that is coupled to the Kir6.2 subunit (encoded by U*KCNJ11*U, also known as the potassium channel, inwardly rectifying subfamily J, member 11). Both genes are involved in the ATP-sensitive potassium channel that plays important role in regulating the release of insulin and glucagon hormones in the β-cell. *ABCC8* and *KCNJ11* are only at a distance of 4.5 kb. Mutations in either gene can affect the potassium 7channel's activity and insulin secretion, ultimately leading to the development of T2DM. Variant forms of *KCNJ11* (Lys) and *ABCC8* (Ala) genes have been associated with T2DM.

CALPAIN-10 encodes an intracellular calcium-dependent cysteine protease that is ubiquitously expressed (Frayling *et al.*, 2001). A haplotype that was initially allowed to T2DM included an intronic A to G mutation at position 43, which appears to be involved *in CALPAIN10* transcription.

Two amino acid polymorphisms (Thr504Ala and Phe200Thr) have also been associated with T2DM risk. However, Physiological studies suggest that variations in calpain 10 activity effects insulin secretion and therefore, susceptibility to T2DM.

1.10. Genetics of Maturity-Onset Diabetes of the Young (MODY)

MODY displays an autosomal dominant pattern inheritance, generally spanning three generations (Stride & Hattersley, 2002). With the progress of molecular genetics, it is identified that there are at least six forms of MODY, each of which caused by a mutation in a different gene that is directly involved with β-cell function (Winter, 2003). Autosomal dominant, early-onset diabetes (known as maturity-onset diabetes of the young, MODY) has been revealed to be a genetically heterogeneous

condition often non-insulin-dependent forms of diabetes that are defined at the molecular genetics level by mutations in different genes (Ellard *et al.*, 2008). Mutations found in the glucokinase gene (Vionnet *et al.*, 1992) and in genes for various transcription factors (hepatocyte nuclear factor (HNF)-1a (Yoshiuchi *et al.*, 1999), HNF- 1b (Horikawa *et al.*, 1997), HNF-4a (Yoshiuchi *et al.*, 1999), insulin promoter factor 1 (IPF1) (Staffers *et al.*, 1997) and neurogenic differentiation 1 (NEUROD1) (Malecki *et al.*, 1999), each capable of producing a phenotype in which β-cell dysfunction predominates. However, there is controversial clinical differentiation between monogenic and multifactorial forms of some genes (e.g. IPF1) seem capable of producing either phenotype (Staffers *et al.*, 1997; Stoffers *et al.*, 1999). A most common cause of MODY is mutations in the GCK and HNF1A genes (about in 70% cases) are found in all populations studied. Rarer forms of MODY include heterozygous mutations in IPF1 (also known as PDX1) and NEUROD1 gene (Ellard *et al.*, 2008).

1.11. Insulin promoter factor (IPF)-1

Insulin promoter factor- 1(IPF-1, also known as PDX-1, IDX-1, and STF-1) is responsible for the development of the pancreas in the embryo and is also a key regulator of insulin gene expression. In developing embryo, the pancreas is formed two buds (exocrine and endocrine) of the primitive gut that eventually fuse to form the pancreas gland. Several pancreas development studies had identified many transcription factors are use in different stages of development. IPF1 is one such transcription factor and is an early pancreatic marker that is also found in adult b cells. In the embryo, the presence of IPF1 is vital to ensure the correct development of the pancreas. Loss of both copies of the gene can cause the pancreatic agenesis. In absence of IPF1gene, the proliferation and differentiation of precursor cells into endocrine and exocrine parts of the pancreas are blocked (Jonsson *et al.*, 1994; Staffers *et al.*, 1997). It has been evident that homozygous null mutation of IPF1 causes lack a pancreas in mice. And IPF1 +/– mice do not show any relevant phenotype at birth, but haploinsuffciency of IPF1 in β-cells causes impaired glucose tolerance in older mice (18 months) (Thomas *et al.*,2001). Which means this gene is also essential for normal pancreatic function in adults. As well as in regulation of various target genes including GLUT2, GCK, insulin, somatostatin and islet amyloid polypeptide (IAPP) (Sujjitjoon *et al.*, 2008). Thus, IPF1 sequence variants might be anticipated to influence both β-cell mass as well as the expression of key β-cell genes. IPF1 gene in mutation screening for permanent neonatal diabetes biallelic PDX1 mutations shows evidence of 2.9% permanent neonatal diabetes (Franco *et al.*, 2013). A dominant-negative frameshift mutation in the IPF-1 gene was identified in a single family and shown to cause pancreatic agenesis when homozygous and maturity-onset diabetes of the young (MODY) when heterozygous (Macfarlane *et al.*,1999). Therefore, IPF1 sequence variants might be anticipated to influence both β-cell mass as well as the expression of key β-cell genes.

1.11.1. IPF1 gene and its polymorphism

The genomic location of the IPF1 gene is on chromosome number 13, later specified and assigned to chromosome number 13q12.1 (Vionnet *et al.*, 1995) and size is 6321 bases. It starts from 27,919,994 bp and end in 27,926,314 bp in this chromosome. This IPF1 gene contains 2 exons with an intronic region spanning about 6 kb and encodes a protein with 283 amino acids.

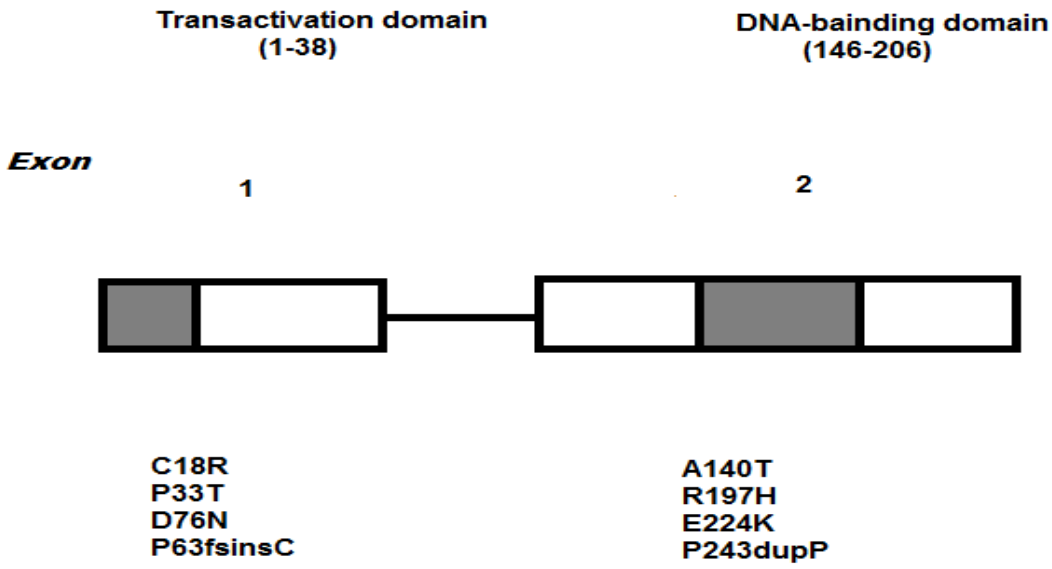


Figure 4: IPF-1 gene contains 2 exons. Location of IPF-1 mutations within the 2exons and functional domains of the IPF-1 protein are shown; the numbers in brackets refer to codons and C18R, P33T, D76N, A140T, R197H, E224K, P63fsinsC, P243dupP are the variants (Source: Winter & Silverstein, 2000).

IPF1 gene contains a lot of variants in its two exonic regions as well as its upstream region. C18R, P33T, D76N, A140T, R197H, E224K (Winter & Silverstein, 2000), E164D, E178K (Schwitzgebel et al., 2003), P63fsinsC, P243dupP (Winter & Silverstein, 2000) are the variants present in its exons. A mutation on those variants could affect its transactivation or DNA –binding activity as well as decrease the protein half-life.

1.11.2. Transactivation of insulin gene using IPF1 gene

An increase in blood glucose level stimulates insulin gene transcription and insulin secretion (Goodison *et al.*, 1992). Insulin gene transcription is mainly controlled by a 340 bp promoter region upstream of the transcription initiation site of the insulin gene. The insulin promoter is organized in a complex arrangement of discrete cis-acting sequence motifs, which serve as binding sites for both

ubiquitous and β-cell-specific transcription factors (Melloul *et al.*, 2002). The coordinated interaction between cis-elements and transacting factors at the promoter contributes to both the β-cell-specific insulin gene expression (Melloul & Cerasi, 1994; Docherty *et al.*, 1996). Much of the glucose responsiveness inherent to the insulin promoter is conferred by the A3, E1 and C1 sites, which are bound by the transcription factors IPF-1 (Hui & Perfeti, 2002; Al-Quabaili & Montenarh, 2008) along with other two transcription factor NeuroD1 (neurogenic differentiation 1)/ β-2 & MafA (V-mafmusculo aponeurotic fibrosarcoma oncogene homologue A) (Kaneto *et al.*, 2007; Aramata *et al.*, 2007), respectively. These three transcription factors act in a coordinated and synergistic manner to stimulate insulin gene expression in response to increased glucose levels (Matsuoka *et al.*, 2007). All three regulatory elements (A3, E1, and C1) of the insulin promoters are important for glucose regulation of insulin gene expression (Hay *et al.*, 2006). It has been projected that glucose modulates the function of IPF-1 by regulating the localization, DNA-binding activity and interaction of IPF1/Pdx-1 with multiple proteins (Harmon *et al.*, 1999; Labrun *et al.*, 2005; Barrow *et al.*, 2006).

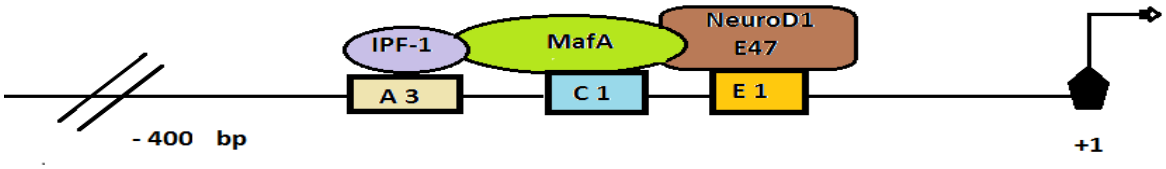


Figure 5: Coordinated and synergistic activation of insulin gene expression by IPF-1, NeuroD1 and MafA. IPF-1 binds to the A boxes, NeuroD1 to the E-boxes and MafA to the C1 element within the 400bp region of the insulin promoter and activate insulin gene expression in a coordinated and synergistic manner in the presence of elevated blood glucose levels (Andrali *et al.*, 2008).

1.11.3. Role of IPF-1 in pancreatic β-cell function

IPF-1/PDX-1, the major regulator of glucose-stimulated insulin gene transcription, is essential for embryonic development of the pancreas (McKinnon & Docherty, 2001; Hui & Perfeti, 2002; Melloul *et al.*, 2002; Al-Quobaili & Montenarh, 2008). The IPF-1 is first expressed in the primitive gut tube during embryonic development of the pancreas around 8.5 embryonic days. IPF-1 homozygous knockout mice fail to develop a pancreas and die shortly after birth, because of a lack of insulin (Jonsson et al., 1994; 1995 & 2003). Heterozygous IPF-1 mice have a normal pancreas but decreased insulin production results in hyperglycemia. Pancreatic agenesis is also observed in a patient carrying a homozygous single nucleotide deletion in the IPF-1 gene (Staffers *et al.*, 1997). Diabetes with aging is caused by β-Cell-specific disruption of the IPF-1 gene. So, IPF-1 is essential for the maintenance of β-cell identity. These findings indicate that IPF-1 is required for early development of the pancreas

and for β-cell maturation and function. In the fully developed pancreas, IPF-1 expression is primarily limited to the insulin-producing β-cells and somatostatin producing δ-cells within the pancreatic islets. However, its expression also has been observed in the duodenum epithelial cells (Miller *et al.*; 1994). IPF-1 mainly acts in the β-cell to up-regulate the transcription of several β-cell-specific genes like insulin, GLUT2 (glucose transporter-2), glucokinase, somatostatin, islet amyloid polypeptide and MafA, and auto-regulating its own expression (Waeber *et al.*, 1996). It also functions as a transcriptional repressor for glucagon, cytokeratin K19 and c-Myc (Ahlgren *et al.*, 1996 & 1998; Liu *et al.*, 2007). This repressor activity is essential to maintain β-cell character by inhibiting the expression of genes specific for α-cells (glucagon) and pancreatic ductal cells (cytokeratin K19). IPF-1transactivation domain is at the N-terminus which mediates unique protein-protein interactions with other transcription factors such as NeuroD1, transcriptional co-regulators, and p300. It also has a central DNA-binding homeodomain including a novel nuclear localization signal (Hessabi *et al.*, 1999; Moede *et al.*, 1999). IPF-1binds to the insulin promoter A-elements and activates insulin gene transcription. Pancreatic agenesis, impaired glucose tolerance and/or diabetes in humans have observed due to various SNP in IPF-1 gene. Specific point mutations in IPF-1 also are related to MODY (maturity-onset diabetes of the young) 4 and late-onset Type 2 diabetes, characterized by reducing in β-cell function (Staffers *et al.*, 1997; Stoffers *et al.*, 1999; Weng *et al.*, 2001).

1.11.4. Insulin gene transcription by IPF-1

Andrali (2008) suggests that one of the mechanisms by which IPF-1 promotes insulin gene transcription is by mediating histone modifications at the insulin promoter. IPF-1 has been demonstrated to recruit the HAT (histone acetyl-transferase) p300 to the insulin promoter only at high levels of glucose (10–30 mM), which leads to increased acetylation of histones (Guo, 2014). The histone methyltransferase Set9 has been recruited to the insulin promoter usingIPF-1, leading to the dimethylation of histone H3 Lys4 (Francis *et al.*, 2005). Both of these histones lead to changes in the chromatin structure that promote insulin gene transcription. Binding of IPF-1 to the insulin promoter is essential for RNA polymerase II (pol II) to adopt the elongation isoform for active transcription (Francis *et al.*, 2005). In the absence of IPF-1, pol II unable to switch the initiation (pSer5) to the elongation (pSer2) isoform, which is essential for active transcriptional elongation (Francis *et al.*, 2005). These findings suggest that IPF-1 regulates insulin gene transcription by promoting a transcriptionally active chromatin structure, which enhances elongation by pol II. Some other evidence suggests IPF-1 involves in transcriptional suppression of glucagon (Petersen *et al.*, 1994), cytokeratin K19 and c-Myc genes (Wescott *et al.*, 2009; Liu et al., 2007).

In summary, IPF-1 regulates insulin gene transcription by promoting the assembly of multiprotein complexes at its promoter region (Andrali *et al.*, 2008).

1.11.5. Regulation of IPF-1 function

Glucose is the physiological regulator of insulin gene transcription, which moderates IPF-1/Pdx-1 function in pancreatic β-cells by multiple mechanisms. It is discovered that glucose regulates the interaction of IPF-1 with various co-regulators in a phosphorylation-dependent manner in the mouse insulinoma cell lineMIN6 (Andrali *et al.*, 2008). When blood glucose level is normal or low in condition (1–3 mM), IPF-1 associated with HDAC-1 and HDAC-2 decrease insulin gene expression (Mosley *et al.*, 2004). On the other hand, increases in glucose levels (10–30 mM) disrupt the interaction of IPF-1 with HDACs. At the same time, it promotes IPF-1 association with the HAT p300, which leads to hyperacetylation of histone H4 and the stimulation of insulin gene transcription (Mosley *et al.*, 2004). In response to high glucose IPF-1 interaction requires a phosphorylation event that causes changes in IPF-1 localization (Mosley *et al.*, 2004). On the contrary, IPF-1 is localized mainly to the nuclear periphery and interacts with HDAC-1 and HDAC- 2 when the glucose level is low. While on high glucose, IPF-1 interacts with p300 and is localized throughout the nucleus. Several signaling pathways including the p38/SAPK (stress-activated protein kinase), PI3K (phosphoinositide 3-kinase), PKC (protein kinase C) and MAPK (mitogen-activated protein kinase) as well as PAS (Per- Arnt-Sim) kinase pathways have been engaged in IPF-1 phosphorylation, nucleocytoplasmic shuttling, DNA binding and transactivation potential (Mcfarlane *et al.*, 1997). Ser61 and Ser66 of IPF-1 have been demonstrated to be phosphorylated in response to p38, MAPK, and PI3K signaling. Thr152 is phosphorylated by PAS kinase and to regulate the nucleocytoplasmic shuttling and transactivation potential of IPF-1(Boucher *et al.*, 2006). However, the exact residues that become phosphorylated by the various kinases and their physiological significance remain to be determined. Various studies indicate that modification by phosphorylation also regulates the stability of the IPF-1 protein. The DNA dependent protein kinase has been shown to phosphorylate Thr11 in IPF-1 in response to radiation-induced DNA damage to promote IPF-1 protein degradation (Lebrun *et al.*, 2005). Another study has demonstrated that GSK3 (glycogen synthase kinase 3) phosphorylates IPF-1 on Ser61 and/or Ser66 in response to oxidative stress, which also results in degradation of the IPF-1 protein (Baucher *et al.*, 2006; Liu *et al.*, 2008). These studies provide a possible mechanism for previous observations where IPF-1 protein levels are reduced under glucotoxic and oxidative stress conditions (Robertson & Harmon, 2007). However, other studies have shown that oxidative stress inhibits IPF-1 nuclear localization and DNA binding through the activation of the JNK (c-Jun N-terminal kinase) pathway (Kawamori *et al.*, 2003). The nuclear exclusion of IPF-1 would also explain the decreases in insulin gene transcription observed under oxidative stress conditions. The interplay of IPF-1 protein stability and intracellular localization requires further investigation to elucidate fully the role of IPF-1 in β-cell dysfunction under oxidative stress conditions. Several studies suggest that IPF-1 may be subject to several other post-translational modifications, including SUMOylation (Kishi *et al.*, 2003) and O-linked glycosylation (Gao *et al.*, 2003). Although the sumoylation of IPF-1 appears

to modulate IPF-1 localization and stability (Kishi *et al*., 2003), glycosylation of IPF-1 has been reported to regulate IPF-1 DNA binding activity (Gao *et al*., 2003). Exposure of pancreatic islets to elevated levels of fatty acids such as palmitate has been shown to reduce the nuclear localization and expression levels of IPF-1 (Hagman *et al.*, 2005 & 2008; Yoshikawa *et al.*, 2001). Other factors that regulate IPF-1 levels include the GLP1 (glucagon-like peptide 1), which is produced in the intestine and promotes glucose-regulated transcription of insulin (Hussain & Habener, 2000).

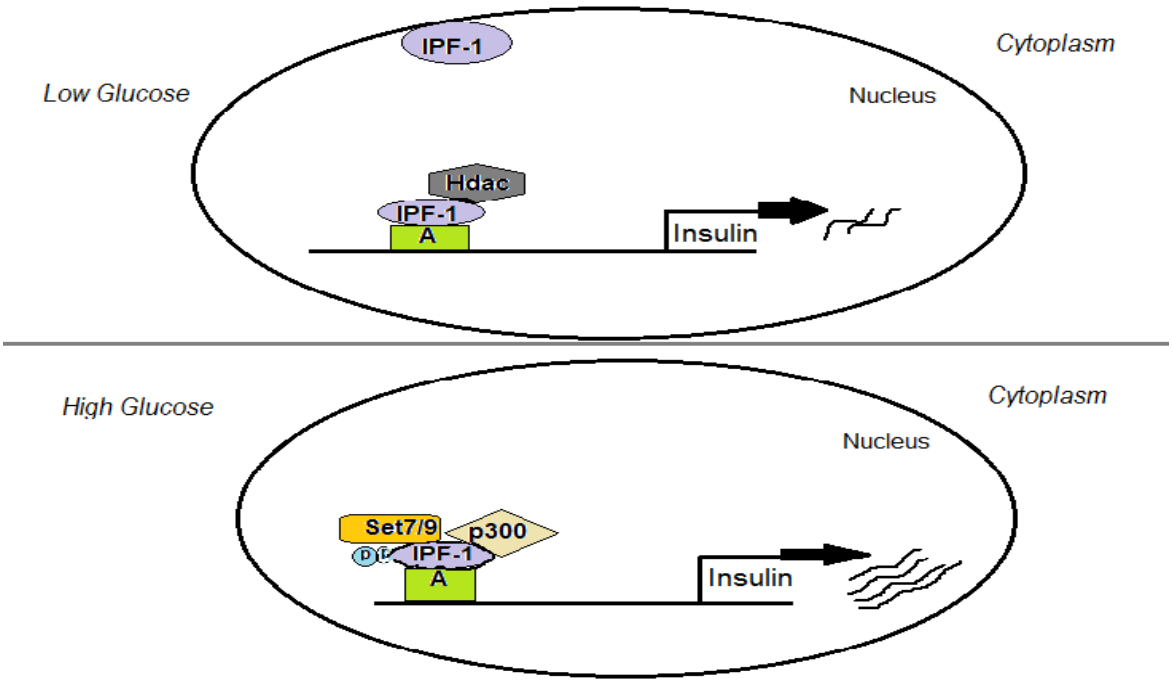


Figure 6: Glucose regulates the sub cellular localization and interaction of IPF-1 with co-regulators (Andali *et al.*, 2008).

Expression of IPF-1 in pancreatic β-cells appears to be independent of the glucose concentration; however, glucotoxic conditions lead to a reduction in IPF-1 levels.

In conclusion, changes in glucose levels have been shown to regulate the nuclear localization, DNA binding and transactivation capacity of IPF-1 and its interaction with co-regulators. Nuclear localization and IPF-1 interaction with co-activators appear to be dependent on phosphorylation of IPF-1. Based on the published results, IPF-1 is mainly in the nuclear periphery under low or normal glucose conditions and interacts with HDACs. In response to high glucose, IPF-1 becomes phosphorylated and translocates into the nucleoplasm and interacts with the HAT p300 to stimulate insulin gene expression.

Chapter Two:

Rationale, Hypothesis & Objectives

2.1. Rationale

A series of genetic studies have undertaken to investigate the pathophysiological basis of T2DM and MODY in the Bangladeshi population. In Bangladeshi subjects of type 2 diabetes impaired insulin secretion & insulin resistance are present (Roy *et al.*, 2007). Leptin, PRARγ, SOD1, CAPN10, INS, TNF-α and many more genes are already investigated in the Department of Physiology & Molecular Biology laboratory of Bangladesh University of Health Sciences (BUHS) however some are associated with T2DM rest are not. For example, Leptin Q223R polymorphism may involve with IGR, T2DM and β-cell secretor defect IGR subjects of Bangladeshi origin. On the other hand, leptin 2548 G/A variant polymorphism of this gene has no association with T2DM pathogenesis. PPARγ Pro12Ala is not directly involved with hyperglycemia or hyperinsulinemia but associated with high WHR (waist-hip ratio) which will possibly predispose to the development of T2DM. SOD1 gene +35 A/C polymorphism has suggested being associated with T2DM pathogenesis. CAPN10 single nucleotide polymorphism (SNP) is not connected with T2DM. INS VNTR (A/T) has involved with central obesity of our population but not with T2DM. TNF-α has shown negative involvement among T2DM patients of the Bangladeshi population.

However, the dual role of insulin promoting factor 1 (IPF1) gene during the development of the pancreas in the embryo and as a regulatory factor of pancreatic genes in adults emphasizes the importance of this gene in glucose homeostasis. Mutations in IPF1 are a rare cause of early-onset T2DM (MODY4) (Chevre *et al.*, 1998; Shepherd *et al.*, 2001; Hansen *et al.*, 2000). Several previous studies have searched for mutations in late-onset T2DM among Caucasians with variable results in different ethnic groups (Macfarlane *et al.*, 1999; Hansen *et al.*, 2000).

Therefore, in this study, we selected IPF-1 as a candidate gene for the common late-onset type 2 diabetes in the Bangladeshi population. And in this study IPF1 gene's two variants are taken C18R and R197H as because both are located in the exonic region.

2.2. Hypothesis

IPF1 gene C18R (T/C) and R197H (G/A) variants are associated with type 2 diabetes mellitus in the Bangladeshi population.

2.3. General Objective

To determine the frequency of both (C18R (T/C) & R197H (G/A)) variants of IPF1 gene and to find the association of this polymorphism with T2DM in the Bangladeshi population.

2.4. Specific Objective

1. To distinguish the BMI, WHR and blood glucose of the healthy individuals (control) & diabetic subjects.
2. To investigate the frequency of both variants of IPF1gene (C18R & R197H) in diabetic subjects in the Bangladeshi Population.
3. Figure out the mutational positions and locations of C18R & R197H in IPF1gene on the coding region in secondary and 3D structure through computational study.
4. To find out the association of this polymorphism with T2DM in the Bangladeshi population.

Chapter Three:

Materials & Methods

3. MATERIALS AND METHODS

3.1. Place of the study

The study was conducted in the laboratory of the Department of Physiology & Molecular Biology, Bangladesh University of Health Sciences (BUHS), Dhaka, Bangladesh.

3.2. Study design

It was a cross-sectional observational study.

3.3. Subjects

A total number of 445 subject's blood sample was used from the repository of the Department of Physiology and Molecular Biology for this study, in which the total 245 subjects were healthy control and 200 T2DM subjects. These depositories were collected previously for genetic studies following the appropriate procedure.

Determination of sample number for this study:

The number of samples for the study was calculated using the following formula:

$$n = \frac{\{t^2 * p(1-p)\}}{d^2}$$

Here,

n = number of samples

t = confidence level at 95% (standard value of 1.96)

p = frequency of allele (p) was taken as 15% of the study population

d = margin of error at 5% (standard value of 0.05)

The calculated number of the study was 196 and to keep away from technical issues among research facility investigation and accomplish statistical analysis extra 50 tests required to be incorporated which turned an aggregate number of subjects to 250 in the study group and healthy controls.

But practical reason, 200 samples from T2DM subjects and 245 samples from non-diabetic subjects were included in this study.

3.3.1. Recruitment of the subjects

Subjects consecutively attending the BUHS hospital, a sister concern of Bangladesh Diabetic Samity (BADAS) for their regular checkup were recruited. Healthy control subjects were recruited through personal communication from the community. Purposes and methods of the study were briefed to each individual and informed consent was taken from the study subjects. Subjects were advised to be

on their usual diet for 3 days and requested to report on a prescheduled morning after overnight fasting (8-10 hours).

3.3.2. Selection criteria

The study subjects were recruited following stringent inclusion and exclusion criteria.

3.3.2.1. Inclusion criteria

- Healthy control subjects, newly diagnosed T2DM subjects.
- Adult subjects with an age range 30 to 60 years.
- Voluntary agreed to include in this study by providing informed consent.

3.3.2.2. Exclusion criteria

- Subjects with co-morbid diseases (infection, stroke, myocardial infarction, major surgery, essential hypertension, malabsorption etc.).
- Subjects with the history of medication, which may significantly affect glucose metabolism (glucocorticoids, oral contraceptives contraceptives containing levonorgestral or high-dose estrogen, phenytoin, high-dose thiazide diuretics etc.)
- Pregnant women and lactating mother.
- Patients had a recent blood transfusion.

3.4. Ethical issues

The precaution was taken at the time of collection to address the ethical aspect. Consent was obtained from the volunteers to the future studies involving molecular analysis in these samples. The result will only be used for scientific purposes. The identity of the individual would not be disclosed. The volunteers in no circumstance would be communicated involving any further analysis.

3.5. Biochemical analysis

The following biochemical parameters were analyzed for the study.

3.5.1. Estimation of glucose

Glucose was estimated by enzymatic colorimetric (GOD-PAP) method in the Hitachi 704 Automatic Analyzer, Hitachi Ltd., Tokyo, Japan using reagents of RANDOX Laboratories Ltd., UK.

Principle

Glucose is determined after enzymatic oxidation in the presence of glucose oxidase. The hydrogen peroxide formed reacts with phenol and 4-aminophenazone under the catalysis of peroxidase to form a red violate quinoneimine dye as an indicator (Trinder, 1969).

$$\text{Glucose} + H_2O \xrightarrow[\text{Oxidase}]{\text{Glucose}} \text{Gluconic acid} + H_2O_2$$

$$2H_2O_2 + \text{4-aminophenazone} + \text{phenol} \xrightarrow{\text{Peroxidase}} \text{Quinoneimine} + 4H_2O$$

Table 5: Concentration of reagent used for ELISA.

Reagent Name	Initial concentration of solution
Phosphate buffer	0.1 mol/L, pH 7.0
Phenol	11 mol/L
4-aminophenazone	0.77 mmol/L
Glucose oxidase	≥1.5kU/L
Peroxidase	≥1.5kU/L
Glucose standard	5.55 mmol/L (100mg/dL)
Uranyl Acetate	0.16%

Materials

- Micro-centrifuge tube
- Micropipettes and pipettes
- Disposable tips
- Automatic Analyzer (Boehringer Mannheim, 704; HITACHI)

Procedure

The method determines glucose without deproteinization. The instrument was calibrated before estimation. Serum and reagents were taken in a specific cup. They were arranged serially into the auto analyzer. The auto analyzer was programmed for the estimation of glucose and allowed to run with the following procedure:

5 µ1 sample and 500µ1 reagent were mixed and incubated at 37° C for 10 minutes. The reaction occurred in the reaction cell or cup. The absorbance of the sample and the standard against the reagent blank were measured at 500 nm within 60 minutes.

Calculation of the result

Optical densities or absorbances were fed into a computer and calculation was done using the software program. Values for the unknown samples were calculated by extrapolating the absorbance for the standard using the following formula.

$$\text{Glucose concentration (mmol/L)} = \frac{A_{\text{sample}}}{A_{\text{standard}}} \times 5.55$$

3.6. Genetic analyses

IPF-1 genotype was determined by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method.

3.6.1. DNA extraction

Extraction of nuclear DNA (n = 445) was performed using Gen Elute Genomic DNA Extraction Kit (QIAGEN, USA) from stored blood.

DNA extraction using QIAGEN kit:

Principle:

The kit uses the principle of silica gel membrane DNA isolation procedure from whole blood adapted into a spin column. The QIAamp spin procedure, which is ideal for simultaneous processing of multiple samples, yield pure DNA ready for direct amplification in just 20 minutes. All centrifugation steps were carried out using a refrigerated microcentrifuge.

Equipment, reagent and accessories

- Water bath
- Vortex mixer
- Centrifuge
- Microcentrifuge tubes (1.5 ml)
- Pipette tips at different capacities (10 µl, 20 µl, 100 µl, 200 µl and 1000 µl)
- Micropipettes (10 µl, 20 µl, l00µl, 200µl, 100-1000 µl)
- Absolute ethanol (95-100 %)

Content of DNA kit

- Extraction Column
- Collection tubes
- Proteinase K
- RNase-A solution
- Lysis Buffer AL
- Wash Buffer AW1
- Wash Buffer AW2
- Elution Buffer AE

Preparation of reagents

Wash buffer Wl: 125 ml absolute ethanol (96-100%) was added to 125 ml Buffer AL.

Wash buffer W2: 160ml absolute ethanol (96-100%) was added to 66 ml buffer AW2.

Extraction procedure

Stored frozen blood samples were brought to room temperature and made homogenous by vigorous vortexing. Aliquot 200 µl blood from each sample, were transferred into a 1.5 ml microcentrifuge tube. 20 µl QIAGEN proteinase K was added inside the cap of the every microcentrifuge tube. Then, 4 µl of an RNase-A stock solution (l00mg/ml) was added to every sample before the addition of Buffer AL. 200 µl Buffer AL was added to each sample tube. To ensure efficient lyses, pulse-vortexed for 15 sec to make a homogenous solution. Samples were incubated at 56 ℃ temperature for 10 minutes in a water bath to lyses. To remove drops from the inside of the lid at the tubes centrifuged at 3000 rpm for 1 minute. Then, 200 µl ethanol (96-100%) was added to each sample and mixed by pulse-vortexing for 15 seconds and briefly spanned. Every blood sample mixture was transferred to the QIAamp Mini Spin columns (mounted on 2 ml collection tube) without wetting the rim and after closing the cap, centrifuged at 8000 rpm for 1 minute. The QIAmp Mini Spin columns were placed on a fresh 2 ml collection tube. Then the mini spin columns opened carefully and 500 µl Buffer AW1 was added and centrifuged at 8000 rpm for 1 minute. The collection tubes were discarded and the spin columns were placed on fresh collection tubes. The QIAamp Mini Spin columns were opened and 500 µl Buffer AW2 was added and centrifuged at 14000 rpm for 3 minutes. At this stage, the columns were appeared to be clean. The QIAamp Mini Spin columns were placed on a 2 ml collection tubes and the filtrate containing tubes discarded. Then the collection tubes were centrifuged at 18000 rpm for 1 minute. The mini spin columns were then placed in 1.5 ml tubes and the filtrate containing tubes were discarded. The QIAamp Mini Spin columns were placed again on a 2 ml collection tube and Buffer AE (200 µl) was added to the columns and incubated at room temperature for 5 minutes. Then, centrifuged at 8000 rpm for 1 minute at 25°C. This elutes supposed to contain DNA. The DNA fragment was stored at 4°C or -20°C.

3.6.2. Check for DNA yield

DNA yield for each sample was checked by agarose gel (1%) electrophoresis.

Reagents

- Agarose (molecular grade)
- TBE (Tris-boric EDTA) buffer.
- Gel loading buffer (5X concentration).
- Ethidium bromide (0.05 µg /ml).

Procedure

To prepare agarose gel, an appropriate amount of agarose was taken into polypropylene conical flasks containing the required volume of working tris-borate EDTA (TBE) buffer. Agarose and buffer

solution were mixed by swirling of the flasks. It was then heated to boiling point in a microwave oven with intermittent mixing until agarose was properly boiled. The gel was cooled nearer to the gelling point and ethidium bromide (0.05 µg /ml) was added. The gel was then poured into a horizontal gel mold, combs inserted, and allowed to polymerize. The gel was subsequently placed in a horizontal electrophoresis tank filled with a working TBE buffer.

To resolve DNA extract, 3 µl of DNA elutes was mixed with an appropriate amount of loading buffer and then the mixer was loaded in agarose gel prepared earlier. The gel was run for at medium voltage for the required time. DNA presence was visualized under UV light and gel image was captured.

The intensity of the bands obtained for the extracted DNA samples was visually compared with that of known concentration of DNA and then the extracted DNA was diluted according to the intensity of the bands to make the concentration of all DNA samples between 10-50 ng/ml, optimum for PCR amplification.

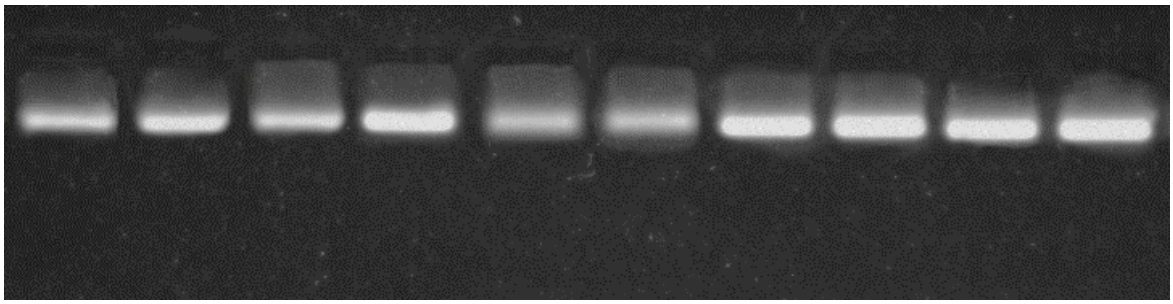


Figure 7: Gel image of DNA electrophoresis to check the extraction yield.

3.6.3. PCR amplification of IPF-1 gene variant allele

The segment of the IPF-1 gene containing the polymorphic marker C18R (T/C) and R197H (G/A) was amplified by polymerase chain reaction (PCR) and the amplified product was checked by agarose gel electrophoresis.

Reagents

- Hot Star Taq Polymerase
- Buffer
- dNTPs
- $MgCl_2$
- Primers
- Dimethyl sulfoxide **(**DMSO)
- Q solution
- Water (Nuclease free UVied)

3.6.3.1. PCR amplification of IPF-1 gene variants

Polymerase chain reaction (PCR) was carried out in 10 µl reaction volume. The PCR was carried out for C18R and R197H. A set of forward and reversed primers were used to amplify the target fragment.

The following primers were used:

For C18R (fragment size 117 bp)

C18R-F: CATGAACGGCGAGGAGCAG

C18R-R: GCCATGTACAGGCACGCAG

For R197H (fragment size 287 bp)

***R197H-F*:** GGTGGAGCTGGCTGTCATGTTG

R197H-R: AGGGCTGTGGCGACGCGTAAG

3.6.3.2. PCR condition

PCR was carried out using HotStarTaq polymerase. Conditions for the amplification for both variant [C18R (T/C) & R197H (G/A)] included initial step of denaturation 95°C for 15 minutes followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 59°C for 30 seconds and elongation at 72°C for 30 seconds and finally a step of final elongation at 72°C for 10 minutes. PCR assays were performed in a DNA thermal cycler (Biometra, USA). A negative control (reagent blank), which contained all components of the reaction mixture except the sample DNA, was included in every PCR procedure. Composition of reaction mix for a PCR of 10 µl contained the following:

Table 6: Concentration of PCR component.

Name of the component	Concentration	Component Volume (µl)
Template DNA		3
PCR buffer		1
$MgCl_2$	25mmol	0.2
Dimithylsulfoxide(DMSO)		0.5
dNTPs	10 mmol	0.2
Primer forward	10 pmol/ µl	0.4
Primer reverse	10 pmol/ µl	0.4
DNA Taq polymerase	5 U/ µl	0.075
Nucleus free Water		4.225
Total		10.00 µl

3.6.4. Evaluation of PCR product

The PCR product was run out on a 10cm x 10cm x 0.6 cm thick 3% agarose gels to assert the amplification of the target DNA fragment. The optimum size of the product was ascertained comparing it with the DNA ladder. The amplified DNA was visualized using under UV light and gel image captured and documented.

Reagents & Equipments:

- Agarose
- TBE (Tris-boric EDTA) buffer
- Ethidium bromide (Sigma, USA)
- DNA Loading buffer (Bromophenol blue 0.25%, Ficoll (Type 400) 15%, Storing at room temperature).
- 100 bp DNA ladder
- Gel electrophoresis system (Biometra, USA)
- Gel documentation system (Bio-Rad, USA)
- Microwave oven
- Electric balance (Scientech, SA 210, USA)

3.6.4.1. 3% agarose gel preparation

Three gram of agarose weighed into polypropylene conical flask. Then 100 ml working TBE (TBE Buffer: Tris-Base 89 mM, Boric Acid 89 mM, EDTA 2 mM) was added. Agarose and buffer solution was mixed by swirling of the flask. Then heated up to the boiling point in a microwave oven for proper mixing. The conical flask with boiling gel was brought out from oven then rotated under tap water to make it slightly cool and ethidium bromide (5µl/100ml) was added. Then the gel poured into a horizontal get mold, combs inserted and allowed to polymerize. The gel was submerged in a horizontal electrophoresis tank filled with a working TBE buffer.

3.6.4.2. Visualization of PCR product under UV light in 3% agarose gel

Three microliter PCR product mixed with 3 µl DNA loading buffer in the microwell plate, mixed using plate centrifuge. Then samples were loaded on to the gel. 100 bp DNA ladder was run together in every lane of the gel. Gel electrophoresis was carried out at 80 V for 30-40 minutes.

The resolved DNA fragments were then visualized under UV ray using a gel documentation system and compared with the target fragment with a 100 bp DNA marker. Visualized gel image then saved as an electronic version.

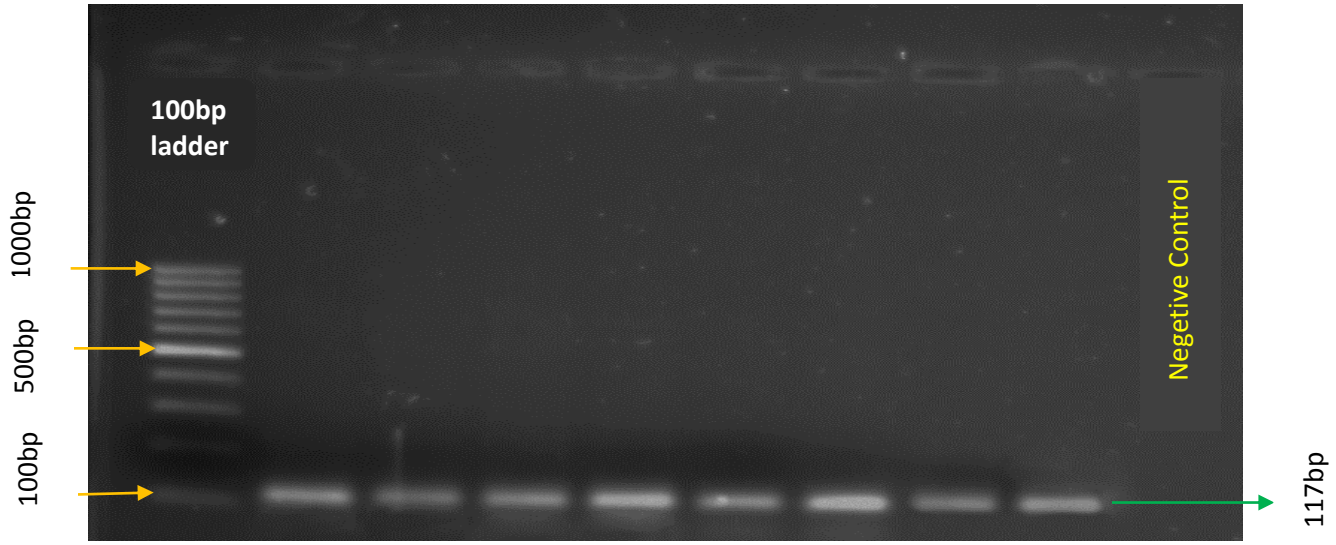


Figure 8: Image of gel analysis of IPF1 C18R (T/C) gene PCR products (in 3% agarose gel).

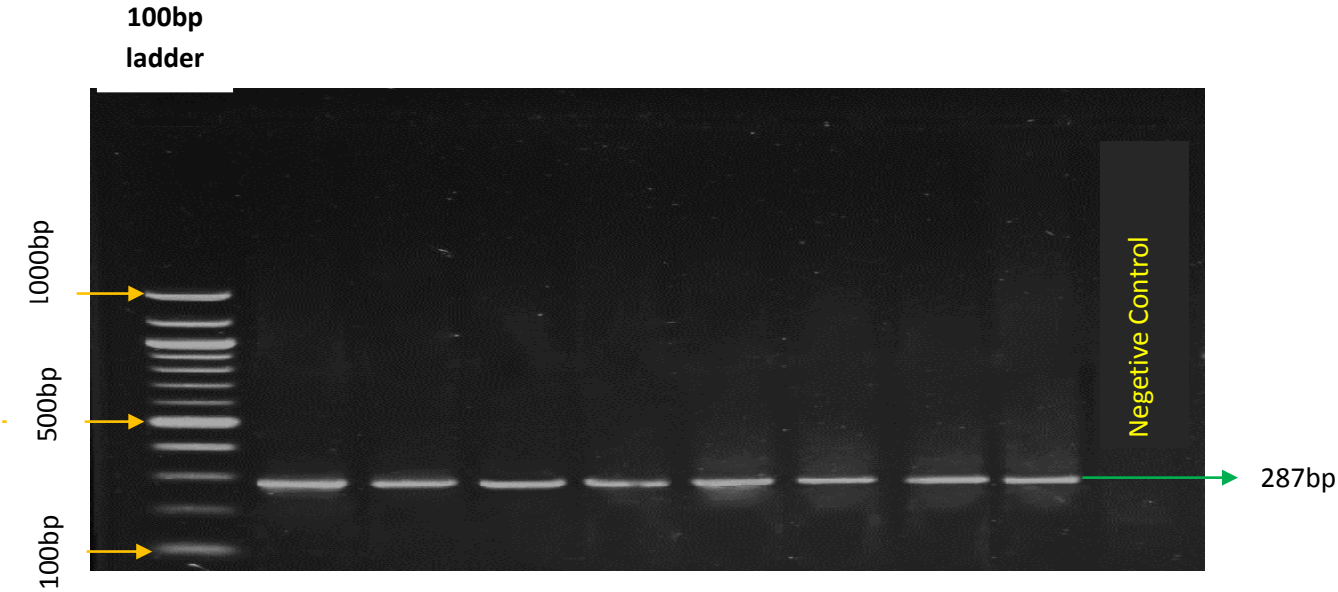


Figure 9: Image of gel analysis of IPF1 R197H (G/A) gene PCR products (in 3% agarose gel).

3.7. RFLP analysis of IPF1 gene candidate variant C18R (T/C) & R197H (G/A)

IPF1 gene candidate polymorphic marker C18R (T/C) & R197H (G/A) were analyzed using site-specific restriction enzyme NlaIII and Fnu4HI respectively. Restriction enzyme digestions performed following standard digestion protocol. Candidate IPF1 variant C18R (T/C) and R197H (G/A) doesn't disrupt restriction enzyme site and were detected by restriction fragment length polymorphism (RFLP) assay.

Equipments

- Water bath (JENCONS, USA)
- Electric balance (Scientech, SA 210, USA)

- Vortex mixer (Gallenkamp, UK)
- Refrigerated microcentrifuge (Hettich, USA)
- Centrifuge (Hanil, South Korea)
- Microwave oven
- Gel electrophoresis system (Biometra, USA)
- Gel documentation system (Bio-Rad, USA)
- Restriction enzyme kit

3.7.1. NlaIII & Fnu4HI digestion

IPF-1 C18R (T/C) & R197H (G/A) variants polymorphic marker was analyzed using NlaIII & Fnu4HI restriction enzyme, respectively. The digestion was carried out in a reaction volume of 10 µl. The enzyme digestion protocol was as follows:

Table 7: NlaIII & Fnu4HI restriction enzyme digestion protocol.

Name of the component	Concentration	Component Volume (µl) for C18R (T/C) digestion	Component Volume (µl) for R197H (G/A) digestion
PCR product		5	5
Restriction enzyme Buffer G*	10x	1	1
RE- *NlaIII*	5U/ µl	0.4	
RE - Fnu4HI	10U/ µl		0.25
H_2O (Nuclease Free water)		3.6	3.75
Total reaction volume		**10**	**10**

*Restriction enzyme buffer G supplied in 10x of working concentration. 1x buffer G contain Tris-HCl (10mM) of 7.5 pH, $MgCl_2$ (10mM), NaCl (50mM) and BSA (0.1mg/ml).

NlaIII and Fnu4HI restriction enzymes digestion were carried at 37°C for 3-4 hours in a water bath. The digestion products were loaded onto a 4% agarose gel for genotyping and run for 4 hours in an electric field at 40V. The digested product was visualized using a gel documentation system following ethidium bromide staining. Fragment sizes were determined compared to undigested PCR product and 100 bp DNA ladder.

3.7.2. The IPF-1 C18R (T/C) polymorphism restricts the NlaIII restriction enzyme site as follows

The IPF-1 gene's C18R (T/C) variant size is 117 bp. NlaIII restriction enzyme has 3 internal cutting site within this variant. NlaIII restriction enzyme recognizes CATG^ sites.

5' C A T G ↓ 3'

3' ↑ G T A C 5'

Normally, the homozygous wild type genotype variant produced 50 bp, 61 bp, 4 bp & 2 bp fragments. However, in case of mutation presence; a restriction site abolishes and 111 bp, 4 bp & 2bp fragmentation occurred in the heterozygous (Ht) genotype. 4 bp & 2 bp fragments size were too small that is why the only 111 bp found in 4% Agarose gel electrophoresis.

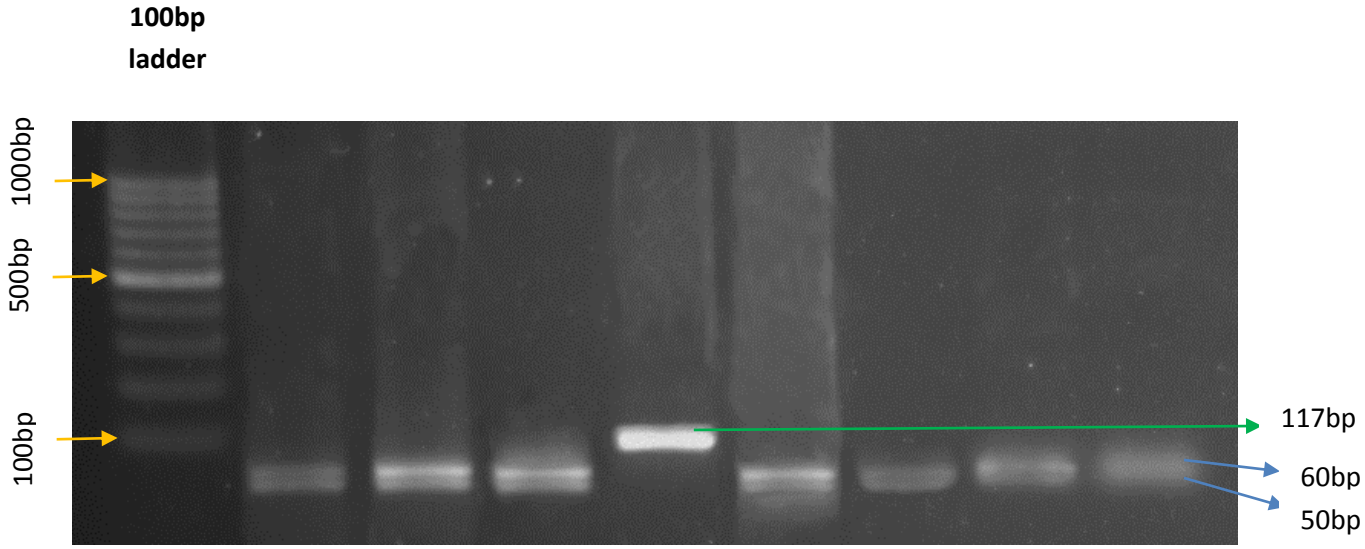


Figure 10: IPF-1 gene C18R (T/C) candidate marker analysis by Nla III restriction enzyme digestion (in 4% agarose gel).

3.7.3. The IPF-1 R197H (G/A) polymorphism restricts the Fnu4HI restriction enzyme site as follows

The IPF-1 gene's R197H (G/A) variant size is 287 bp. Fnu4HI restriction enzyme has no internal cutting site within this variant. Fnu4HI restriction enzyme recognizes GC^NGC sites.

5' G C ↓ N G C 3'

3' C G N ↑ C G 5'

Normally, the homozygous wild type variant does not have any restriction site. However, mutation creates a restriction site, which produces 187 bp, & 100 bp fragments in heterozygous (Ht) genotype.

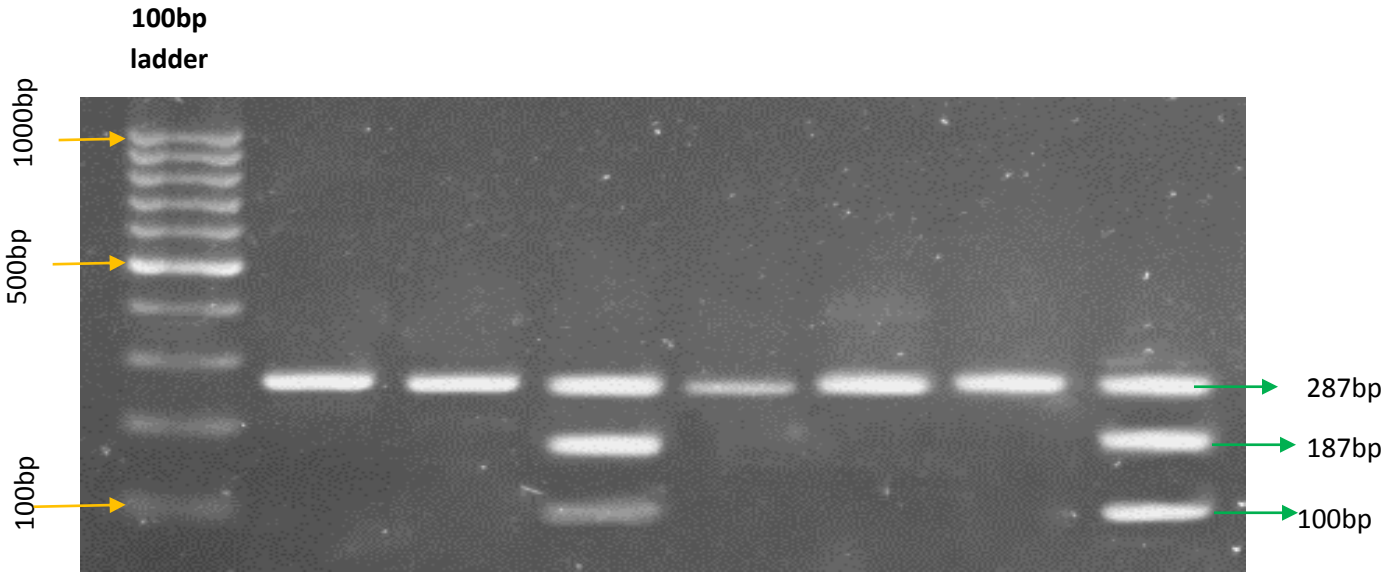


Figure 11: IPF-1 gene R197H (G/A) candidate marker analysis by Fnu4HI restriction enzyme digestion (in 4% agarose gel).

3.8. Computational study

3.8.1. Retrieval of IPF-1 Sequence from the NCBI Nucleotide Database

The nucleotide sequence of IPF-1/ PDX1 was retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/).

Retrieval of the IPF-1 nucleotide sequence done by providing the name of the gene "Homo Sapiens PDX1" sequence in the NCBI search bar. (http://www.ncbi.nlm.nih.gov/nucleotide/). Downloading of full-length sequence and the coding sequence of PDX1 was done by clicking send to option.

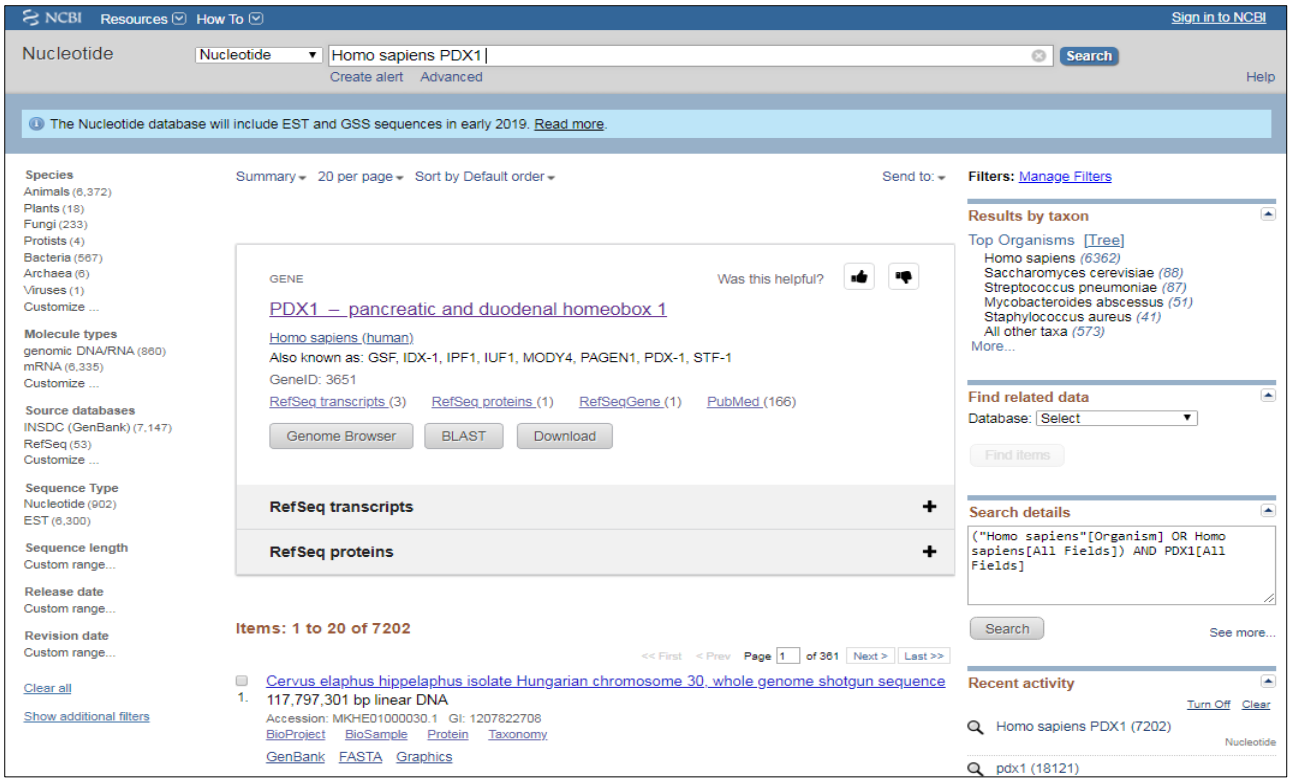


Figure 12: Main search window of NCBI Nucleotide

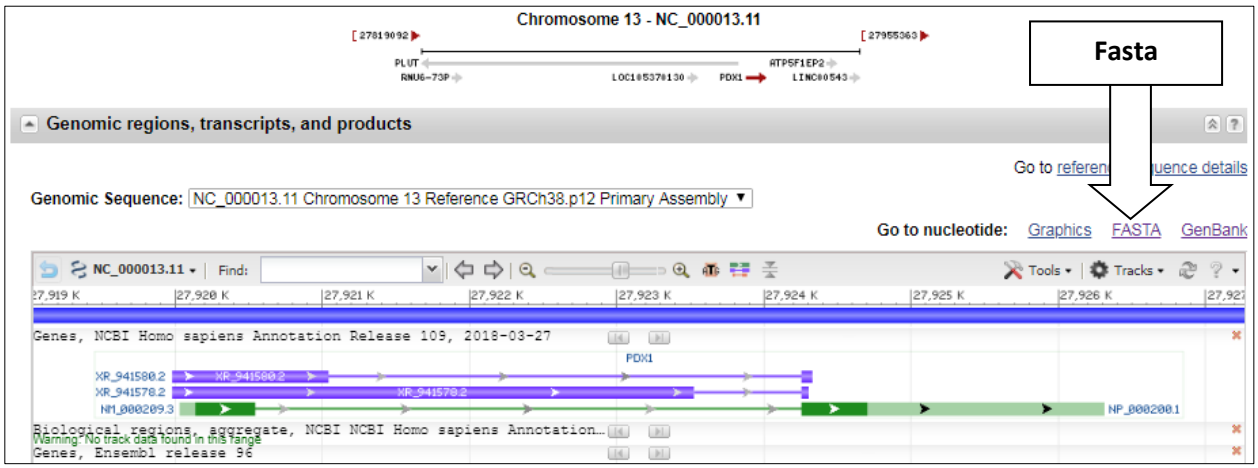


Figure 13: Redirect to the Fasta file page for the IPF-1 gene (NCBI)

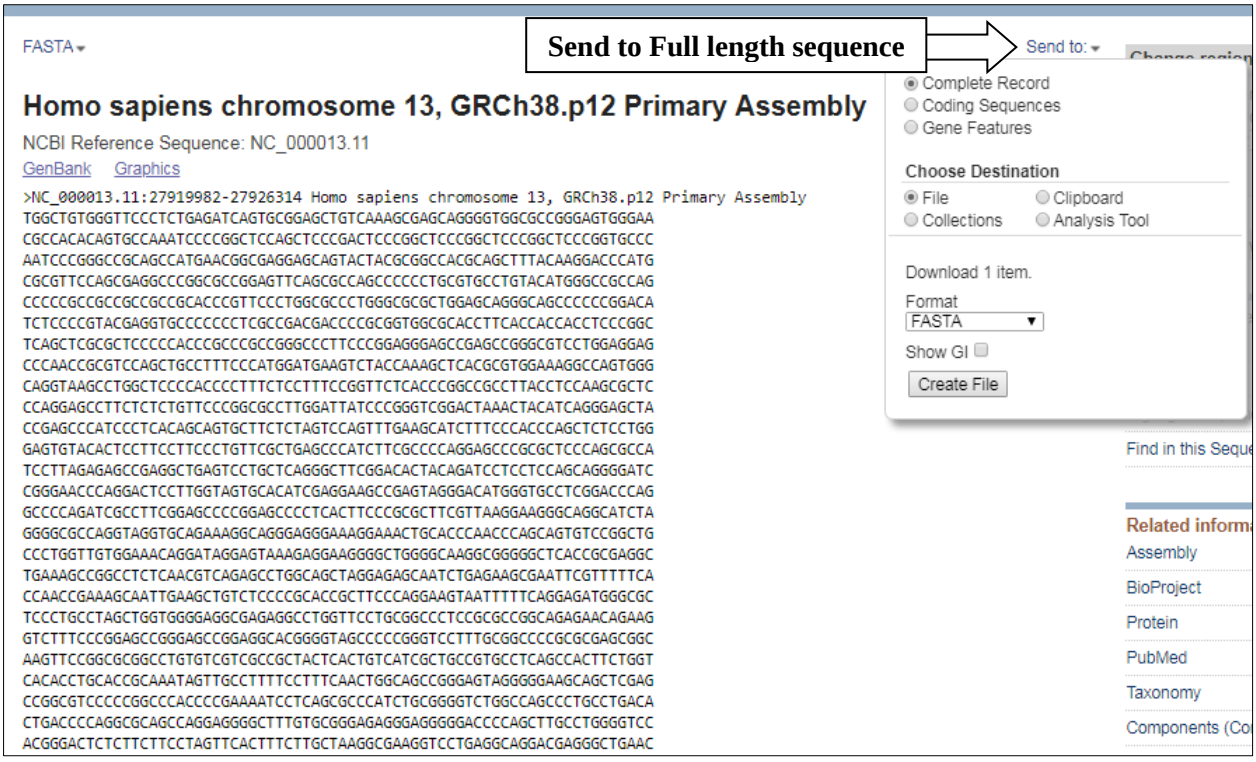


Figure 14: Downloading IPF-1 full length fasta sequence from NCBI

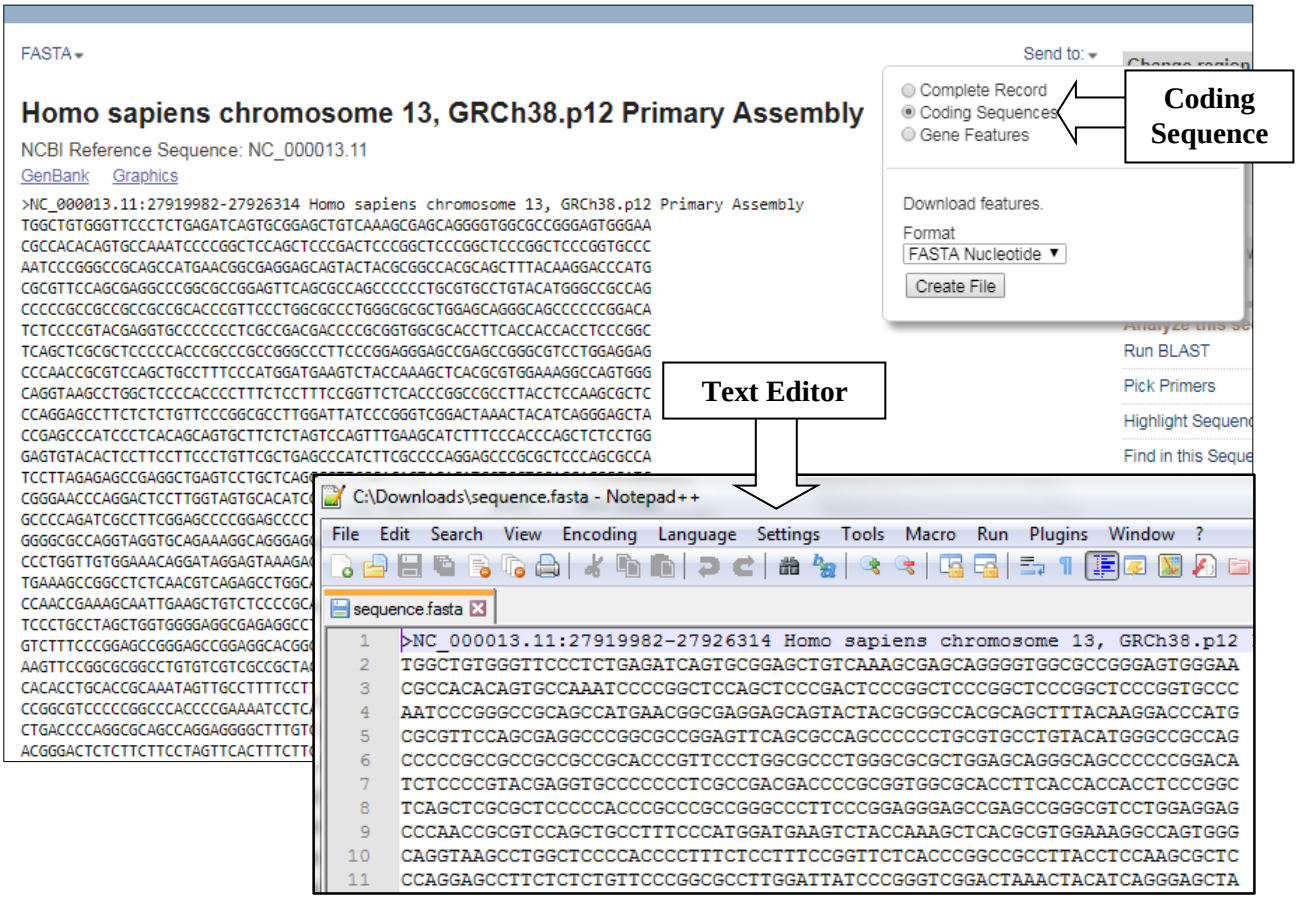


Figure 15: Downloading IPF-1 coding sequence in fasta format and saved in Text editor Notepad++

3.8.2. Sequence analysis and alignment

The full length and the coding sequence of IPF-1 (Insulin Promoter Factor 1); also known as PDX1 (pancreatic and duodenal homeobox 1) retrieved from the National Center for Biotechnology Information (NCBI) GenBank database and saved in .fasta format using Text Editor Notepad++ with the header "PDX1_Wild". The amino acid sequence of IPF1 protein was obtained using the software Molecular Evolutionary Genetic Analysis Tool MEGA X (64-bit).

The product length of C18R and R197H primer pair was identified from the retrieved sequences. Both of the reverse primers were converted to reverse complement using the online Reverse Complement tool (https://www.bioinformatics.org/sms/rev_comp.html). Forward and reverse primer sequences are highlighted with green color, in between them highlighted with yellow color indicates the primer product length.

Table 8: Full length nucleotide sequence of IPF-1 (PDX1) obtained from NCBI. Highlighted green area indicates Primer position and in between highlighted yellow indicates primer product length.

PDX1 (pancreatic and duodenal homeobox 1) Full Sequence

>NC_000013.11:27919982-27926314 Homo sapiens chromosome 13, GRCh38.p12 Primary Assembly

TGGCTGTGGGTTCCCTCTGAGATCAGTGCGGAGCTGTCAAAGCGAGCAGGGGTGGCGCCGGGAGTGGGAA
CGCCACACAGTGCCAAATCCCCGGCTCCAGCTCCCGACTCCCGGCTCCCGGCTCCCGGCTCCCGGTGCCC
AATCCCGGGCCGCAGC**CATGAACGGCGAGGAGCAG**TACTACGCGGCCACGCAGCTTTACAAGGACCCATG
CGCGTTCCAGCGAGGCCCGGCGCCGGAGTTCAGCGCCAGCCCCCCTGCGTGCCTGTACATGGGCCGCCAG
CCCCCGCCGCCGCCGCCGCACCCGTTCCCTGGCGCCCTGGGCGCGCTGGAGCAGGGCAGCCCCCCGGACA
TCTCCCCGTACGAGGTGCCCCCCCTCGCCGACGACCCCGCGGTGGCGCACCTTCACCACCACCTCCCGGC
TCAGCTCGCGCTCCCCCACCCGCCCGCCGGGCCCTTCCCGGAGGGAGCCGAGCCGGGCGTCCTGGAGGAG
CCCAACCGCGTCCAGCTGCCTTTCCCATGGATGAAGTCTACCAAAGCTCACGCGTGGAAAGGCCAGTGGG
CAGGTAAGCCTGGCTCCCCACCCCTTTCTCCTTTCCGGTTCTCACCCGGCCGCCTTACCTCCAAGCGCTC
CCAGGAGCCTTCTCTCTGTTCCCGGCGCCTTGGATTATCCCGGGTCGGACTAAACTACATCAGGGAGCTA
CCGAGCCCATCCCTCACAGCAGTGCTTCTCTAGTCCAGTTTGAAGCATCTTTCCCACCCAGCTCTCCTGG
GAGTGTACACTCCTTCCTTCCCTGTTCGCTGAGCCCATCTTCGCCCCAGGAGCCCGCGCTCCCAGCGCCA
TCCTTAGAGAGCCGAGGCTGAGTCCTGCTCAGGGCTTCGGACACTACAGATCCTCCTCCAGCAGGGGATC
CGGGAACCCAGGACTCCTTGGTAGTGCACATCGAGGAAGCCGAGTAGGGACATGGGTGCCTCGGACCCAG
GCCCCAGATCGCCTTCGGAGCCCCGGAGCCCCTCACTTCCCGCGCTTCGTTAAGGAAGGGCAGGCATCTA
GGGGCGCCAGGTAGGTGCAGAAAGGCAGGGAGGGAAAGGAAACTGCACCCAACCCAGCAGTGTCCGGCTG
CCCTGGTTGTGGAAACAGGATAGGAGTAAAGAGGAAGGGGCTGGGGCAAGGCGGGGGCTCACCGCGAGGC
TGAAAGCCGGCCTCTCAACGTCAGAGCCTGGCAGCTAGGAGAGCAATCTGAGAAGCGAATTCGTTTTTCA
CCAACCGAAAGCAATTGAAGCTGTCTCCCCGCACCGCTTCCCAGGAAGTAATTTTTCAGGAGATGGGCGC
TCCCTGCCTAGCTGGTGGGGAGGCGAGAGGCCTGGTTCCTGCGGCCCTCCGCGCCGGCAGAGAACAGAAG
GTCTTTCCCGGAGCCGGGAGCCGGAGGCACGGGGTAGCCCCCGGGTCCTTTGCGGCCCCGCGCGAGCGGC
AAGTTCCGGCGCGGCCTGTGTCGTCGCCGCTACTCACTGTCATCGCTGCCGTGCCTCAGCCACTTCTGGT
CACACCTGCACCGCAAATAGTTGCCTTTTCCTTTCAACTGGCAGCCGGGAGTAGGGGGAAGCAGCTCGAG
CCGGCGTCCCCCGGCCCACCCCGAAAATCCTCAGCGCCCATCTGCGGGGTCTGGCCAGCCCTGCCTGACA
CTGACCCCAGGCGCAGCCAGGAGGGGCTTTGTGCGGGAGAGGGAGGGGGACCCCAGCTTGCCTGGGGTCC
ACGGGACTCTCTTCTTCCTAGTTCACTTTCTTGCTAAGGCGAAGGTCCTGAGGCAGGACGAGGGCTGAAC
TGCGCTGCAATCGTCCCCACCTCCAGCGAAACCCAGTTGACAGGGGCGCCAGAAGCTGCCGCGGCGCCTC
TGCAAATTTATCCAGCTCGCGCAGCCCGGGCCAAAGGCCTTGAAGTCTCCGGAAATGCGGGGTTCTTAGG
AGGCGGGAGGACAGTCCCTCGAACAAAGGTGGGGGGCTCCTCGTCCTCACCCAGTTTTCTTCCAGGGCTG
CCTCCCCTCCAGACCTCTCTTCTGGCCTCCTAGGCCCTCGGAGCTCCTGCTTTCCCACCCTGGGCCTTCC
TCAGGAAATGGGCGACATCAGGGTCCCGAAAGAGGATTTGTGAGGTGGAGTAACTTCCCTATCCCAACCC
AAGGGGTGATACCTCTGCTCTGGAGGACTTGGGCTTAGGCTGACCCAAGAAGCCAGAAAGTAAAACCAGA
AGGCAAATCAGCAGCCTTGGCGAGGGTTCGGGGACCCAAGGAGGGCGACACTCTCGGGCTGGAGTTGGCC
CCAGGCCTTTGCTGGCGCCCTCTAACCCGCTGCATGCTCGACTCTCGGGGAAGGAGACGACCTCCCCTCT
CTTCCCCTGGAAGCCGTCTGCGGGGCCGGCTGCTATCCCCGCGTTCCTCTAGGGGAAACTTCGATGGAGC
CGAAATTCAAAAATTGCAAACCCACCTGCCCCTGGGAAGAGCGAAGTGACAAAAGGGCTCTCACTGGCAG
TACGAATCTGAATGCTAATGACAACAGAGGTTTTGAAAAACATTGACCCCCAAATGCTTCAGCAGCGCTG

TCCAGCTGGCACCTAAACTGCATCACTCTGCGCCTTGGGGAAGGGCCCAGGCTTGGCGACCTTGACCTTT
TCCCACCATCCTCAACCTCCACCCCTGCCGCGTCGCGCTGAGCACAGGTCCCCCGGGAATAGTGCACCCC
AGGAAGTCTCTCCCTGAGCAGTCTCTCGCAGGGACTTCACGAAGCCCTCTCGCAGGGACTATACGAAGCC
CGCAGCCTAAGGCAGGAACCCAGAGACATGTCGGTTTAATGTAAAAACTTTGGAGAGCCTTTCAAAATGT
TTATTGAAGGCCCGTCTCGCTTCTCTCCCAGGCGTGGGATGCCAGGTAGATTCGGGGATGCCCCCAGGGA
GTAGAACTCTCCCTGGACTAGGGTTTGAGCCTCTGCTTCAGCTTCTGGCGCCTCTTCTCGACCTGGGGGG
AAACCCAGTAGGGTTCCATCGCGAAATTAAACCCGCCCCCAACACACACACACTCGCCTTTCAATTCCTT
AAGGCTTAGCCAACATTCACAGGAGAAATGTCCCCTGCCTTTGCTCTAAGACAAGCCTCTCCCCGGAACT
TTGGTGGAACTTCCCGCGCCAGCGTCCACAGCCTGGGTGCAGTCAGTATTTTCCACAGAAAAGAAAAGAT
TGGGACCTGGCTGAGCGCAGCGGCAAACAGTGAATGTGGGTCTCCAACCTCCTGGGCCAGGGCGTCCTGT
TGCCTCTTGGAGACACGAGAGGCTTGTTTCTGCACCACTACCACCTCCTCCGTAGGGCTGTCGGTTCTGC
AGCTGGGCTAGGGCCCCTGTGTCTCCCCTCAACACCTCTGAGGGCATTTGGGATCCAGGGCGTAGAGTCT
GGAGCTGCCAGAGTTCTGCCCTGGCCAACGTGACCCCCAGAACAATATTCCTTCACTTCGCGGGCAGAAG
TCCGGCTGAAGTTAAAACAATTATGGAGAATTTGCTGGCTCTCAGGTTGGGACTAATTACGATATAACTA
TAGAGAGAGGAAACACATGGTCAGATATAACAAAATGTGTCACAGTCTCCATTAGCACAAAGATTTTCAA
ACTGCAGGTTGCACCCATTCGCAGGTCATAAAATCAATTTACTAGGTTGAGATTAGTATTTTTTAAACGA
AATAGCAGATAATGGAGAGAAAAGTAGATAGCATCATACGTGGTAAACGTTTGTTTTATGTCCTTAAGAT
TTGTCAGTATAACTGACCTGCAGTGTCCGTGTGTGAACTACACAACGATCCGAAATGTATTTCTCACATT
TGTGGGTCACCATCAGGAGGTTTTTTTTAGCCCTGGATTAAAGGCGTTTTATTGCCTTTGTAGGATCCAG
CCGGTTTAACTATTATATACATTTAAATCAACATCAATCAGTTGATTAACACCGATTATATGAGCGCATT
CAAGGACCACTCATTGGCAGAGCCAAGCTTAGGCTCACGGCGAGAGCTGACTCGAGTTTGGTCTCCAATA
AAAAGGCTATCTTTATTAGGAAGGGCTTGAGTTACTAGGGAAGAGCTTCGCGCGCCTACACTAGGCGCTG
AAATGGGATGCTGGGGCTTGGTGGCTCCGGCGGGAGCAGCTGGTAGGGCTAGGGCTCCCTGGCCCCCCTT
GAAGGGGTTGGGCTGCGTGGGTGGGGGCTGTGCGGGGCTCCGGGGGCCACACTCACGCCCTGTGTCGCCC
GCAGGCGGCGCCTACGCTGCGGAGCCGGAGGAGAACAAGCGGACGCGCACGGCCTACACGCGCGCACAGC
TGCTAGAGCTGGAGAAGGAGTTCCTATTCAACAAGTACATCTCACGGCCGCGCCG**GGTGGAGCTGGCTGT**
CATGTTGAACTTGACCGAGAGACACATCAAGATCTGGTTCCAAAACCGCCGCATGAAGTGGAAAAAGGAG
GAGGACAAGAAGCGCGGCGGCGGGACAGCTGTCGGGGGTGGCGGGGTCGCGGAGCCTGAGCAGGACTGCG
CCGTGACCTCCGGCGAGGAGCTTCTGGCGCTGCCGCCGCCGCCGCCCCCCGGAGGTGCTGTGCCGCCCGC
TGCCCCCGTTGCCGCCCGAGAGGGCCGCCTGCCGCCTGGCCTTAG**CGCGTCGCCACAGCCCT**CCAGCGTC
GCGCCTCGGCGGCCGCAGGAACCACGATGAGAGGCAGGAGCTGCTCCTGGCTGAGGGGCTTCAACCACTC
GCCGAGGAGGAGCAGAGGGCCTAGGAGGACCCCGGGCGTGGACCACCCGCCCTGGCAGTTGAATGGGGCG
GCAATTGCGGGGCCCACCTTAGACCGAAGGGGAAAACCCGCTCTCTCAGGCGCATGTGCCAGTTGGGGCC
CCGCGGGTAGATGCCGGCAGGCCTTCCGGAAGAAAAAGAGCCATTGGTTTTTGTAGTATTGGGGCCCTCT
TTTAGTGATACTGGATTGGCGTTGTTTGTGGCTGTTGCGCACATCCCTGCCCTCCTACAGCACTCCACCT
TGGGACCTGTTTAGAGAAGCCGGCTCTTCAAAGACAATGGAAACTGTACCATACACATTGGAAGGCTCCC
TAACACACACAGCGGGGAAGCTGGGCCGAGTACCTTAATCTGCCATAAAGCCATTCTTACTCGGGCGACC
CCTTTAAGTTTAGAAATAATTGAAAGGAAATGTTTGAGTTTTCAAAGATCCCGTGAAATTGATGCCAGTG
GAATACAGTGAGTCCTCCTCTTCCTCCTCCTCCTCTTCCCCCTCCCCTTCCTCCTCCTCCTCTTCTTTTC
CCTCCTCTTCCTCTTCCTCCTGCTCTCCTTTCCTCCCCCTCCTCTTTTCCCTCCTCTTCCTCTTCCTCCT
GCTCTCCTTTCCTCCCCCTCCTCTTTCTCCTCCTCCTCCTCTTCTTCCCCCTCCTCTCCCTCCTCCTCTT
CTTCCCCCTCCTCTCCCTCCTCCTCTTCTTCTCCCTCCTCTTCCTCTTCCTCCTCTTCCACGTGCTCTCC
TTTCCTCCCCCTCCTCTTGCTCCCCTTCTTCCCCGTCCTCTTCCTCCTCCTCCTCTTCTTCTCCCTCCTC
TTCCTCCTCCTCTTTCTTCCTGACCTCTTTCTTTCTCCTCCTCCTCCTTCTACCTCCCCTTCTCATCCCT
CCTCTTCCTCTTCTCTAGCTGCACACTTCACTACTGCACATCTTATAACTTGCACCCCTTTCTTCTGAGG
AAGAGAACATCTTGCAAGGCAGGGCGAGCAGCGGCAGGGCTGGCTTAGGAGCAGTGCAAGAGTCCCTGTG
CTCCAGTTCCACACTGCTGGCAGGGAAGGCAAGGGGGGACGGGCCTGGATCTGGGGGTGAGGGAGAAAGA
TGGACCCCTGGGTGACCACTAAACCAAAGATATTCGGAACTTTCTATTTAGGATGTGGACGTAATTCCTG
TTCCGAGGTAGAGGCTGTGCTGAAGACAAGCACAGTGGCCTGGTGCGCCTTGGAAACCAACAACTATTCA
CGAGCCAGTATGACCTTCACATCTTTAGAAATTATGAAAACGTATGTGATTGGAGGGTTTGGAAAACCAG
TTATCTTATTTAACATTTTAAAAATTACCTAACAGTTATTTACAAACAGGTCTGTGCATCCCAGGTCTGT
CTTCTTTTCAAGGTCTGGGCCTTGTGCTCGGGTTATGTTTGTGGGAAATGCTTAATAAATACTGATAATA
TGGGAAGAGATGAAAACTGATTCTCCTCACTTTGTTTCAAACCTTTCTGGCAGTGGGATGATTCGAATTC
ACTTTTAAAATTAAATTAGCGTGTTTTGTTTTG

Table 9: Coding nucleotide sequence of IPF-1 (PDX1) obtained from NCBI. Highlighted green area indicates Primer position and in between highlighted yellow indicates primer product length.

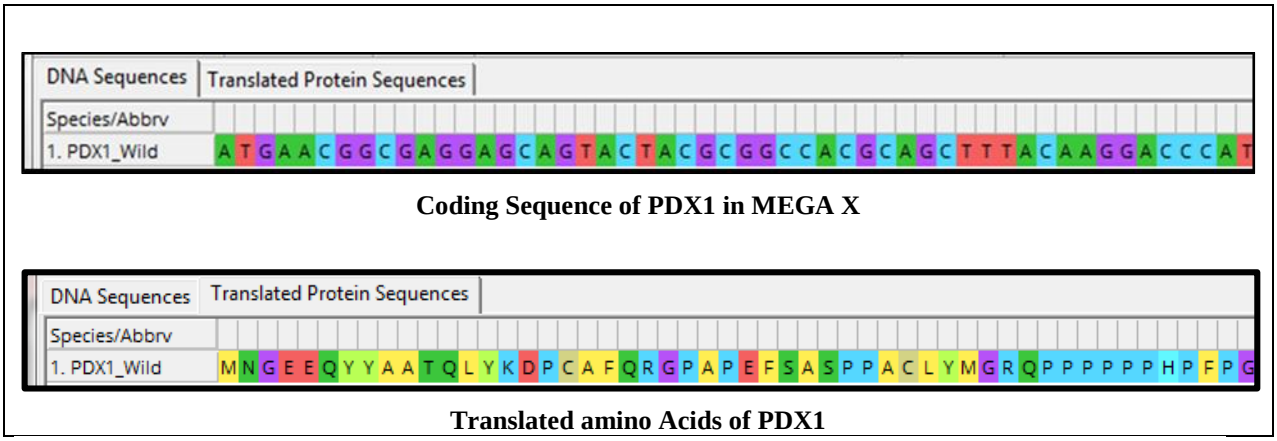
PDX1 (pancreatic and duodenal homeobox 1) Coding Sequence

>lcl|NC_000013.11_cds_NP_000200.1_1 [gene=PDX1] [db_xref=CCDS:CCDS9327.1,GeneID:3651] [protein=pancreas/duodenum homeobox protein 1] [protein_id=NP_000200.1] [location=join(27920139..27920544,27924256..27924701)] [gbkey=CDS]

ATGAACGGCGAGGAGCAGTACTACGCGGCCACGCAGCTTTACAAGGACCCATGCGCGTTCCAGCGAGGCC
CGGCGCCGGAGTTCAGCGCCAGCCCCCCTGCGTGCCTGTACATGGGCCGCCAGCCCCCGCCGCCGCCGCC
GCACCCGTTCCCTGGCGCCCTGGGCGCGCTGGAGCAGGGCAGCCCCCCGGACATCTCCCCGTACGAGGTG
CCCCCCCTCGCCGACGACCCCGCGGTGGCGCACCTTCACCACCACCTCCCGGCTCAGCTCGCGCTCCCCC
ACCCGCCCGCCGGGCCCTTCCCGGAGGGAGCCGAGCCGGGCGTCCTGGAGGAGCCCAACCGCGTCCAGCT
GCCTTTCCCATGGATGAAGTCTACCAAAGCTCACGCGTGGAAAGGCCAGTGGGCAGGCGGCGCCTACGCT
GCGGAGCCGGAGGAGAACAAGCGGACGCGCACGGCCTACACGCGCGCACAGCTGCTAGAGCTGGAGAAGG
AGTTCCTATTCAACAAGTACATCTCACGGCCGCGCCGGGTGGAGCTGGCTGTCATGTTGAACTTGACCGA
GAGACACATCAAGATCTGGTTCCAAAACCGCCGCATGAAGTGGAAAAAGGAGGAGGACAAGAAGCGCGGC
GGCGGGACAGCTGTCGGGGGTGGCGGGGTCGCGGAGCCTGAGCAGGACTGCGCCGTGACCTCCGGCGAGG
AGCTTCTGGCGCTGCCGCCGCCGCCGCCCCCCGGAGGTGCTGTGCCGCCCGCTGCCCCCGTTGCCGCCCG
AGAGGGCCGCCTGCCGCCTGGCCTTAGCGCGTCGCCACAGCCCTCCAGCGTCGCGCCTCGGCGGCCGCAG
GAACCACGATGA

3.8.2.1. Translation of Nucleotide Sequence

Amino acid sequence was obtained from IPF-1 or PDX1 nucleotide sequence by MEGA X alignment explorer Translation tool.



Coding Sequence of PDX1 in MEGA X

Translated amino Acids of PDX1

Figure 16: Translation of the Coding sequences into Amino acid sequence using MEGA X.

Table 10: Translated amino acid sequence of IPF-1

MNGEEQYYAATQLYKDPCAFQRGPAPEFSASPPACLYMGRQPPPPPPHPFPGALGALEQGSPPDISPYEVPPLADDPAVAHLHHHLPAQLAL PHPPAGPFPEGAEPGVLEEPNRVQLPFPWMKSTKAHAWKGQWAGGAYAAEPEENKRTRTAYTRAQLLELEKEFLFNKYISRPRRVELAVML NLTERHIKIWFQNRRMKWKKEEDKKRGGGTAVGGGGVAEPEQDCAVTSGEELLALPPPPPPGGAVPPAAPVAAREGRLPPGLSASPQPSSV APRRPQEPR

Mutation position C18R and R197H was find out from the obtained amino acid sequence by DNA/RNA protein sequence character counter C-Count v 2.0(http://www.biogem.org/cgibin/count.pl).

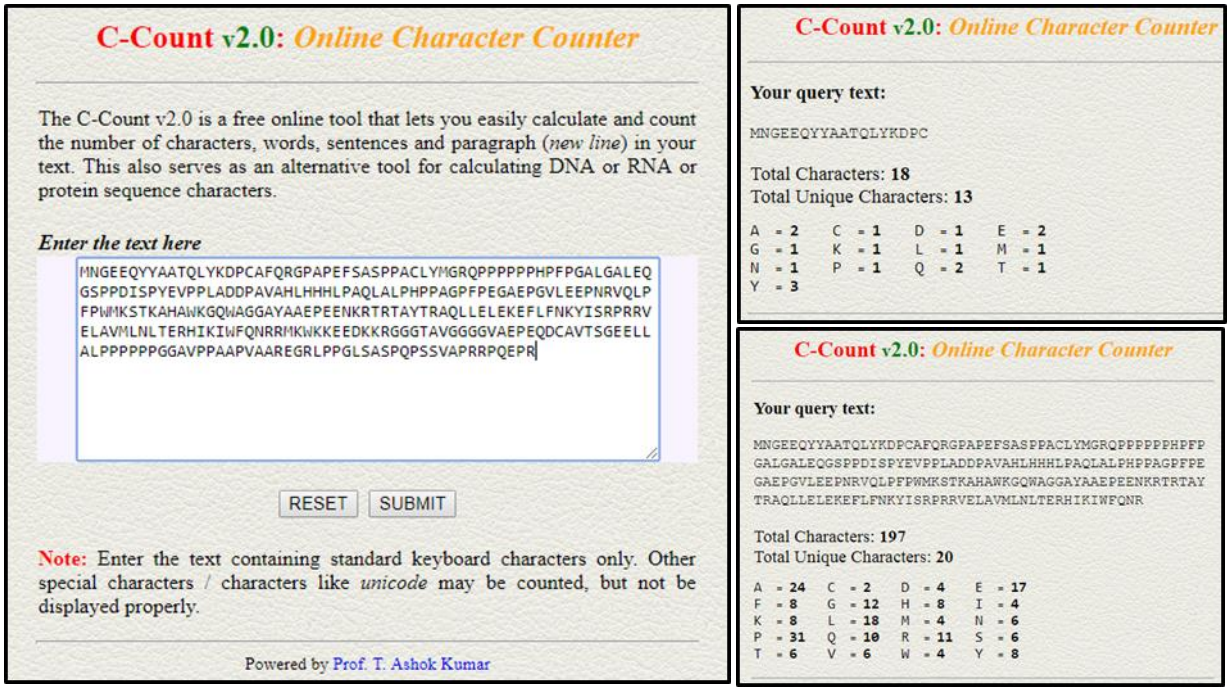


Figure 17: DNA/RNA protein sequence character counter C-Count v 2.0

Table 11: DNA/RNA protein sequence character counter C-Count v 2.0 used to locate the mutational position C18R and R197H.

C18R

MNGEEQYYAATQLYKDPRAFQRGPAPEFSASPPACLY
MGRQPPPPPPHPFPGALGALEQGSPPDISPYEVPPLAD
DPAVAHLHHHLPAQLALPHPPAGPFPEGAEPGVLEEPN
RVQLPFPWMKSTKAHAWKGQWAGGAYAAEPEENKR
TRTAYTRAQLLELEKEFLFNKYISRPRRVELAVMLNLTER
HIKIWFQNRRMKWKKEEDKKRGGGTAVGGGGVAEPE
QDCAVTSGEELLALPPPPPPGGAVPPAAPVAAREGRLP
PGLSASPQPSSVAPRRPQEPR

Mutation position C18R

R197H

MNGEEQYYAATQLYKDPCAFQRGPAPEFSASPPACLY
MGRQPPPPPPHPFPGALGALEQGSPPDISPYEVPPLAD
DPAVAHLHHHLPAQLALPHPPAGPFPEGAEPGVLEEPN
RVQLPFPWMKSTKAHAWKGQWAGGAYAAEPEENKR
TRTAYTRAQLLELEKEFLFNKYISRPRRVELAVMLNLTER
HIKIWFQNHRMKWKKEEDKKRGGGTAVGGGGVAEPE
QDCAVTSGEELLALPPPPPPGGAVPPAAPVAAREGRLP
PGLSASPQPSSVAPRRPQEPR

Mutation position R197H

>PDX1_W
MNGEEQYYAATQLYKDPCAFQRGPAPEFSASPPACLYMGRQPPPPPPHPFPGALGALEQGSPPDISPYEVPPLADDPAVAHLHHHLP
AQLALPHPPAGPFPEGAEPGVLEEPNRVQLPFPWMKSTKAHAWKGQWAGGAYAAEPEENKRTRTAYTRAQLLELEKEFLFNKYISRP
RRVELAVMLNLTERHIKIWFQNRRMKWKKEEDKKRGGGTAVGGGGVAEPEQDCAVTSGEELLALPPPPPPGGAVPPAAPVAAREG
RLPPGLSASPQPSSVAPRRPQEPR

PDX1 wild Fasta amino acids sequence

>PDX1_M
MNGEEQYYAATQLYKDPRAFQRGPAPEFSASPPACLYMGRQPPPPPPHPFPGALGALEQGSPPDISPYEVPPLADDPAVAHLHHHLP
AQLALPHPPAGPFPEGAEPGVLEEPNRVQLPFPWMKSTKAHAWKGQWAGGAYAAEPEENKRTRTAYTRAQLLELEKEFLFNKYISRP
RRVELAVMLNLTERHIKIWFQNHRMKWKKEEDKKRGGGTAVGGGGVAEPEQDCAVTSGEELLALPPPPPPGGAVPPAAPVAAREG
RLPPGLSASPQPSSVAPRRPQEPR

PDX1 mutated Fasta amino acids sequence

3.8.2.2. Amino Acid Sequence alignment

Amino acid sequence PDX1_W and PDX1_M saved in Fasta format and aligned using MEGA X to verify the mutation position C18R and R197H.

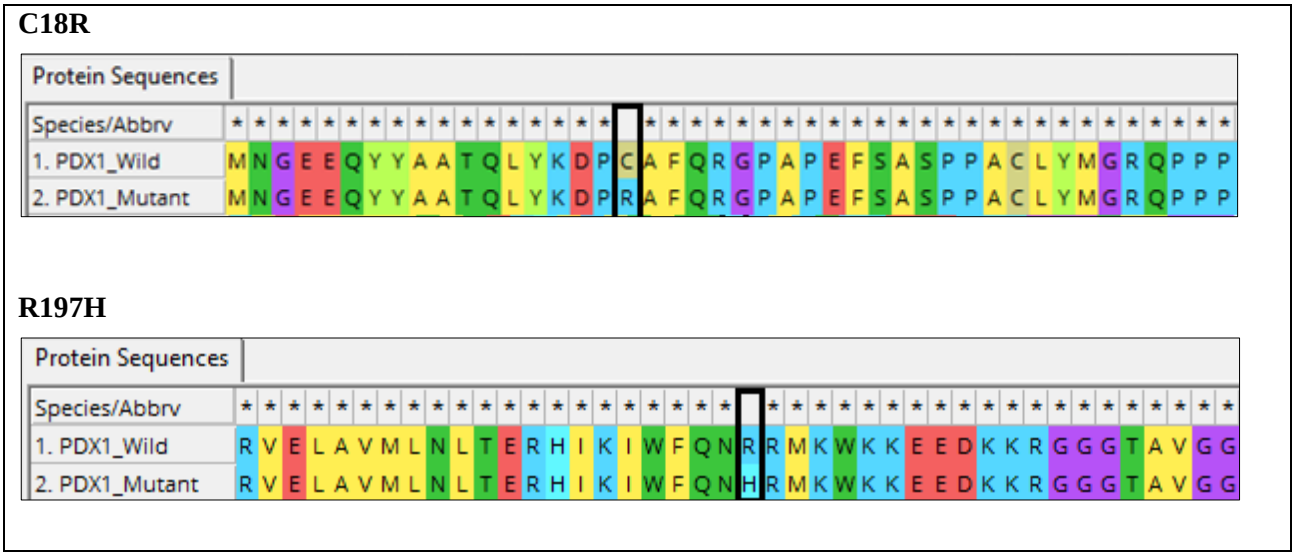


Figure 18: Translation & alignment of the Coding sequences into Amino acid sequence using MEGA X.

3.8.3. Constructing alignments of coding DNA from aligned amino acid sequences

The RevTrans 2.0b Server (http://www.cbs.dtu.dk/services/RevTrans-2.0/web/) was used virtually translate the sequences and align the resulting peptide sequences using MAFFT method with default settings. Finally, RevTrans constructs a multiple DNA alignment using the peptide alignment as a scaffold. The translation process is done by default using the Standard Genetic Code (alternative translation tables can be selected) and has support for the full IUPAC alphabet of degenerate nucleotides.

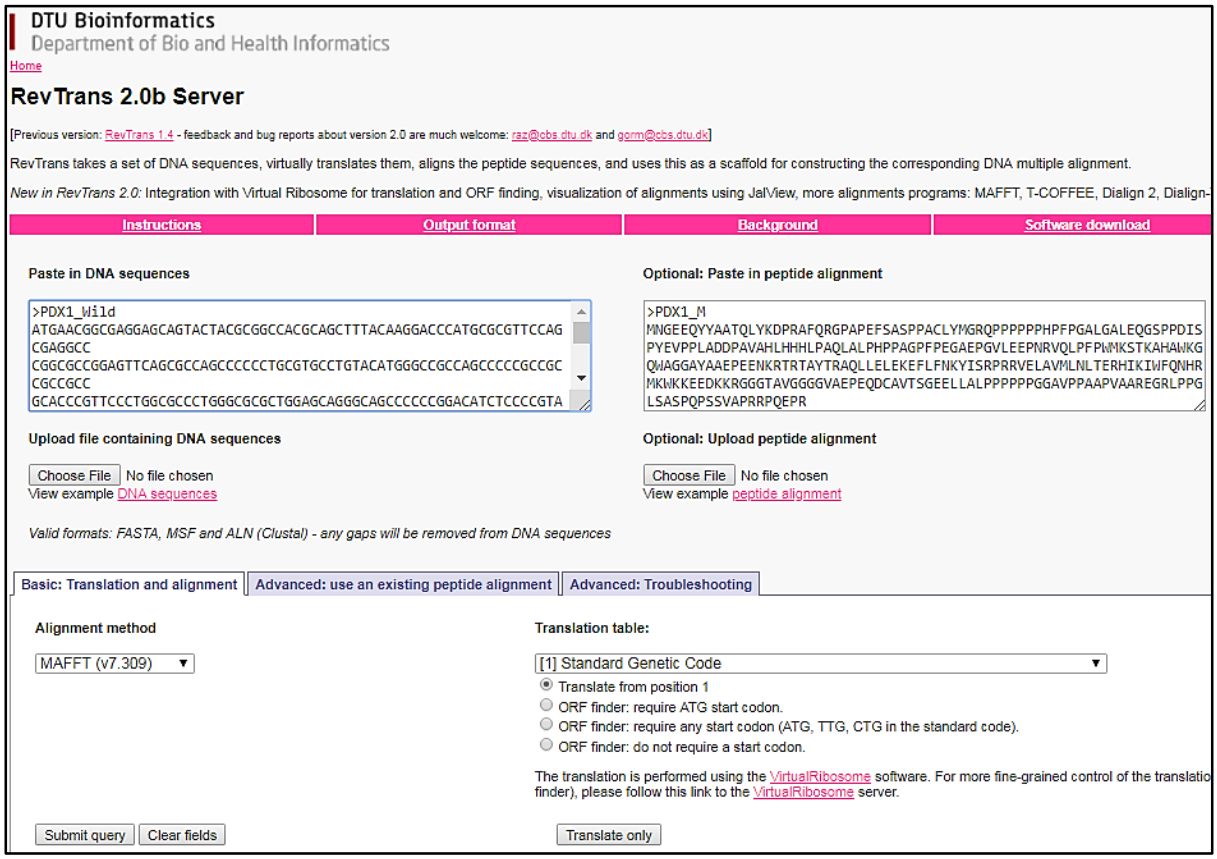


Figure 19: RevTrans 2.0b Server virtual translation tool.

This alignment was made through the RevTrans server to find out C18R (T/C) and R197H (G/A) position from the amino acids and corresponding nucleotides alignment. RevTrans takes a set of DNA sequences, virtually translates them, aligns the peptide sequences, and uses this as a scaffold for constructing the corresponding DNA multiple alignments.

Table 12: IPF-1_Wild amino acid sequence and corresponding nucleotide sequence translation table by RevTrans 2.0b Server.

```
VIRTUAL RIBOSOME
----------------
Translation table: Standard SGC0

>>PDX1_W_AASeq
Reading frame: 1

    M  N  G  E  E  Q  Y  Y  A  A  T  Q  L  Y  K  D  P  C  A  F  Q  R  G  P  A  P  E  F  S  A
5' ATGAACGGCGAGGAGCAGTACTACGCGGCCACGCAGCTTTACAAGGACCCATGCGCGTTCCAGCGAGGCCCGGCGCCGGAGTTCAGCGCC 90
   >>>.....................................................................................

    S  P  P  A  C  L  Y  M  G  R  Q  P  P  P  P  P  P  H  P  F  P  G  A  L  G  A  L  E  Q  G
5' AGCCCCCCTGCGTGCCTGTACATGGGCCGCCAGCCCCCGCCGCCGCCGCCGCACCCGTTCCCTGGCGCCCTGGGCGCGCTGGAGCAGGGC 180
   ..............)))...>>>.............................................)))......)))........

    S  P  P  D  I  S  P  Y  E  V  P  P  L  A  D  D  P  A  V  A  H  L  H  H  H  L  P  A  Q  L
5' AGCCCCCCGGACATCTCCCCGTACGAGGTGCCCCCCCTCGCCGACGACCCCGCGGTGGCGCACCTTCACCACCACCTCCCGGCTCAGCTC 270
   .....................................................................................

    A  L  P  H  P  P  A  G  P  F  P  E  G  A  E  P  G  V  L  E  E  P  N  R  V  Q  L  P  F  P
5' GCGCTCCCCCACCCGCCCGCCGGGCCCTTCCCGGAGGGAGCCGAGCCGGGCGTCCTGGAGGAGCCCAACCGCGTCCAGCTGCCTTTCCCA 360
   ..........................................................)))......................)))........

    W  M  K  S  T  K  A  H  A  W  K  G  Q  W  A  G  G  A  Y  A  A  E  P  E  E  N  K  R  T  R
5' TGGATGAAGTCTACCAAAGCTCACGCGTGGAAAGGCCAGTGGGCAGGCGGCGCCTACGCTGCGGAGCCGGAGGAGAACAAGCGGACGCGC 450
   ...>>>.....................................................................................

    T  A  Y  T  R  A  Q  L  L  E  L  E  K  E  F  L  F  N  K  Y  I  S  R  P  R  R  V  E  L  A
5' ACGGCCTACACGCGCGCACAGCTGCTAGAGCTGGAGAAGGAGTTCCTATTCAACAAGTACATCTCACGGCCGCGCCGGGTGGAGCTGGCT 540
   ..................)))......)))..........................................................)))...

    V  M  L  N  L  T  E  R  H  I  K  I  W  F  Q  N  R  R  M  K  W  K  K  E  E  D  K  K  R  G
5' GTCATGTTGAACTTGACCGAGAGACACATCAAGATCTGGTTCCAAAACCGCCGCATGAAGTGGAAAAAGGAGGAGGACAAGAAGCGCGGC 630
   ...>>>)))...)))..........................................>>>.............................

    G  G  T  A  V  G  G  G  G  V  A  E  P  E  Q  D  C  A  V  T  S  G  E  E  L  L  A  L  P  P
5' GGCGGGACAGCTGTCGGGGGTGGCGGGGTCGCGGAGCCTGAGCAGGACTGCGCCGTGACCTCCGGCGAGGAGCTTCTGGCGCTGCCGCCG 720
   ..........................................................................)))...)))......

    P  P  P  P  G  G  A  V  P  P  A  A  P  V  A  A  R  E  G  R  L  P  P  G  L  S  A  S  P  Q
5' CCGCCGCCCCCCGGAGGTGCTGTGCCGCCCGCTGCCCCCGTTGCCGCCCGAGAGGGCCGCCTGCCGCCTGGCCTTAGCGCGTCGCCACAG 810
   ..............................................................)))............................

    P  S  S  V  A  P  R  R  P  Q  E  P  R
5' CCCTCCAGCGTCGCGCCTCGGCGGCCGCAGGAACCACGATGA 852
   ....................................
```

Table 13: IPF-1_Mutant amino acid sequence and corresponding nucleotide sequence translation table by RevTrans 2.0b Server.

```
VIRTUAL RIBOSOME
----------------
Translation table: Standard SGC0

>>PDX1_M_AASeq
Reading frame: 1

    M  N  G  E  E  Q  Y  Y  A  A  T  Q  L  Y  K  D  P  R  A  F  Q  R  G  P  A  P  E  F  S  A
5' ATGAACGGCGAGGAGCAGTACTACGCGGCCACGCAGCTTTACAAGGACCCACGCGCGTTCCAGCGAGGCCCGGCGCCGGAGTTCAGCGCC 90
   >>>.....................................................................................

    S  P  P  A  C  L  Y  M  G  R  Q  P  P  P  P  P  P  H  P  F  P  G  A  L  G  A  L  E  Q  G
5' AGCCCCCCTGCGTGCCTGTACATGGGCCGCCAGCCCCCGCCGCCGCCGCCGCACCCGTTCCCTGGCGCCCTGGGCGCGCTGGAGCAGGGC 180
   ..............)))...>>>.............................................)))......))).........

    S  P  P  D  I  S  P  Y  E  V  P  P  L  A  D  D  P  A  V  A  H  L  H  H  H  L  P  A  Q  L
5' AGCCCCCCGGACATCTCCCCGTACGAGGTGCCCCCCCTCGCCGACGACCCCGCGGTGGCGCACCTTCACCACCACCTCCCGGCTCAGCTC 270
   .....................................................................................

    A  L  P  H  P  P  A  G  P  F  P  E  G  A  E  P  G  V  L  E  E  P  N  R  V  Q  L  P  F  P
5' GCGCTCCCCCACCCGCCCGCCGGGCCCTTCCCGGAGGGAGCCGAGCCGGGCGTCCTGGAGGAGCCCAACCGCGTCCAGCTGCCTTTCCCA 360
   ..................................................)))......................)))..........

    W  M  K  S  T  K  A  H  A  W  K  G  Q  W  A  G  G  A  Y  A  A  E  P  E  E  N  K  R  T  R
5' TGGATGAAGTCTACCAAAGCTCACGCGTGGAAAGGCCAGTGGGCAGGCGGCGCCTACGCTGCGGAGCCGGAGGAGAACAAGCGGACGCGC 450
   ...>>>.....................................................................................

    T  A  Y  T  R  A  Q  L  L  E  L  E  K  E  F  L  F  N  K  Y  I  S  R  P  R  R  V  E  L  A
5' ACGGCCTACACGCGCGCACAGCTGCTAGAGCTGGAGAAGGAGTTCCTATTCAACAAGTACATCTCACGGCCGCGCCGGGTGGAGCTGGCT 540
   ......................)))......)))...........................................)))...

    V  M  L  N  L  T  E  R  H  I  K  I  W  F  Q  N  H  R  M  K  W  K  K  E  E  D  K  K  R  S
5' GTCATGTTGAACTTGACCGAGAGACACATCAAGATCTGGTTCCAAAACCACCGCATGAAGTGGAAAAAGGAGGAGGACAAGAAGCGCAGC 630
   ...>>>)))...)))..........................................>>>.............................

    G  G  T  A  V  G  G  G  G  V  A  E  P  E  Q  D  C  A  V  T  S  G  E  E  L  L  A  L  P  P
5' GGCGGGACAGCTGTCGGGGGTGGCGGGGTCGCGGAGCCTGAGCAGGACTGCGCCGTGACCTCCGGCGAGGAGCTTCTGGCGCTGCCGCCG 720
   ...........................................................................)))...)))......

    P  P  P  P  G  G  A  V  P  P  A  A  P  V  A  A  R  E  G  R  L  P  P  G  L  S  A  S  P  Q
5' CCGCCGCCCCCCGGAGGTGCTGTGCCGCCCGCTGCCCCCGTTGCCGCCCGAGAGGGCCGCCTGCCGCCTGGCCTTAGCGCGTCGCCACAG 810
   ...........................................................)))..........................

    P  S  S  V  A  P  R  R  P  Q  E  P  R
5' CCCTCCAGCGTCGCGCCTCGGCGGCCGCAGGAACCACGATGA 852
   ........................................
```

3.8.4. Physiochemical Properties

The ProtParam tool in the ExPASy server (http://web.expasy.org/protparam) was used to predict the physicochemical properties of PDX1. The parameters included: Molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, estimated half‑life, instability index, aliphatic index and grand average of hydropathy (GRAVY). To compute the parameters, we had submitted the amino acid sequence of IPF1 (wild) and the mutation inserted sequence, separately.

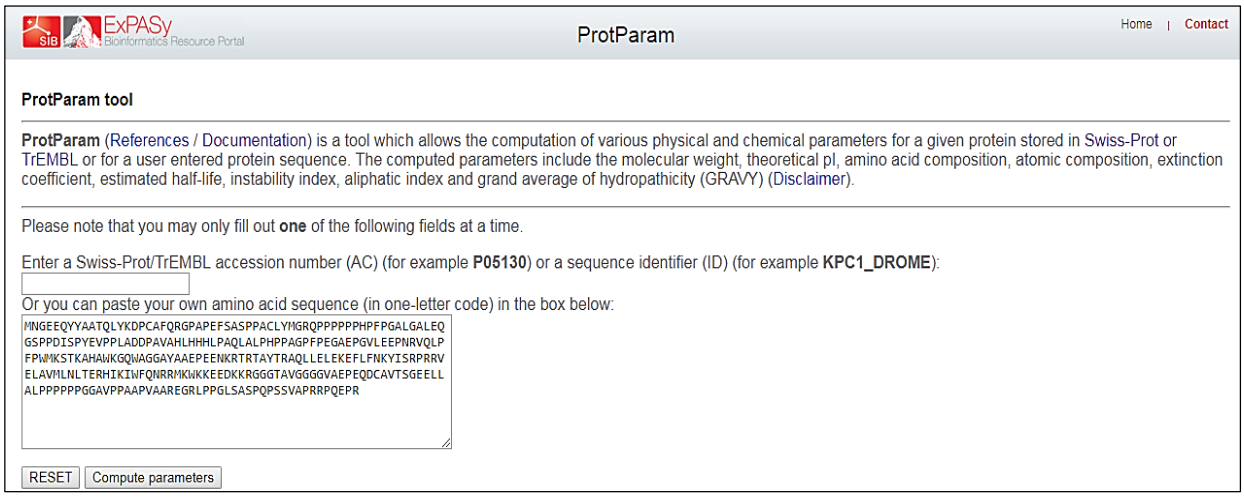


Figure 20: ProtParam Tool used to predict the physiochemical properties of IPF-1.

Table 14: ProtParam tool results for IPF-1_Wild amino acid sequence

Number of amino acids: 283
Molecular weight: 30770.99
Theoretical pI: 7.10

Amino acid composition:

Ala (A)	35	12.4%
Arg (R)	18	6.4%
Asn (N)	6	2.1%
Asp (D)	6	2.1%
Cys (C)	3	1.1%
Gln (Q)	13	4.6%
Glu (E)	25	8.8%
Gly (G)	24	8.5%
His (H)	8	2.8%
Ile (I)	4	1.4%
Leu (L)	23	8.1%
Lys (K)	13	4.6%
Met (M)	5	1.8%
Phe (F)	8	2.8%
Pro (P)	48	17.0%
Ser (S)	11	3.9%
Thr (T)	8	2.8%
Trp (W)	5	1.8%
Tyr (Y)	8	2.8%
Val (V)	12	4.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 31
Total number of positively charged residues (Arg + Lys): 31

Atomic composition:

Carbon	C	1385
Hydrogen	H	2128
Nitrogen	N	390
Oxygen	O	392
Sulfur	S	8

Formula: $C_{1385}H_{2128}N_{390}O_{392}S_8$
Total number of atoms: 4303

Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 39545
Abs 0.1% (=1 g/l) 1.285, assuming all pairs of Cys residues form cystines

Ext. coefficient 39420
Abs 0.1% (=1 g/l) 1.281, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 78.54
This classifies the protein as unstable.
Aliphatic index: 61.87
Grand average of hydropathicity (GRAVY): -0.671

Table 15: ProtParam tool results for IPF-1_Mutated amino acid sequence

Number of amino acids: 283
Molecular weight: 30804.99
Theoretical pI: 7.14

Amino acid composition:

Ala (A)	35	12.4%
Arg (R)	18	6.4%
Asn (N)	6	2.1%
Asp (D)	6	2.1%
Cys (C)	2	0.7%
Gln (Q)	13	4.6%
Glu (E)	25	8.8%
Gly (G)	24	8.5%
His (H)	9	3.2%
Ile (I)	4	1.4%
Leu (L)	23	8.1%
Lys (K)	13	4.6%
Met (M)	5	1.8%
Phe (F)	8	2.8%
Pro (P)	48	17.0%
Ser (S)	11	3.9%
Thr (T)	8	2.8%
Trp (W)	5	1.8%
Tyr (Y)	8	2.8%
Val (V)	12	4.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 31
Total number of positively charged residues (Arg + Lys): 31

Atomic composition:

Carbon	C	1388
Hydrogen	H	2130
Nitrogen	N	392
Oxygen	O	392
Sulfur	S	7

Formula: $C_{1388}H_{2130}N_{392}O_{392}S_7$
Total number of atoms: 4309

Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 39545
Abs 0.1% (=1 g/l) 1.284, assuming all pairs of Cys residues form cystines

Ext. coefficient 39420
Abs 0.1% (=1 g/l) 1.280, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 76.52
This classifies the protein as unstable.
Aliphatic index: 61.87
Grand average of hydropathicity (GRAVY): -0.691

3.9. Functional annotation of IPF1 protein

Based on sequence similarity searches with orthologous family members, several bioinformatics tools were used to analyze the IPF1 protein sequence. The UniProt Knowledgebase (UniProtKB) is a database of protein sequence and functional information. UniProtKB Peptide search tool (https://www.uniprot.org/peptidesearch/) was used to find out functional annotation (Transactivation domain, Binding domain) of the protein.

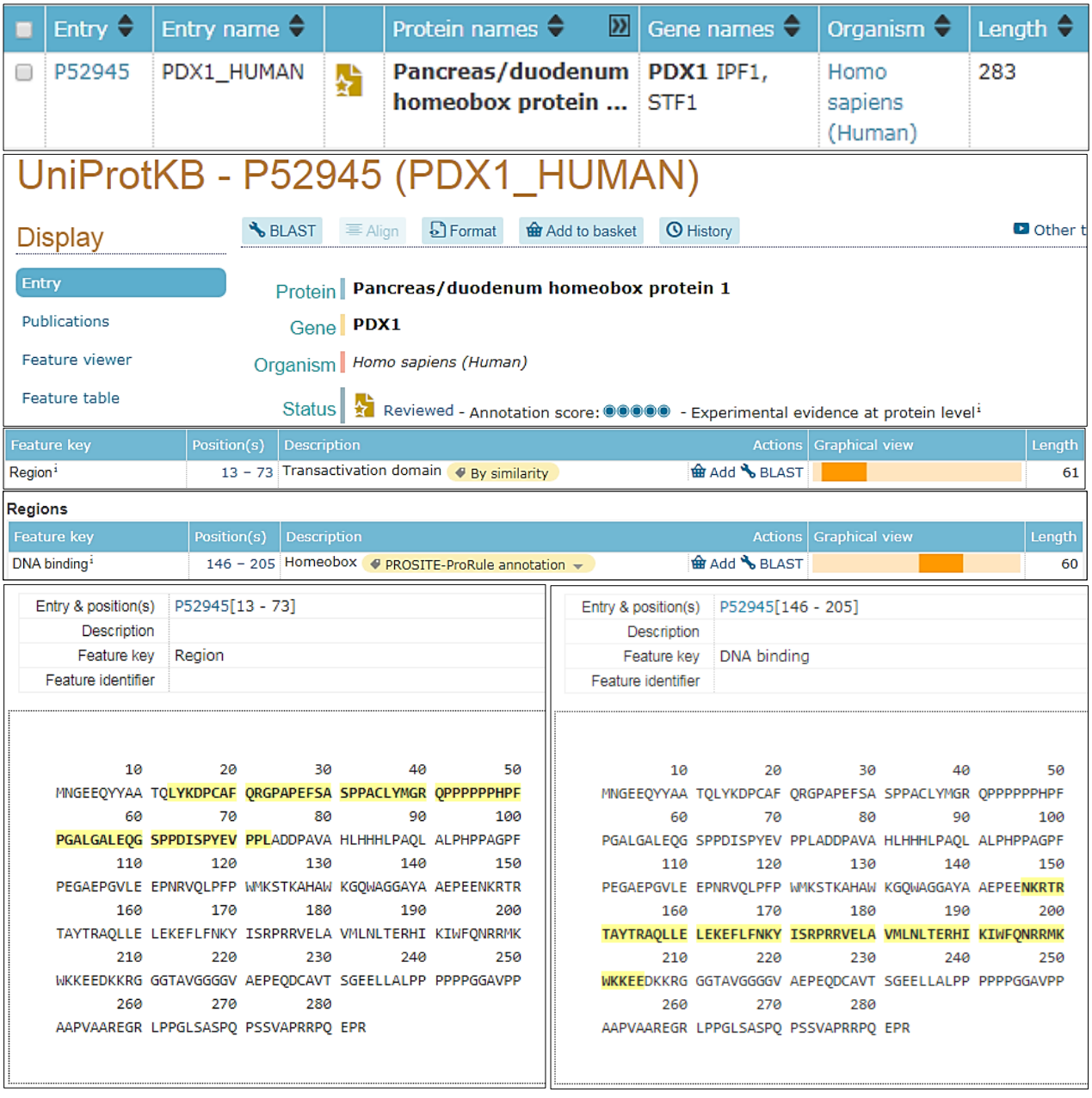


Figure 21: UniprotKB results of IPF-1 (PDX1) protein.

3.10. Finding out Binding Domain from CDD (Conserved Domain Database)

The NCBI Conserved Domains Database NCBI-CDD;(https://www.ncbi.nlm.nih.gov/cdd/wrpsb.cgi), which is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins (Marchler-Bauer *et.al.*, 2014).The NCBI-CDD shows a graphical summary of IPF1 conserved homeobox DNA binding site with specific base contacts. Curated CD Hierarchy cd00086 homeodomain 3D structure was viewed by Cn3D 4.3.

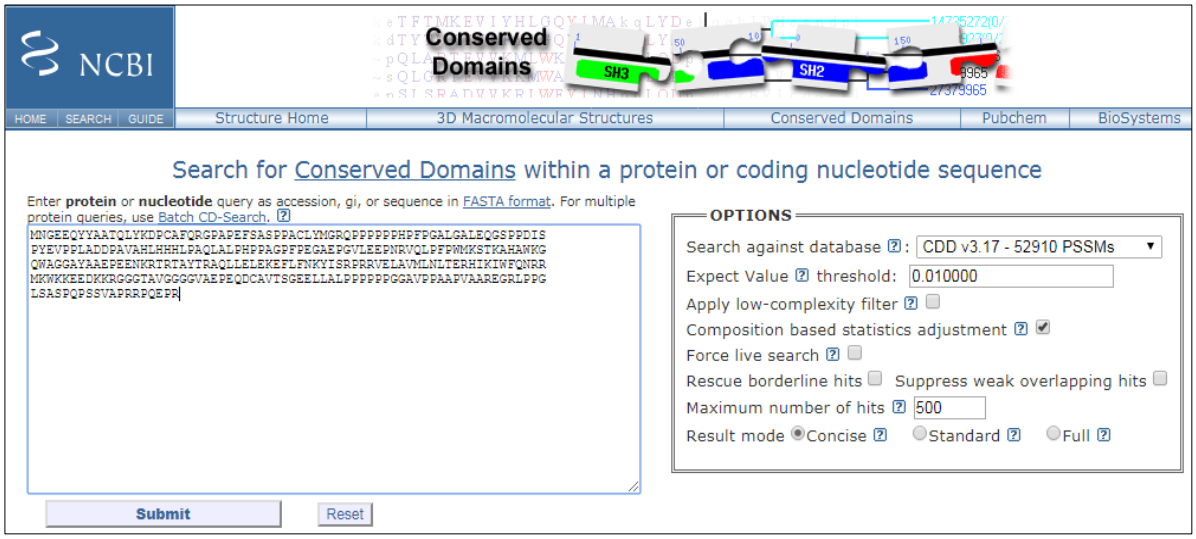


Figure 22: NCBI-Conserved domain search tool.

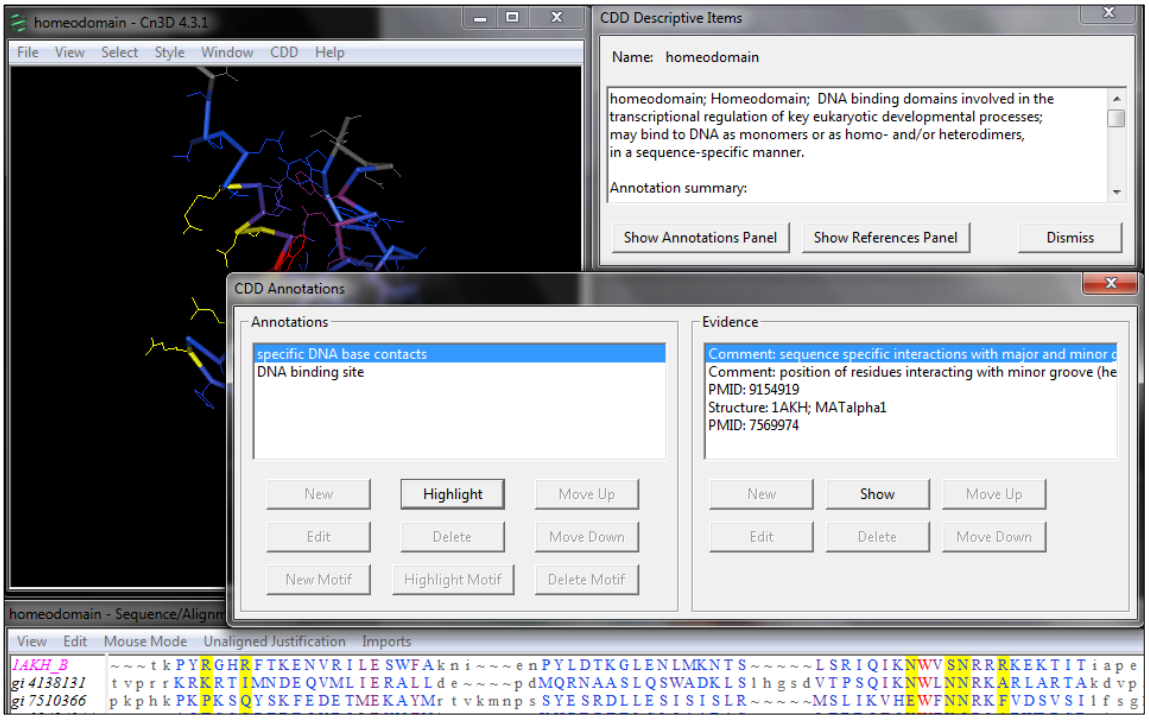


Figure 23: Cn3D 4.3

3.11. Structure prediction

3.11.1. Secondary Structure prediction

The secondary structure of IPF1 protein is predicted by PSIPRED protein structure prediction server (http://bioinf.cs.ucl.ac.uk/psipred), which is based on a neural network algorithm for predicting structural information about a protein from its amino acid sequence alone (McGuffin *et. al*.,2000).

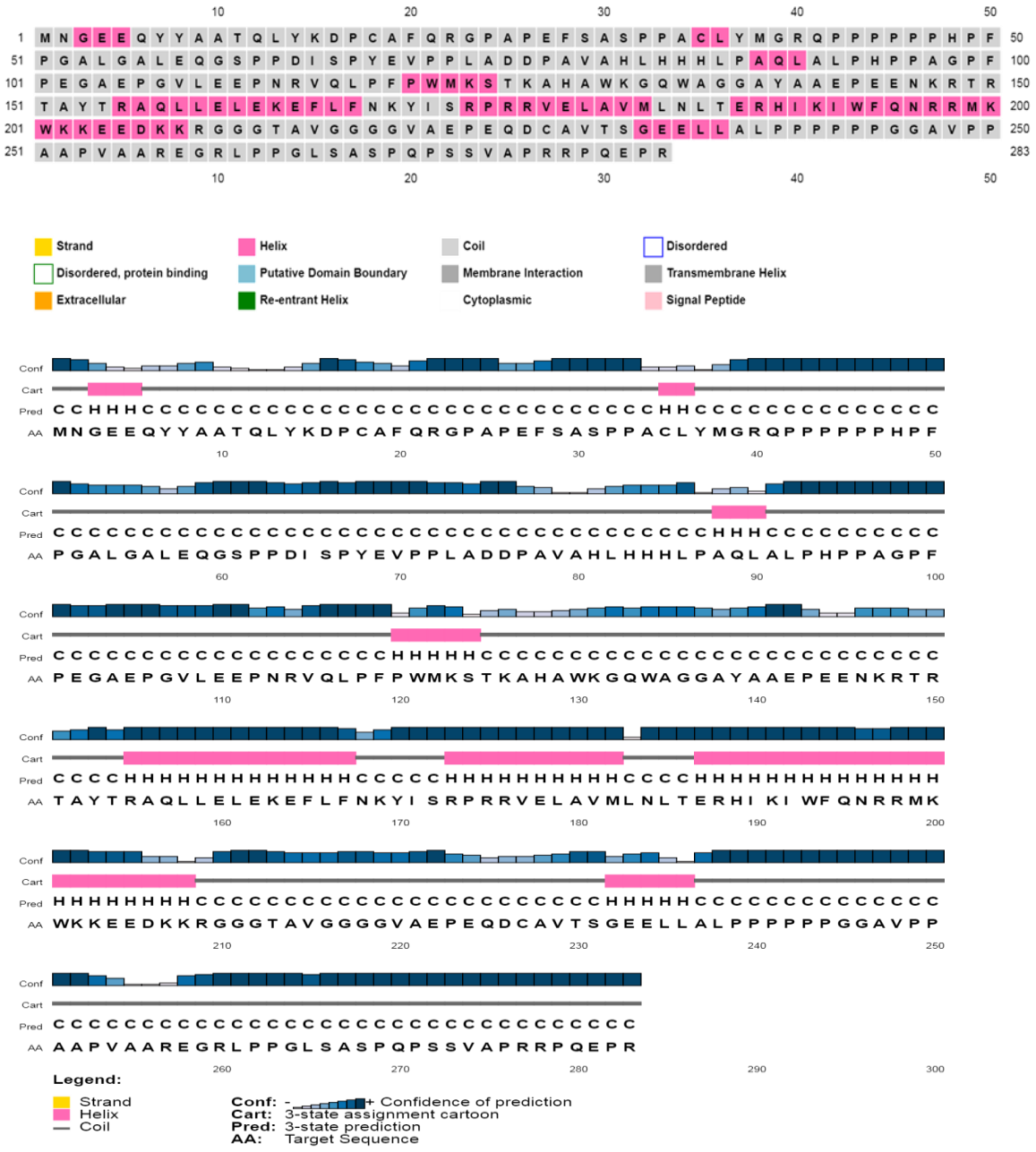


Figure 24: PSIPRED secondary structure prediction of IPF1 protein.

The PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB). It shows the molecule(s) that make up the structure (i.e. protein chains, DNA, ligands and metal ions) and schematic diagrams of their interactions. The sequence search option of the PDBSum provides the secondary structure of the IPF1 DNA binding domain.

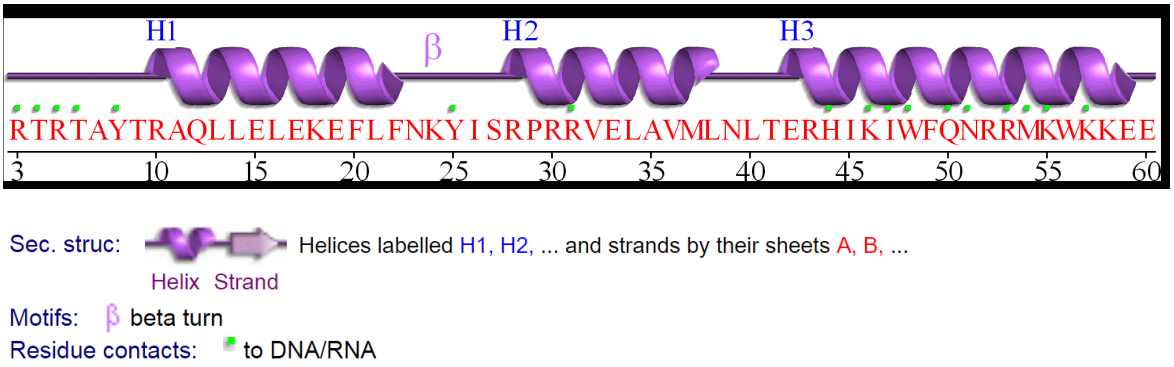


Figure 25: PDB sum schematic secondary structure of IPF1 Homeobox DNA binding domain with amino acid residues, specific base contacts, Helix, Strand and ß –turn location.

3.11.2. 3D modelling of the IPF1 and mutating amino acid residue using UCSF chimera

A tertiary or 3D structure of the IPF1 protein was modeled by comparative protein modeling methods using the SWISS-MODEL server (http://swissmodel.expasy.org). A suitable template for homology modeling was selected using NCBI Position Specific Iterated-Basic Local Alignment Search Tool (PSI-BLAST) against the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) (http://rcsb.org). The ExPASy Swiss-PdbViewer 3.5 (http://spdbv.vital-it.ch) was used to produce a structure-based alignment and SWISS-MODEL was used in the optimized model to minimize energy. 2h1k.pdb file was downloaded as 3D template for chimera (Long *et al.*, 2007). For structural evaluation and stereochemical analyses, the 3D model was submitted to the EBI PDBsum pictorial database. The IPF1 protein model was further evaluated using the University of California, Los Angeles-Department of Energy Institute programs PROCHECK, Verify3D, and ERRAT (http://services.mbi.ucla.edu). Chimera 1.11.2 was used for mutagenesis (Pettersen *et al.*, 2004).

Align.	DB:ID	Source	Length	Score (Bits)	Identities %	Positives %	E()
1 New	SP:PDX1_HUMAN	Pancreas/duodenum homeobox protein 1 OS=Homo sapiens OX=9606 GN=PDX1 PE=1 SV=1	283	602.0	100.0	100.0	1.0E-169
2 New	TR:H2NJH7_PONAB	PDX1 isoform 1 OS=Pongo abelii OX=9601 GN=PDX1 PE=3 SV=1	283	599.0	98.0	100.0	1.0E-168
3 New	TR:G1QHP4_NOMLE	Uncharacterized protein OS=Nomascus leucogenys OX=61853 GN=PDX1 PE=3 SV=1	283	599.0	99.0	99.0	1.0E-168

Figure 26: PSI BLAST Result against the PDX1 amino acid sequence.

Template Results

Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates | More

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
	2h1k.2.C	Pancreatic and duodenal homeobox 1		0.14	-	100.00	X-ray, 2.4Å	monomer ✓	None
	2h1k.1.C	Pancreatic and duodenal homeobox 1		0.14	-	100.00	X-ray, 2.4Å	monomer ✓	None
	2h1k.2.C	Pancreatic and duodenal homeobox 1		0.12	-	100.00	X-ray, 2.4Å	monomer ✓	None
	2h1k.1.C	Pancreatic and duodenal homeobox 1		0.11	-	100.00	X-ray, 2.4Å	monomer ✓	None

Figure 27: Swiss -Moeller template search results for the 3D model of IPF1 protein.

In silico mutation will not drastically change the 3D structure Chimera 1.11.2 was used to introduce mutation in the 3D model. During mutation, a particular rotamer is selected based on its probability (given alongside in UCSF Chimera). High probability means the chance of one residue in that conformation is dominating over others. UCSF Chimera was used to studying any unfavorable contacts/clashes due to mutation.

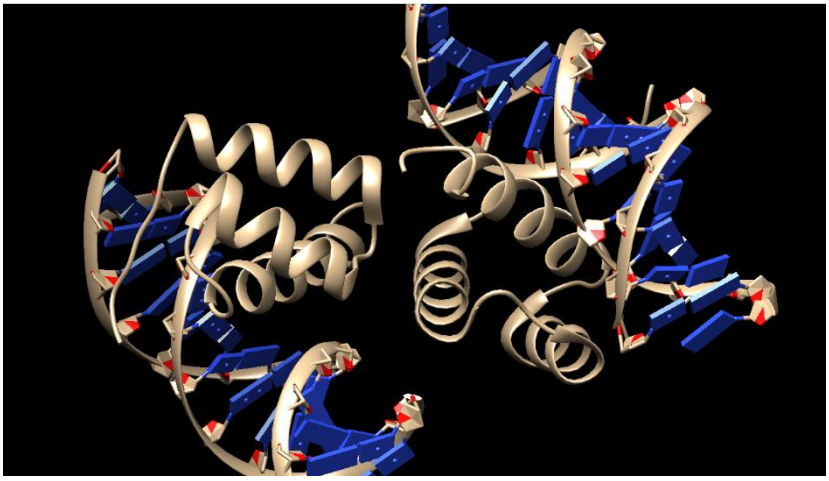


Figure 28: UCSF Chimera 1.11.2 visualizing 2h1k.pdb file.

At first 2h1k PDB file was opened using UCSF chimera. The amino acid sequence was visualized from the chain A. This was done by selecting the sequence option from the favorite menu. The amino acid Arg 52 was selected and shown by clicking the Atom/Bond show option from the Action menu. Residues of around <5A are visualized by clicking the Zone option. From the structure editing sub-menu Rotamer option was selected and amino acid Histidine was selected for introducing mutation and the highest probable rotameric position was applied. Possible clash and contacts were evaluated due to the mutation from the option Surface/binding analysis.

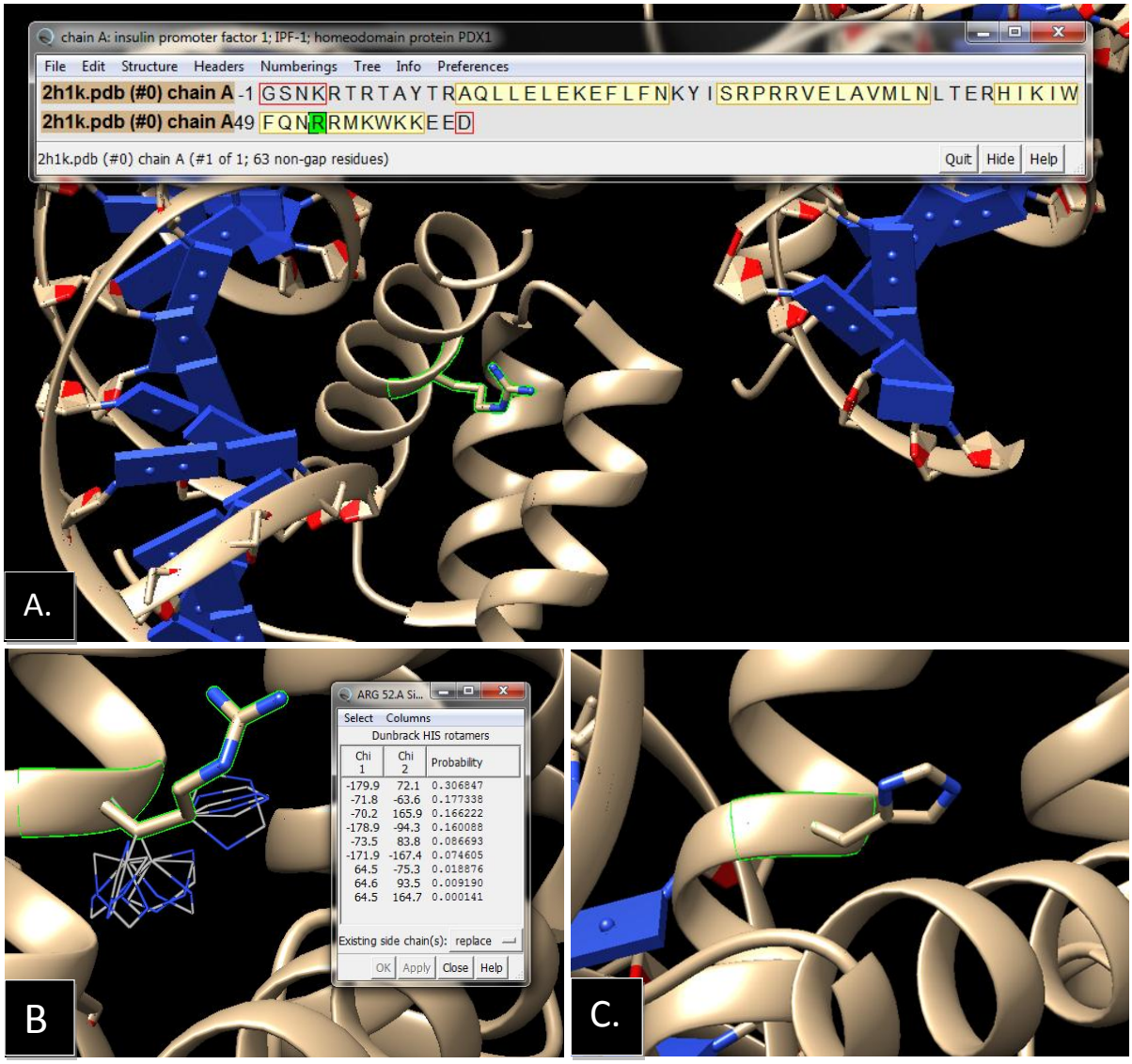


Figure 29: Introducing mutation into 2h1k.pdb file.
(A) Mutating residue Arginen 52 to Histidine 52.
(B) Rotamer possibilty of Histidine 52.
(C) Rotermeric state of Histidine 52.

3.11.3. Predicting protein–DNA binding free energy change upon point mutations

Protein–DNA interactions are essential for regulating many cellular processes, such as transcription, replication, recombination, and translation. Amino acid mutations occurring in DNA-binding proteins have profound effects on protein–DNA binding and are linked with many diseases. SAMPDI web server (http://compbio.clemson.edu/SAMPDI) used to predict the binding free energy change due to mutation. This system uses modified molecular mechanics Poisson-Boltzmann Surface Area (MM/PBSA) method along with an additional set of knowledge-based terms delivered from the investigation of the physicochemical properties of protein-DNA complexes. The important feature is applying DelPhi Gaussian-based smooth dielectric function to calculate the change of solvation energy (Peng *et al.*, 2017).

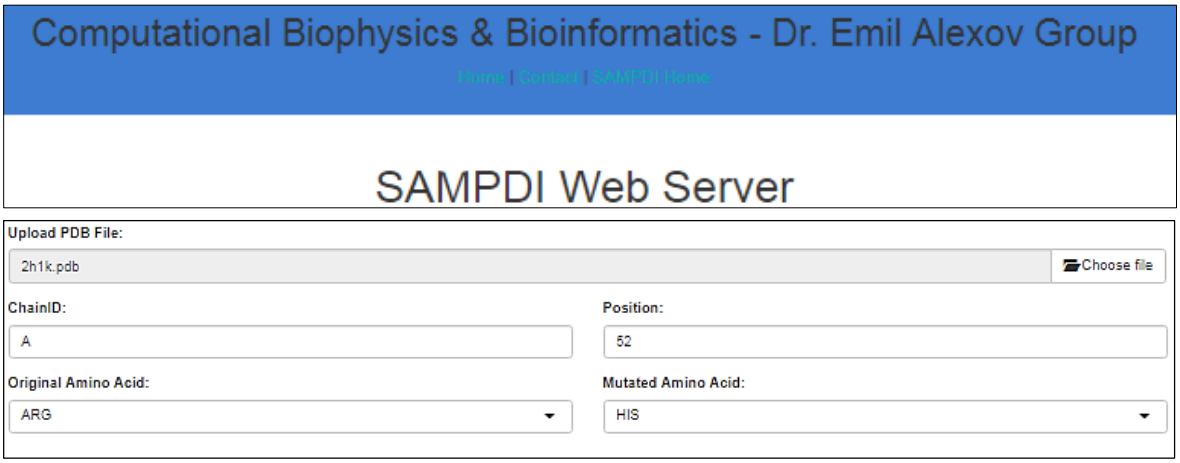


Figure 30: SAMPDI Web server for the prediction of protein DNA binding free energy changes due to missense mutation.

Chapter Four:
Result & Discussion

4.1. RESULT

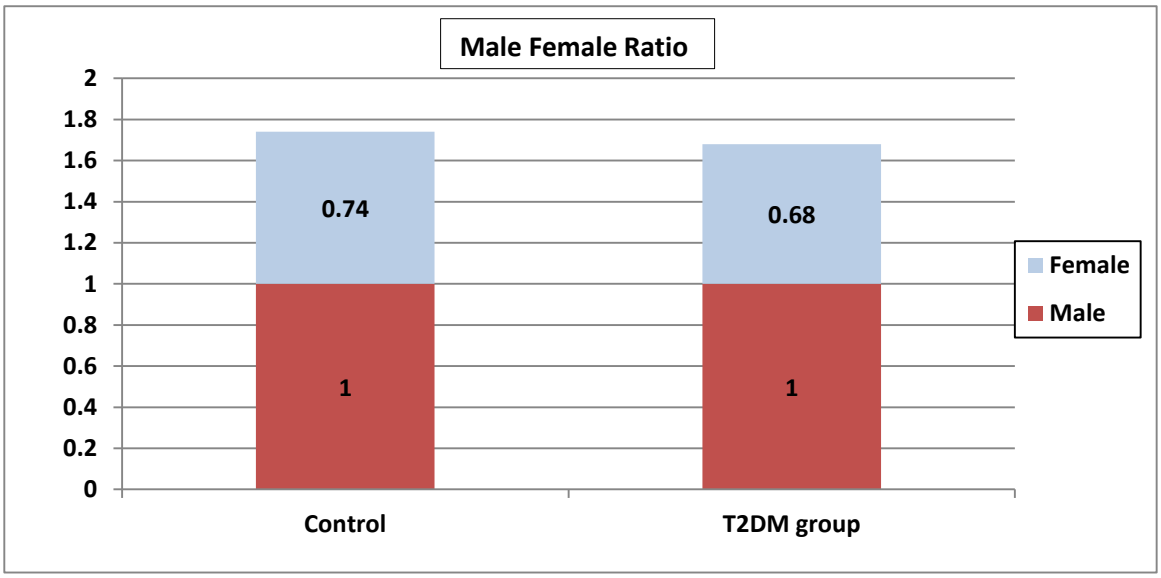
4.1.1. Characteristics of the study subjects

4.1.1.1. Age

The T2DM affected group of subject's average age was 36 to54 which was higher than the average age of the control group 33 to 53 (p=0.016).

4.1.1.2. Gender

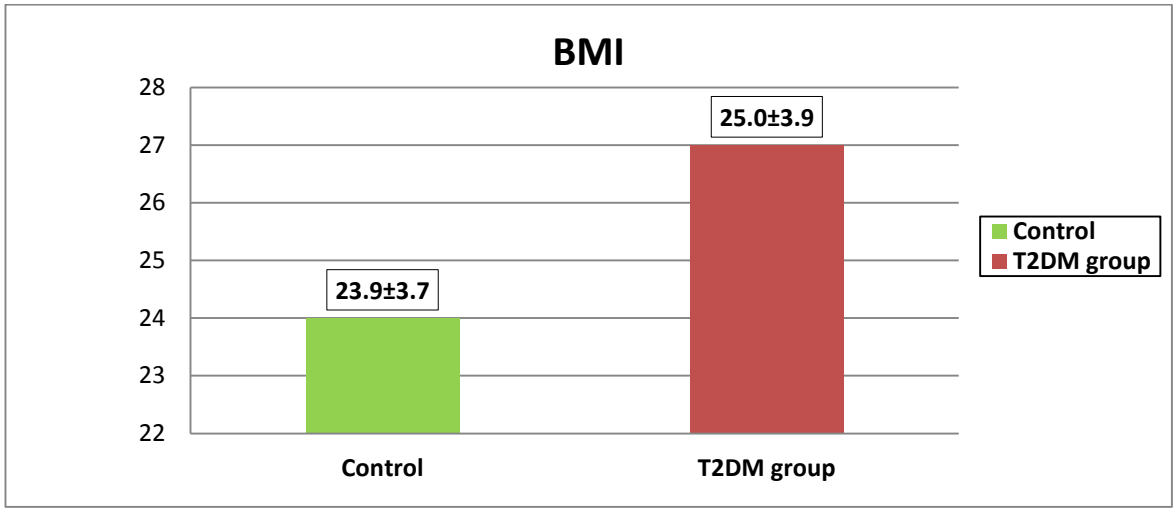
The male-female ratio in the T2DM group was 1: 0.68 and controls 1: 0.74 which did not show any statistical association (p = 0.767).



Bar diagram 1: Male Female ratio among control & T2DM group.

4.1.1.3. Body Mass Index (BMI)

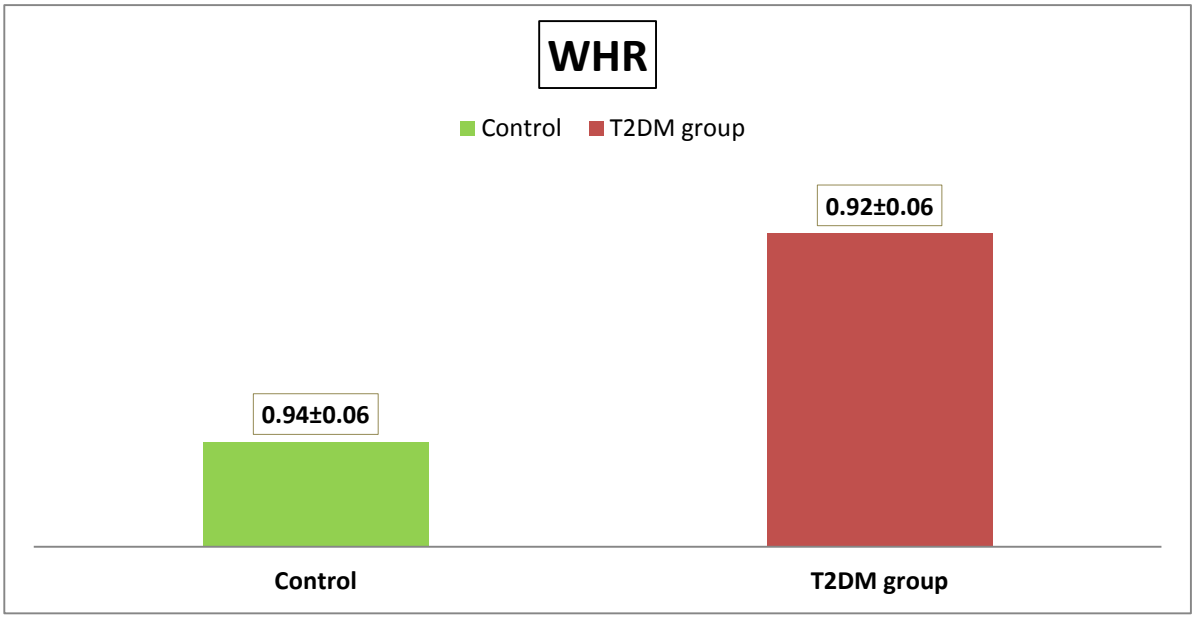
Mean (±SD) BMI (kg/m^2) of the T2DM group was 25.0±3.9 which was significantly higher compared to the control group 23.9±3.7 (p=0.002). Among the total subject 74(17%) had body mass index ≥ 27.5kg/m^2.



Bar diagram 2: Mean (±SD) BMI comparison among control & T2DM group.

4.1.1.4. Waist Hip Ratio (WHR)

Mean (±SD) waist-hip ratio in the T2DM and control group was 0.94±0.06 and 0.92±0.06 respectively. The WHR was significantly (p=0.002) higher in the diabetic group compared to the control counterpart.



Bar diagram 3: Mean (±SD) WHR comparison among control & T2DM group.

Table 16: Clinical characteristics of the study subjects

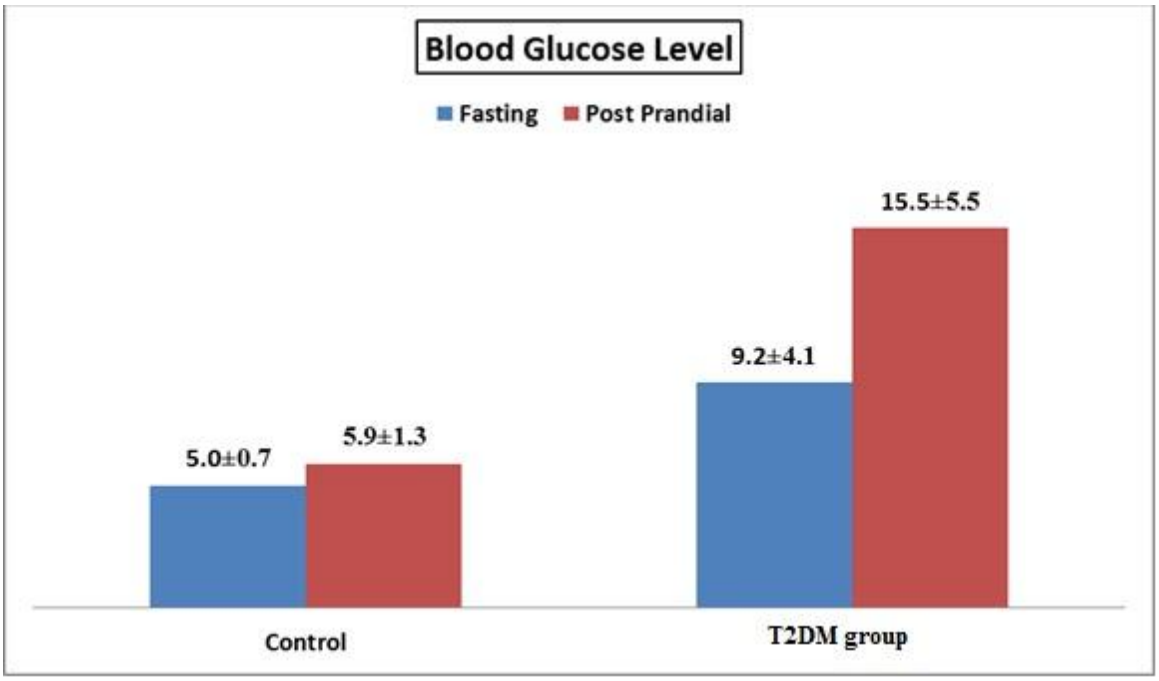
Variables	T2DM (n=200)	Control (n=245)	P-value
Age (Years)	45±9	43±10	0.0156
Sex (Male/Female)	119/81	141/104	0.6997
Body Mass Index (BMI) (kg/m2)	25.0±3.9	23.9±3.7	0.0021
Waist-hip Ratio (WHR)	0.94±0.06	0.92±0.06	0.0015

Results were expressed as mean ±SD. Students unpaired T-test and Fisher's exact test were performed to calculate statistical significance between two groups as appropriate. A P value <0.05 was taken as level of significance.

4.1.2. Biochemical Characteristics of the study subjects

4.1.2.1. Blood glucose

Fasting blood pressure of the T2Dm group was 9.2±4.1 and controls 5.0±0.7 and post prandial (2h glucose tolerance test) glucose in the T2DM and controls was 15.5±5.5 & 5.9±1.3, respectively.



Bar diagram 4: Mean (±SD) blood glucose level comparison among control & T2DM group.

4.1.3. Molecular Characteristics of the study subjects

4.1.3.1. Candidate allele analysis

In all 445 subjects, restriction enzyme digestion produced 117 bp and 287 bp bands were wild type respectively for C18R (T/C) and R197H (G/A) variants. Of the 200 subjects of type 2 diabetes subjects in only one case 111 bp was found, 4 bp and 2 bp bands were too small that those pass through the gel. 50 bp, 61 bp, 4 bp, and 2 bp bands were revealed. The prevalence of IPF-1 C18R (T/C) mutation was 0.5% in the diabetic group. On the other hand, on 200 subjects of type 2 diabetes subjects in only seven cases of 187 bp and 100 bp were found. The prevalence of IPF-1 R197H (G/A) mutation was 3.5% in the diabetic group.

Table 17: Genotype frequency of the candidate IPF-1 gene C18R (T/C) variants in the study subjects

Genotype	Control n (%)	T2DM n (%)
Wild (**T**GC)	244(99.5)	199(99.5)
Polymorphic (**C**GC)	1(0.41)	1(0.51)

*Results expressed as the number (percentage).

Table 18: Genotype frequency of the candidate IPF-1 gene R197H (G/A) variants in the study subjects

Genotype	Control n (%)	T2DM n (%)
Wild (C**G**C)	245(100)	193(96.5)
Polymorphic (C**A**C)	-	7(3.5)

*Results expressed as the number (percentage).

4.1.4. Computational analysis of the IPF1 gene

4.1.4.1. Sequence analysis and alignment

Several bioinformatics tools were used to analyze the IPF1 nucleotide and protein sequence to locate the mutation position of C18R (T/C) and R197H (G/A). The ExPASy ProtParam tool was used to analyze various physiochemical properties from the amino acid sequence of the IPF-1 (or PDX1) protein.

ProtParam is a tool that allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user-entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic

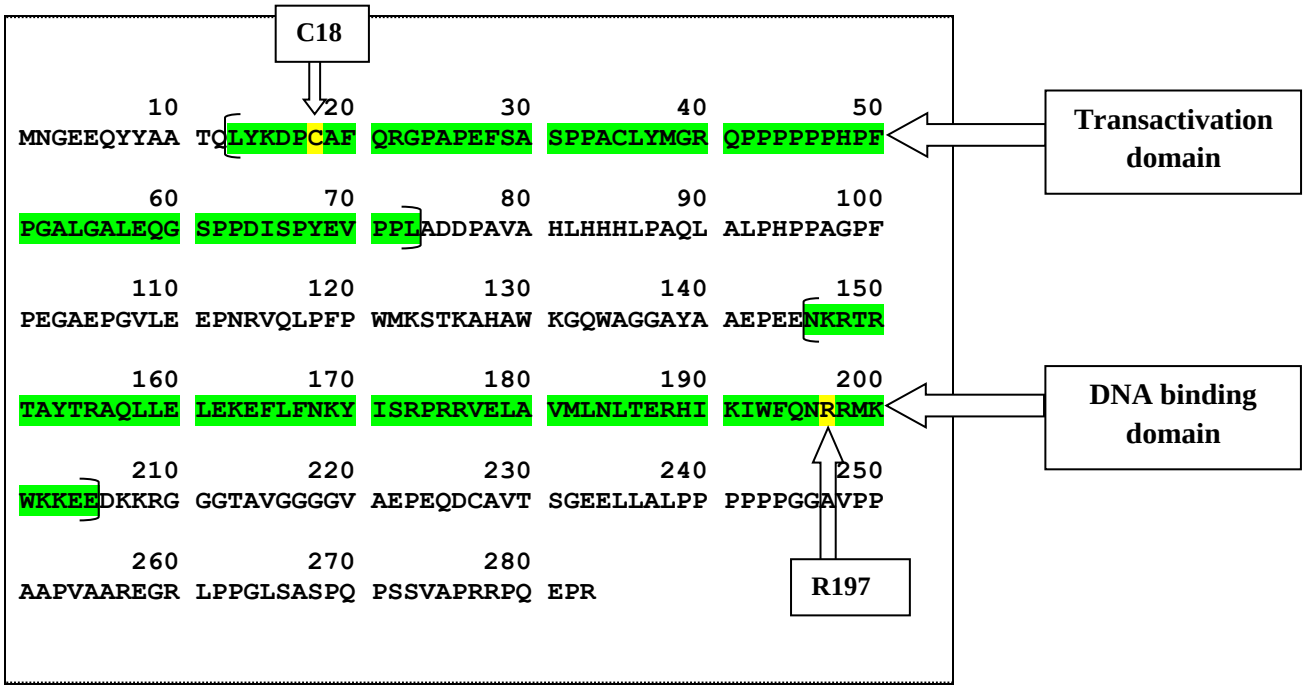
composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

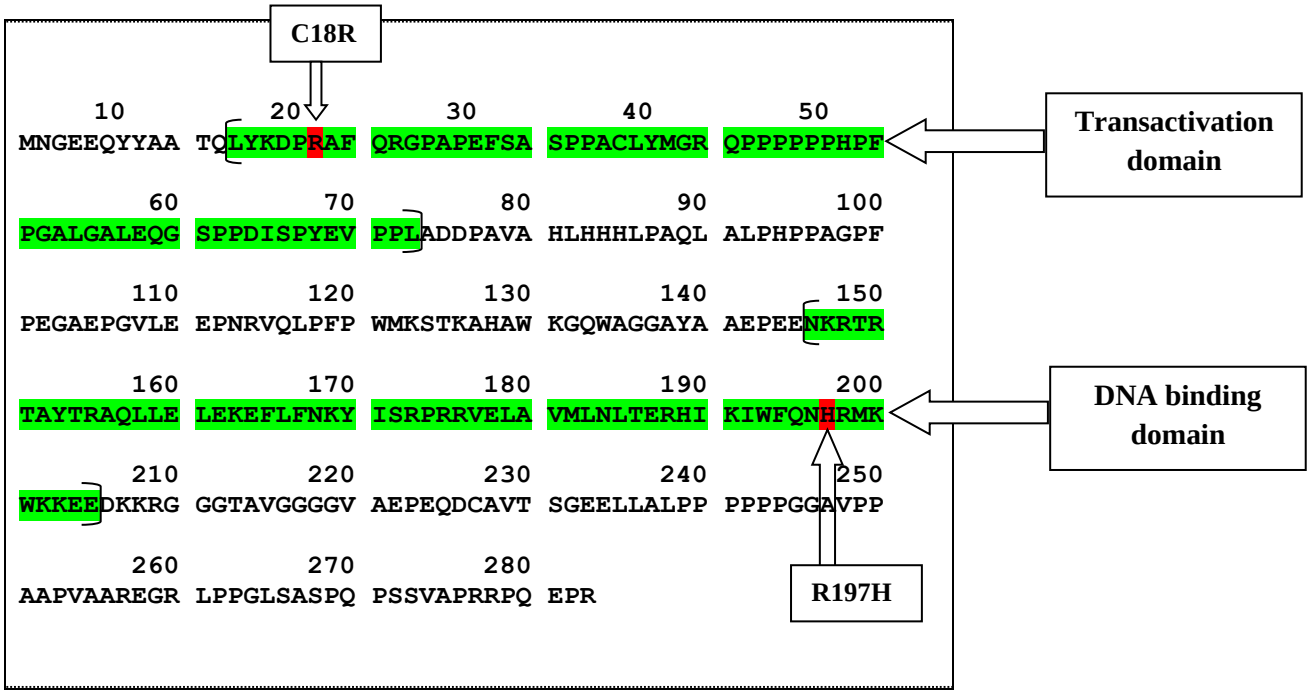
The molecular weight of PDX1 wild type and Mutant protein were predicted 30770.99 Da and 30804.99 Da. and pI of the 283-amino-acid-long PDX1 wild type and mutant sequence were predicted as 7.10 and 7.14 respectively. Instability index for both PDX wild type and mutant protein was predicted 78.54 and 76.52 respectively indicated that both of the protein was stable.

The UniProt Knowledgebase (UniProtKB) is a database of protein sequence and functional information. To find out the functional domain location UniprotKB Peptide search tool (https://www.uniprot.org/peptidesearch/) shows that IPF1 (UniprotKB entry P52945, PDX1_HUMAN) protein has Transactivation domain position from 13-73 and DNA binding Homeobox domain position from 146- 205 in the peptide sequence.

The NCBI Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd/wrpsb.cgi), The NCBI-CDD shows a graphical summary of IPF1 conserved homeobox DNA binding site with specific base contacts.

Table 19: UniProtKB shows domain location of the PDX1 protein (Highlighted green) mutation position C18R and R197H (Highlighted yellow & red).





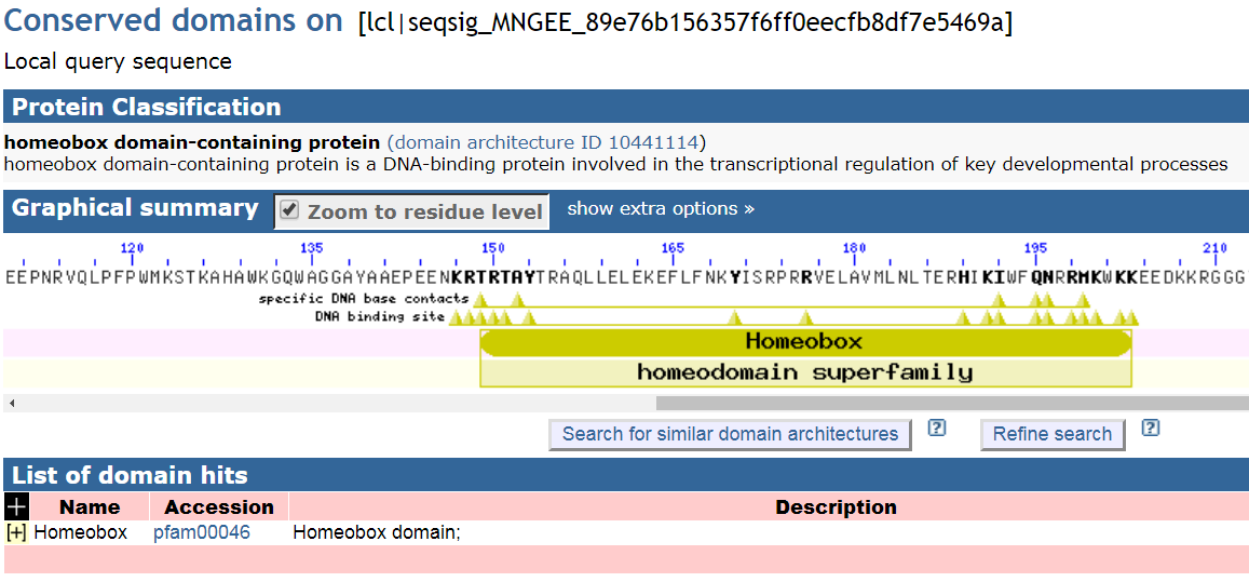


Figure 31: NCBI-CDD graphical summary of IPF1 protein.

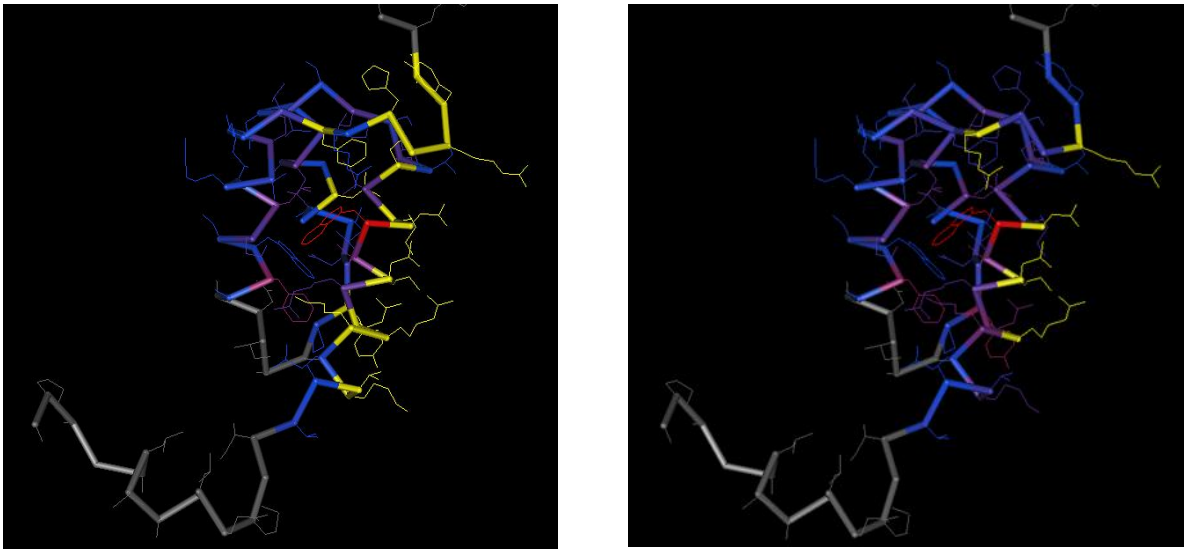


Figure 32: DNA binding site and specific DNA bases contact of PDX1 homeodomain highlighted in yellow viewed by Cn3D.

4.1.4.2. Secondary Structure Prediction

The bioinformatics tools PSIPRED were used to predict the secondary structure of the PDX1 protein. The results revealed that the PDX1 protein possessed a left-handed-β-helix (LβH) domain and belonged to the homeodomain family.

The sequence search option of the PDBSum provides the secondary structure of the IPF1 DNA binding domain.

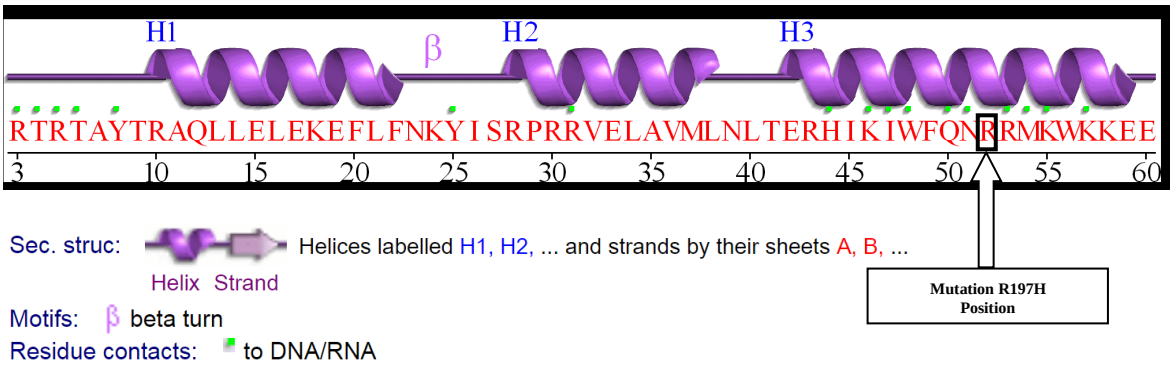


Figure 33: PDB sum schematic secondary structure of IPF1 Homeobox DNA binding domain with Amino acid residues, specific base contacts, Helix, Strand and β –turn location with mutation Position to the corresponding base.

4.1.4.3. 3D modelling of the IPF1 protein

Homology modeling is a standard method for structure prediction and contributes to the understanding of the relationship between protein structure and function. The SWISS-MODEL server was used to model proteins based on protein-structure homology (Arnold *et al.*, 2006). A SWISS-MODEL template search revealed "2h1k" (PDB ID) as an available template that can be used for homology modeling of the IPF1 protein. 2h1k is a refined X‑ray diffraction model of the A‑chain and B- Chain of IPF1 protein. The homology modeling of IPF1 was constructed using 2h1k as a template.

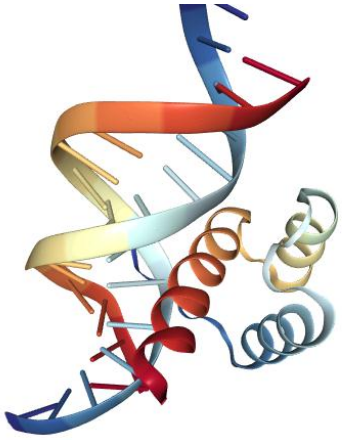


Figure 34: Predicted 3D structure of IPF1 protein.

4.1.4.4. Assessment of 3D structure & PROCHECK summary

The reliability of the predicted 3D model of the "2h1k" was assessed by Ramachandran plot, using PROCHECK, and a quality assessment confirmed using PDBsum. The Ramachandran plot revealed a distribution of 178 amino acid residues of IPF1, with the majority of residues (92.9%) in the favored and allowed regions, 6.2% in the additional allowed region and only one of the residues (0.9%) in the generously allowed region.

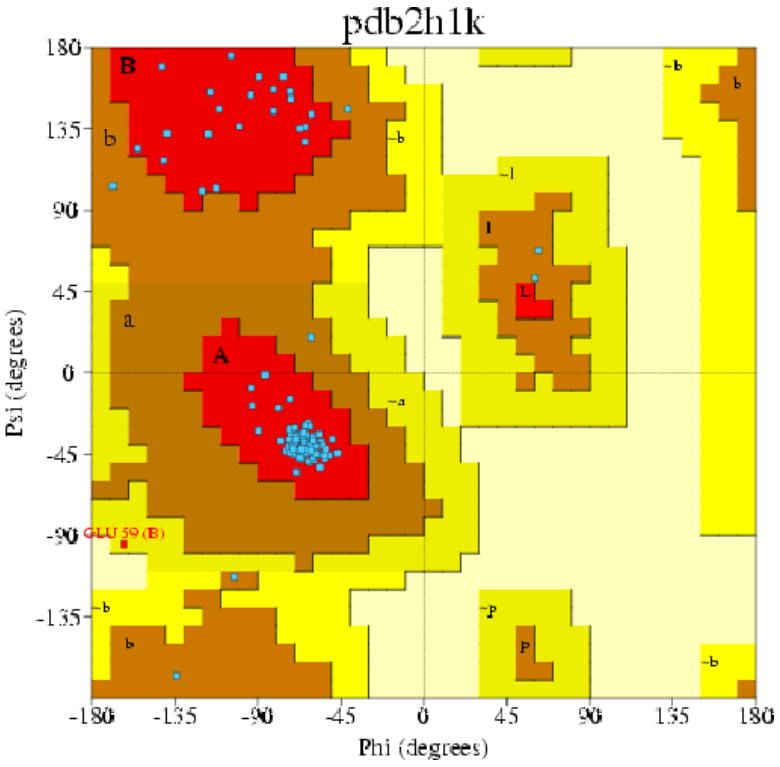


Figure 35: Ramachandran Plot for 2h1k

Table 20: Ramachandran Plot statistics for predicted 3D model.

Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	104	92.9%
Additional allowed regions	[a,b,l,p]	7	6.2%
Generously allowed regions	[~a,~b,~l,~p]	1	0.9%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		112	100.0%
End-residues (excl. Gly and Pro)		64	
Glycine residues		0	
Proline residues		2	
Total number of residues		178	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A, B, L].

4.1.4.5. Mutating Residue in UCSF Chimera 1.11.2

UCSF Chimera (or simply Chimera) is an extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles (Pettersen *et. al.*, 2004).

In silico mutation will not drastically change the 3D structure Chimera 1.11.2 was used to introduce mutation in the 3D model. During mutation, a particular rotamer is selected based on its probability (given alongside in Chimera). High probability means the chance of one residue in that conformation is dominating over others. Chimera was used to study any unfavorable contacts/clashes due to mutation.

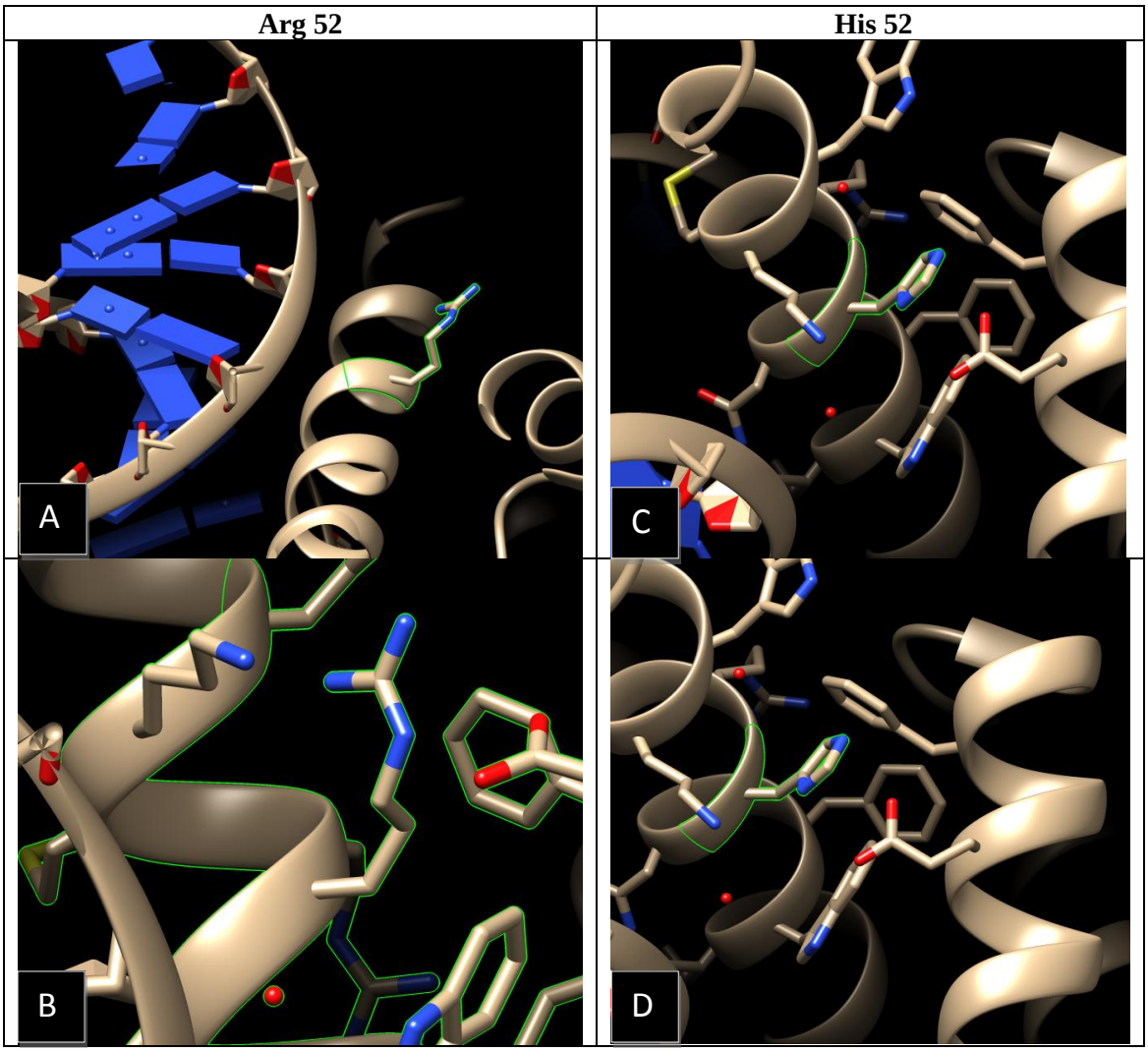


Figure 36: Introduce mutation to the position Arg 52 to His 52 and finding any possible Clash and contacts due to mutation. (A) & (B) shows the clash/contacts with neighbouring residues. (C) & (D) shows the clash/contacts with neibouring residues.

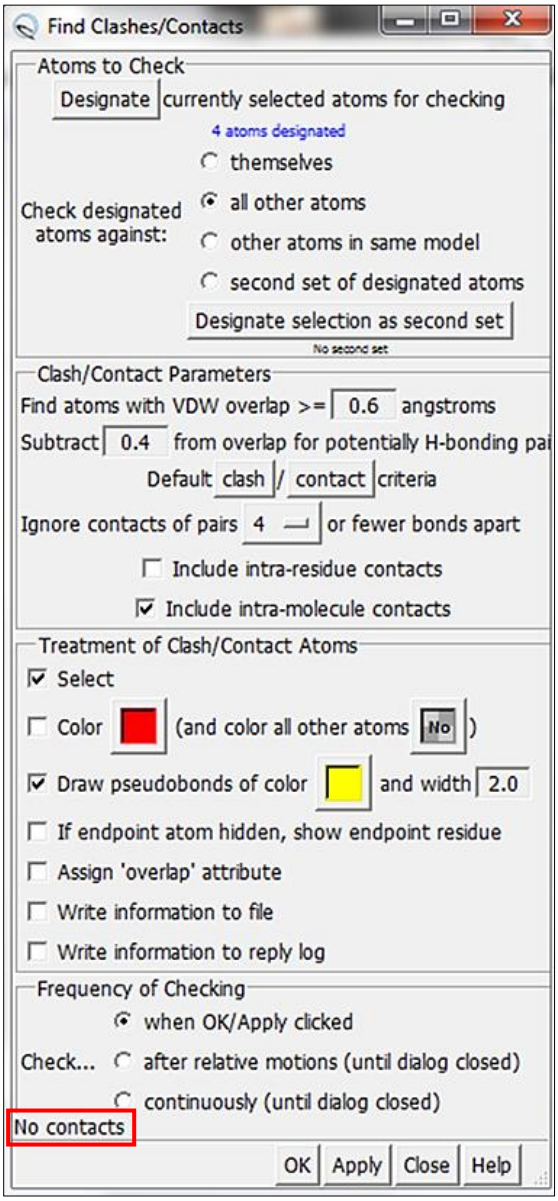


Figure 37: UCSF Chimera result box for possible clash/contacts due to introducing mutation.

4.1.4.6. Predicting protein–DNA binding free energy change

SAMPDI Web Server gives quick and correct predictions of the protein-DNA complex for the effects of single amino acid substitution on the binding free energy (Peng *et al.*, 2017).

Prediction of protein DNA binding free energy change shows the total change in binding free energy is 0.887977 kcal/mol.

Result:

Locations	52
Chain	A
WT Residue	ARG
MT Residue	HIS
Change of Coulombic energy	58.4304
Change of vdW energy	0.5393
Change of Polar component solvation energy	-32.2226
Change of Solvent Accessiable Surface Area	-34
Change of Entropy	0.234994
Change of Hydrogen bonds within 6 angstrom of mutations occuring position	-1
The total change in binding free energy due to mutation is	0.887977

##Positive change in binding free energy means that the mutation makes binding weaker: ddG=dG(MT)-dG(WT), unit in kcal/mol

Figure 38: SAMPDI Web Server results for the PDX1 protein due to R197H mutation.

4.2. DISCUSSION

As a key transcription factor in the pathways that control both β-cell mass and essential insulin biosynthesis and secretion genes, IPF1 is a strong candidate for the inherited insulin secretion defect that characterizes T2DM and the pre-diabetic state. It has been estimated that IPF1 may represent a diabetes-predisposing gene in a non-negligible proportion (~6%) of the common forms of T2DM (Hani *et al.*, 1999).

Cell line studies indicated that the three novel IPF-1 mutations are reported to be associated with T2DM in some population (Hansen *et al.*, 2000; Macfarlane *et al.*, 1999; Hani *et al.*, 1999; Reis *et al.*, 2000). In this study, 445 samples from Bangladesh subjects have been studied for the IPF-1 gene's C18R (T/C) and R197H (G/A) variants mutation by PCR-RFLP, among the subjects 200 samples were obtained from type-2 diabetes mellitus subjects.

Macfarlane *et al.* (1999) had investigated that mutations on C18R, D76N, and R197H in the IPF-1 gene in diabetic and non-diabetic in UK subjects are rare but predispose the subjects to T2DM (who were having the mutation) with a relative risk of 3.0. But these mutations are a rare cause of MODY and pancreatic agenesis (Hansen *et al.*, 2000; Mcfarlane *et al.*, 1999; Hani *et al.*, 1999). Furthermore, these variants can be as high as 6% in subjects with a strong family of T2DM (Reis *et al.*, 2000). Other reports from Taiwan suggested that IPF- 1 mutations are not commonly associated with T2DM (Shiau *et al.*, 2010). The mutations, however, do not highly penetrate, as the carrier has the mutation for 25-53 years of age and remained non-diabetic. Hani *et al.* (1999) had investigated 192 French families with T2DM and identified 3 novel IPF1 mutations, including 2 substitutions (Q59L and D76N) and in-frame proline insertion (insCCG243). The CCG243 mutation linked in two families in autosomal dominant-like late-onset form of type 2 DM was revealed by functional analysis. The most prevalent D76N (G/A) and Q59L mutations were associated with an increased relative risk for diabetes and decreased glucose-stimulated insulin secretion in non-diabetic subjects. In another study, 200 Danish late-onset T2DM and, 44 Danish and Italian MODY patients are investigated for IPPI mutations and observed noncoding-108G Indel, silent G45C substitution as well as rare variant D75N in the Danish T2DM patients. Moreover, one Danish MODY patient has carried an A140T variant. D75N and A140T variants are not segregated with diabetes (Fiansen *et al.*, 2000). In another one, 26 variant of IPF1 gene are analyzed through genetics sequencing on 1788 UK individuals (where 910 individuals were patients and 878 in control group) and found ~5% of the individual has this variant. No evidence of association of T2DM was observed (Edghill *et al.*, 2011). A cohort study on Sudanese T2DM patients presents evidence that C18R & D76N variants of the IPF1 gene's mutation have <1% frequency for T2DM development but R197H was not. Only one C18R SNP was found, out of 31 patients and 3 out of 61 non-diabetic individuals were carrying this SNP. Karim *et al.* 2005 had

examined 341 African American individual cases and identified 8 variant in which InsCCG243 variant contribute to diabetes susceptibility in this population in low penetrance about 3.7%. In many countries lack association for the IPF1 gene C18R (T/C) and R197H (G/A) mutations have also reported by the number of researchers. In the British population, the C18R variant 0.51% frequency observed in T2DM patients (Macfarlane *et al.*, 1999). In another study among Caucasian, it was found 0.37% frequency of C18R mutation in diabetic patients (Reis *et al.*, 2000), it also plays a small part in T2DM development in Sudanese individuals.

Mutation in the β-Cell transcription factor IPF-1 can result in reduced binding of the protein to the insulin gene promoter and decreased insulin gene transcription in response to hyperglycemia. But the reduction in the function of the missense mutations identified in type 2 diabetic patients is not as severe as that seen with the P63fsdelC frameshift mutation, which resulted in MODY (Staffers *et al.*, 1997). But the prevalence of IPF-1 mutations in diabetic and control populations from the United Kingdom among 206 diabetic study subjects only 1% C18R and 0.5% R197H mutation was found (Macfarlane *et al.*, 1999). Evidence showed that the R197H mutations contribution is very rare.

In the present study, multiple factors related to T2DM were taken into consideration. Obesity was one of them. BMI determines as a predictor of obesity and in this study, BMI of T2DM group has shown significantly higher ($p=0.0021$) compared to the control group. Also, WHR was higher ($p=0.0015$) in T2DM patients. The serum glucose level was high in T2DM subjects ($p\leq0.001$) compared to the healthy control group. These are common phenomena among diabetic patients.

As the evidence showed that the R197H mutations contribution is very rare. In our molecular study, we found a 3.5% mutation of R197H among T2DM patients, which is below 5%. That indicates this variant contribution is negligible to develop T2DM in the Bangladeshi population. On the other hand, we found 1 mutation in healthy individuals among 245 as well as 1 in 200 diabetic subjects for C18R. Which shows, C18R mutation is also uninfluential.

Also, we have found a similarity with the observable evidence in our computational study. ExPASy ProtParam tool showed, IPF1 having C18R and R197H mutations have no or very little impact on protein stability. From UniProtKB we have found, C18R SNP occurs within the transactivation domain which primarily plays a role in transcriptional activation. Several studies have indicated, a mutation within arginine to cysteine (or vice versa) residue located near the NH_2-terminal and first proline-rich domains of IPF-1 has less severe or less impact on protein functionality (Chakkalakal *et al.*, 2018; Macfarlane *et al.*, 1999). From the NCBI Conserved Domains Database; NCBI Conserved domain (CD) search tool recognize and illustrate the DNA binding Homeobox domain along with DNA binding site with specific base contacts. PSIPRED

predicted and PDBsum (Structural summaries of PDB entries) secondary structure of the IPF1 protein shown that the mutation R197H location belongs to the DNA binding homeodomain and also belongs to the Helix H3, but not belong to the residues those have specific base contacts with DNA. Three dimensional (3D) modeling and mutating residue with UCSF Chimera have shown introducing mutation R197H on the 3D structure of the IPF1 binding domain with the highest possible rotameric state for placing Histidine in place of Arginine to study possible clashes/contacts and showed no clashes with neighboring residues. SAMPDI (Single Amino acid Mutation binding free energy change of Protein–DNA Interaction) performed predictions of binding free energy changes of protein–DNA complexes caused by single mutations on the proteins. SAMPDI Web Server (Peng *et al.*, 2017) result showed binding free energy change due to mutation R197H was negligible.

T2DM is multi-factorial disorders that are often strongly associated with environmental factors at one end of the spectrum and genetic factors at the other end. In the present study, the prevalence of dyslipidemia was very common among T2DM subjects compared to non-diabetic subjects. Furthermore, obesity, as assessed by BMI or WHR, was higher in type 2 diabetes subjects. Thus at the spectrum of genetic factors, obesity-related genes or other promising genetic variations that can affect carbohydrate metabolism may predispose among T2DM subjects of the Bangladeshi origin.

4.3. CONCLUSION

Molecular analysis of the present study represents a small number of C18R and R197H mutations in T2DM patients and the computational study also showed, both of these variants are not much influencing. It might be due to the C18R and R197H mutations do not make significant changes in IPF1 protein structure or functionality. So, it can be suggested that C18R, as well as, R197H mutations of the IPF1 gene, do not contribute to develop T2DM in Bangladeshi population.

4.4. LIMITATION

- Small Number of Subjects was recruited.
- Insulin sensitivity and secretory capacity of these diabetic and non-diabetic subjects have not been carried out.

4.5. RECOMMENDATION

- Study of the other variants of IPF1 gene possibly give better explanation of the role of these gene variants in the pathogenesis of the type 2 diabetes mellitus in the population of Bangladeshi origin.
- Recruitment of the additional diabetic cases from population might be studied to validate the findings.

REFERENCE:

Abdul-Ghani, M.A., Williams, K., DeFronzo, R. and Stern, M., 2006. Risk of progression to type 2 diabetes based on relationship between postload plasma glucose and fasting plasma glucose. Diabetes Care, 29(7), pp.1613-1618.

Acharjee, S., Ghosh, B., Al-Dhubiab, B.E. and Nair, A.B., 2013. Understanding type 1 diabetes: etiology and models. Canadian journal of diabetes, 37(4), pp.269-276.

Ahlgren, U., Jonsson, J. and Edlund, H., 1996. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. Development, 122(5), pp.1409-1416.

Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K. and Edlund, H., 1998. β-Cell specific inactivation of the mouseIpf1/Pdx1 gene results in loss of the β-cell phenotype and maturity onset diabetes. Genes & development, 12(12), pp.1763-1768.

Aitman, T.J. and Todd, J.A., 1995. Molecular genetics of diabetes mellitus. Bailliere's clinical endocrinology and metabolism, 9(3), pp.631-656.

Allende-Vigo, M., 2010. Pathophysiologic mechanisms linking adipose tissue and cardiometabolic risk. Endocrine Practice, 16(4), pp.692-698.

Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M.R., Lyssenko, V., Tuomi, T., Groop, L. and Botnia Study Group, 2011. Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia, 54(11), p.2811.

Al-Quobaili, F. and Montenarh, M., 2008. Pancreatic duodenal homeobox factor-1and diabetes mellitus type 2. International journal of molecular medicine, 21(4), pp.399-404.

American Diabetes Association (ADA), 1997. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes care, 20, pp.1183-1197.

American Diabetes Association, 2003. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2003 (suppl 1), p.S2-5.

American Diabetes Association, 2014. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2014, 37 (Suppl 1): p.S81-90.

Andrali, S.S., Sampley, M.L., Vanderford, N.L. and Özcan, S., 2008. Glucose regulation of insulin gene expression in pancreatic β-cells. Biochemical Journal, 415(1), pp.1-10.

Aramata, S., Han, S.I. and Kataoka, K., 2007. Roles and regulation of transcription factor MafA in islet β-cells. Endocrine journal, 54(5), pp.659-666.

Attiya, K. and Sahar, F., 2012. 'Maturity-onset Diabetes of the Young (MODY) Genes□: Literature Review', 1(1), pp. 4–11.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22(2), pp.195-201.

Barroso, I., 2005. Genetics of type 2 diabetes. Diabetic Medicine, 22(5), pp.517-535.

Barrow, J., Hay, C.W., Ferguson, L.A., Docherty, H.M. and Docherty, K., 2006. Transcription factor cycling on the insulin promoter. FEBS letters, 580(2), pp.711-715.

Belkina, A.C., Nikolajczyk, B.S. and Denis, G.V., 2013. BET protein function is required for inflammation: Brd2 genetic disruption and BET inhibitor JQ1 impair mouse macrophage inflammatory responses. The Journal of Immunology, 190(7), pp.3670-3678.

Bell, G.I. and Polonsky, K.S., 2001. Diabetes mellitus and genetically programmed defects in β-cell function. Nature, 414(6865), p.788.

Bisschop, P.H., de Metz, J., Ackermans, M.T., Endert, E., Pijl, H., Kuipers, F., Meijer, A.J., Sauerwein, H.P. and Romijn, J.A., 2001. Dietary fat content alters insulin-mediated glucose metabolism in healthy men. The American journal of clinical nutrition, 73(3), pp.554-559.

Bogardus, C., Lillioja, S., Mott, D., Zawadzki, J., Young, A. and Abbott, W., 1985. Evidence for reduced thermic effect of insulin and glucose infusions in Pima Indians. The Journal of clinical investigation, 75(4), pp.1264-1269.

Bonnefond, A. and Froguel, P., 2015. Rare and common genetic events in type 2 diabetes: what should biologists know?. Cell metabolism, 21(3), pp.357-368.

Bouatia-Naji, N., Rocheleau, G., Van Lommel, L., Lemaire, K., Schuit, F., Cavalcanti-Proença, C., Marchand, M., Hartikainen, A.L., Sovio, U., De Graeve, F. and Rung, J., 2008. A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. Science, 320(5879), pp.1085-1088.

Boucher, M.J., Selander, L., Carlsson, L. and Edlund, H., 2006. Phosphorylation marks IPF1/PDX1 protein for degradation by glycogen synthase kinase 3-dependent mechanisms. Journal of Biological Chemistry, 281(10), pp.6395-6403.

Butler, A.E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R.A. and Butler, P.C., 2003. β-cell deficit and increased β-cell apoptosis in humans with type 2 diabetes. Diabetes, 52(1), pp.102-110.

Campbell, A.J., Busby, W.J., Horwath, C.C. and Robertson, M.C., 1993. Relation of age, exercise, anthropometric measurements, and diet with glucose and insulin levels in a population aged 70 years and over. American journal of epidemiology, 138(9), pp.688-696.

Chan, J.M., Rimm, E.B., Colditz, G.A., Stampfer, M.J. and Willett, W.C., 1994. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes care, 17(9), pp.961-969.

Chen, W.M., Erdos, M.R., Jackson, A.U., Saxena, R., Sanna, S., Silver, K.D., Timpson, N.J., Hansen, T., Orrù, M., Piras, M.G. and Bonnycastle, L.L., 2008. Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. The Journal of clinical investigation, 118(7), pp.2620-2628.

Cheng, K., Andrikopoulos, S. and E Gunton, J., 2013. First phase insulin secretion and type 2 diabetes. Current molecular medicine, 13(1), pp.126-139.

Chakkalakal, S., Heilig, J., Baumann, U., Paulsson, M. and Zaucke, F., 2018. Impact of arginine to cysteine mutations in collagen II on protein secretion and cell survival. International journal of molecular sciences, 19(2), p.541.

Collins, F.S., Morgan, M. and Patrinos, A., 2003. The Human Genome Project: lessons from large-scale biology. Science, 300(5617), pp.286-290.

Dahlquist, G., 1998. The aetiology of type 1 diabetes: an epidemiological perspective. Actapaediatrica, 87(s425), pp.5-10.

DeFronzo, R.A., 2004. Pathogenesis of type 2 diabetes mellitus. Medical Clinics, 88(4), pp.787-835.

DeFronzo, R.A., 2009. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. Diabetes, 58(4), pp.773-795.

De Franco, E., Shaw-Smith, C., Flanagan, S.E., Edghill, E.L., Wolf, J., Otte, V., Ebinger, F., Varthakavi, P., Vasanthi, T., Edvardsson, S. and Hattersley, A.T., 2013. Biallelic PDX1 (insulin promoter factor 1) mutations causing neonatal diabetes without exocrine pancreatic insufficiency. Diabetic Medicine, 30(5), pp.e197-e200.

Docherty, K., Macfarlane, W.M., Read, M.L., Smith, S.B., Wilson, M.E., Bujalska, I. and Gilligan, M., 1996. Regulation of insulin gene transcription by nutrients.

Dorajoo, R., Liu, J. and Boehm, B.O., 2015. Genetics of type 2 diabetes and clinical utility. Genes, 6(2), pp.372-384.

Doria, A., 2010. Genetics of diabetes complications. Current diabetes reports, 10(6), pp.467-475.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L. and Lindgren, C.M., 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics, 42(2), p.105.

Edghill, E.L., Khamis, A., Weedon, M.N., Walker, M., Hitman, G.A., McCarthy, M.I., Owen, K.R., Ellard, S., T Hattersley, A. and Frayling, T.M., 2011. Sequencing PDX1 (insulin promoter factor 1) in 1788 UK individuals found 5% had a low frequency coding variant, but these variants are not associated with Type 2 diabetes. Diabetic Medicine, 28(6), pp.681-684.

Ellard, S., Bellanné-Chantelot, C., Hattersley, A.T. and European Molecular Genetics Quality Network (EMQN) MODY group, 2008. Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia, 51(4), pp.546-553.

Fagot-Campagna, A., Pettitt, D.J., Engelgau, M.M., Burrows, N.R., Geiss, L.S., Valdez, R., Beckles, G.L., Saaddine, J., Gregg, E.W., Williamson, D.F. and Narayan, K.V., 2000. Type 2 diabetes among North adolescents: An epidemiologic health perspective. The Journal of pediatrics, 136(5), pp.664-672.

Fajans, S.S. and Vinik, A.I., 1989. Insulin-producing islet cell tumors. Endocrinology and metabolism clinics of North America, 18(1), pp.45-74.

Florez, J.C., Hirschhorn, J. and Altshuler, D., 2003. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annual review of genomics and human genetics, 4(1), pp.257-291.

Francis, J., Chakrabarti, S.K., Garmey, J.C. and Mirmira, R.G., 2005. Pdx-1 links histone H3-Lys-4 methylation to RNA polymerase II elongation during activation of insulin transcription. Journal of Biological Chemistry, 280(43), pp.36244-36253.

Franks, P.W., Pearson, E. and Florez, J.C., 2013. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. Diabetes care, 36(5), pp.1413-1421.

Frayling, T.M., Evans, J.C., Bulman, M.P., Pearson, E., Allen, L., Owen, K., Bingham, C., Hannemann, M., Shepherd, M., Ellard, S. and Hattersley, A.T., 2001. beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. Diabetes, 50(suppl 1), p.S94.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. and Shields, B., 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science, 316(5826), pp.889-894.

Gambineri, A., Patton, L., Altieri, P., Pagotto, U., Pizzi, C., Manzoli, L., & Pasquali, R., 2012. Polycystic ovary syndrome is a risk factor for type 2 diabetes: results from a long-term prospective study. Diabetes, 61(9), pp. 2369–2374.

Gao, Y., Miyazaki, J.I. and Hart, G.W., 2003. The transcription factor PDX-1 is post-translationally modified by O-linked N-acetylglucosamine and this modification is correlated with its DNA binding activity and insulin secretion in min6 β-cells. Archives of biochemistry and biophysics, 415(2), pp.155-163.

Gloyn, A. L. 2003. Glucokinase (GCK) mutations in hyper- and hypoglycemia: maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemia of infancy. Hum Mutat22:353-362.

Goodison, S., Kenna, S. and Ashcroft, S. J., 1992. Control of insulin gene expression byglucose. Biochem. J. 285, 563–568.

Guo, S., 2014. Insulin signaling, resistance, and the metabolic syndrome: insights from mouse models to disease mechanisms. The Journal of endocrinology, 220(2), p.T1.

Hagman, D.K., Hays, L.B., Parazzoli, S.D. and Poitout, V., 2005. Palmitate inhibits insulin gene expression by altering PDX-1 nuclear localization and reducing MafA expression in isolated rat islets of Langerhans. Journal of Biological Chemistry, 280(37), pp.32413-32418.

Hagman, D.K., Latour, M.G., Chakrabarti, S.K., Fontes, G., Amyot, J., Tremblay, C., Semache, M., Lausier, J.A., Roskens, V., Mirmira, R.G. and Jetton, T.L., 2008. Cyclical and alternating infusions of glucose and intralipid in rats inhibit insulin gene expression and Pdx-1 binding in islets. Diabetes, 57(2), pp.424-431.

Haines, L., Wan, K.C., Lynn, R., Barrett, T.G. and Shield, J.P., 2007. Rising incidence of type 2 diabetes in children in the UK. Diabetes care, 30(5), pp.1097-1101.

Hales, C.N., Barker, D.J., Clark, P.M., Cox, L.J., Fall, C., Osmond, C. and Winter, P.D., 1991. Fetal and infant growth and impaired glucose tolerance at age 64. Bmj, 303(6809), pp.1019-1022.

Hansen, L., Urioste, S., Petersen, H.V., Jensen, J.N., Eiberg, H., Barbetti, F., Serup, P., Hansen, T. and Pedersen, O., 2000. Missense mutations in the human insulin promoter factor-1 gene and their relation to maturity-onset diabetes of the young and late-onset type 2 diabetes mellitus in caucasians. The Journal of Clinical Endocrinology & Metabolism, 85(3), pp.1323-1326.

Harmon, J.S., Gleason, C.E., Tanaka, Y., Oseid, E.A., Hunter-Berger, K.K. and Robertson, R.P., 1999. In vivo prevention of hyperglycemia also prevents glucotoxic effects on PDX-1 and insulin gene expression. Diabetes, 48(10), pp.1995-2000.

Health Survey for England 2004: The Health of Minority Ethnic Groups- header table. In: NHS Health and Social Care Information Centre, Public Health Statistics 2005.

Hetherington, M.M. and Cecil, J.E., 2010. Gene-environment interactions in obesity. In Frontiers in Eating and Weight Regulation (Vol. 63, pp. 195-203). Karger Publishers.

Hessabi, B., Ziegler, P., Schmidt, I., Hessabi, C. and Walther, R., 1999. The nuclear localization signal (NLS) of PDX-1 is part of the homeodomain and represents a novel type of NLS. European journal of biochemistry, 263(1), pp.170-177.

Hivert, M.F., Jablonski, K.A., Perreault, L., Saxena, R., McAteer, J.B., Franks, P.W., Hamman, R.F., Kahn, S.E., Haffner, S., Meigs, J.B. and Altshuler, D., 2011. Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. Diabetes, 60(4), pp.1340-1348.

Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B.N., Lindner, T., Yamagata, K., Ogata, M., Tomonaga, O. and Kuroki, H., 1997. Mutation in hepatocyte nuclear factor–1β gene (TCF2) associated with MODY. Nature genetics, 17(4), p.384.

Hui, H. and Perfetti, R., 2002. Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood. European Journal of Endocrinology, 146(2), pp.129-141.

Hussain, M.A. and Habener, J.F., 2000. Glucagon-like peptide 1 increases glucose-dependent activity of the homeoprotein IDX-1 transactivating domain in pancreatic β-cells. Biochemical and biophysical research communications, 274(3), pp.616-619.

International Diabetes Foundation (IDF), 2011. IDF Diabetes Atlas, 5th edition. Brussels, Belgium: International Diabetes Federation. Retrieve from http://www.idf.org/diabetesatlas.

International Diabetes Foundation (IDF), 2013. IDF Diabetes Atlas, 6th edition. Brussels, Belgium: International Diabetes Federation. Retrieve from http://www.idf.org/diabetesatlas.

International Diabetes Federation (IDF), 2015. IDF Diabetes Atlas, 8th edn. Brussels, Belgium: International Diabetes Federation. Retrieve from http://www.idf.org/diabetesatlas.

Ioja, S., Chasens, E.R., Ng, J., Strollo, P.J. and Korytkowski, M.T., 2017. Obstructive sleep apnea in adults with type 1 and type 2 diabetes: perspectives from a quality improvement initiative in a university-based diabetes center. BMJ Open Diabetes Research and Care, 5(1), p.e000433.

Johansen, J.S., Harris, A.K., Rychly, D.J. and Ergul, A., 2005. Oxidative stress and the use of antioxidants in diabetes: linking basic science to clinical practice. Cardiovascular diabetology, 4(1), p.5.

Jonsson, J., Carlsson, L., Edlund, T. and Edlund, H., 1994. Insulin-promoter-factor 1 is required for pancreas development in mice. Nature, 371(6498), p.606.

Jonsson, J., Ahlgren, U., Edlund, T. and Edlund, H., 1995. IPF1, a homeodomain protein with a dual function in pancreas development. Int. J. Dev. Biol. 39, 789–798

Jonsson, J., Ahlgren, U., Edlund, T. and Edlund, H., 2003. IPF1, a homeodomain protein with a dual function in pancreas development. International Journal of Developmental Biology, 39(5), pp.789-798.

Kahn, S.E., 2003. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. Diabetologia, 46(1), pp.3-19.

Kakehi, T., Hisatomi, A., Kuzuya, H., Yoshimasa, Y., Okamoto, M., Yamada, K., Nishimura, H., Kosaki, A., Nawata, H. and Umeda, F., 1988. Defective processing of insulin-receptor precursor in

cultured lymphocytes from a patient with extreme insulin resistance. The Journal of clinical investigation, 81(6), pp.2020-2022.

Kaneto, H., Miyatsuka, T., Fujitani, Y., Noguchi, H., Song, K.H., Yoon, K.H. and Matsuoka, T.A., 2007. Role of PDX-1 and MafA as a potential therapeutic target for diabetes. Diabetes Research and Clinical Practice, 77(3), pp.S127-S137.

Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., Stengård, J. and Kesäniemi, Y.A., 1992. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. Diabetologia, 35(11), pp.1060-1067.

Karim, M.A., Wang, X., Hale, T.C. and Elbein, S.C., 2005. Insulin promoter factor 1 variation is associated with type 2 diabetes in African Americans. BMC medical genetics, 6(1), p.37.

Kawamori, D., Kajimoto, Y., Kaneto, H., Umayahara, Y., Fujitani, Y., Miyatsuka, T., Watada, H., Leibiger, I.B., Yamasaki, Y. and Hori, M., 2003. Oxidative stress induces nucleo-cytoplasmic translocation of pancreatic transcription factor PDX-1 through activation of c-Jun NH2-terminal kinase. Diabetes, 52(12), pp.2896-2904.

King, H. and Rewers, M., 1993. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. Diabetes care, 16(1), pp.157-177.

Kishi, A., Nakamura, T., Nishio, Y., Maegawa, H. and Kashiwagi, A., 2003. Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation. American Journal of Physiology-Endocrinology and Metabolism, 284(4), pp.E830-E840.

Knip, M. and Åkerblom, H.K., 1999. Environmental factors in the pathogenesis of type 1 diabetes mellitus. Experimental and clinical endocrinology & diabetes, 107(S 03), pp.S93-S100.

Kyvik, K.O., Green, A. and Beck-Nielsen, H., 1995. Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. Bmj, 311(7010), pp.913-917.

Lebrun, P., Montminy, M.R. and Van Obberghen, E., 2005. Regulation of the pancreatic duodenal homeobox-1 protein by DNA-dependent protein kinase. Journal of Biological Chemistry, 280(46), pp.38203-38210.

Liu, L., Furuta, H., Minami, A., Zheng, T., Jia, W., Nanjo, K. and Xiang, K., 2007. A novel mutation, Ser159Pro in the NeuroD1/BETA2 gene contributes to the development of diabetes in a Chinese potential MODY family. Molecular and cellular biochemistry, 303(1-2), pp.115-120.

Liu, Z., Tanabe, K., Bernal-Mizrachi, E. and Permutt, M.A., 2008. Mice with beta cell overexpression of glycogen synthase kinase-3β have reduced beta cell mass and proliferation. Diabetologia, 51(4), pp.623-631.

Longo, A., Guanga, G.P. and Rose, R.B., 2007. Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. Biochemistry, 46(11), pp.2948-2957.

Lovejoy, J.C., Most, M.M., Lefevre, M., Greenway, F.L. and Rood, J.C., 2002. Effect of diets enriched in almonds on insulin action and serum lipids in adults with normal glucose tolerance or type 2 diabetes. The American journal of clinical nutrition, 76(5), pp.1000-1006.

Lyssenko, V., Nagorny, C.L., Erdos, M.R., Wierup, N., Jonsson, A., Spégel, P., Bugliani, M., Saxena, R., Fex, M., Pulizzi, N. and Isomaa, B., 2009. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nature genetics, 41(1), p.82.

Macfarlane, W.M., Frayling, T.M., Ellard, S., Evans, J.C., Allen, L.I., Bulman, M.P., Ayres, S., Shepherd, M., Clark, P., Millward, A. and Demaine, A., 1999. Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. The Journal of clinical investigation, 104(9), pp.R33-R39.

Mahtani, M.M., Widén, E., Lehto, M., Thomas, J., McCarthy, M., Brayer, J., Bryant, B., Chan, G., Daly, M., Forsblom, C. and Kanninen, T., 1996. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. Nature genetics, 14(1), p.90.

Majithia, A.R. and Florez, J.C., 2009. Clinical translation of genetic predictors for type 2 diabetes. Current opinion in endocrinology, diabetes, and obesity, 16(2), p.100.

Malecki, M.T., Jhala, U.S., Antonellis, A., Fields, L., Doria, A., Orban, T., Saad, M., Warram, J.H., Montminy, M. and Krolewski, A.S., 1999. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. Nature genetics, 23(3), p.323.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I. and Lanczycki, C.J., 2014. CDD: NCBI's conserved domain database. Nucleic acids research, 43(D1), pp.D222-D226.

Matsuoka, T.A., Kaneto, H., Stein, R., Miyatsuka, T., Kawamori, D., Henderson, E., Kojima, I., Matsuhisa, M., Hori, M. and Yamasaki, Y., 2007. MafA regulates expression of genes important to islet β-cell function. Molecular Endocrinology, 21(11), pp.2764-2774.

McCarthy, M.I., Froguel, P. and Hitman, G.A., 1994. The genetics of non-insulin-dependent diabetes mellitus: tools and aims. Diabetologia, 37(10), pp.959-968.

McCarthy, M.I. and Zeggini, E., 2009. Genome-wide association studies in type 2 diabetes. Current diabetes reports, 9(2), pp.164-171.

McGuffin, L.J., Bryson, K. and Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics, 16(4), pp.404-405.

McKinnon, C.M. and Docherty, K., 2001. Pancreatic duodenal homeobox–1, PDX-1, a major regulator of beta cell identity and function. Diabetologia, 44(10), pp.1203-1214.

Meigs, J.B., Cupples, L.A. and Wilson, P.W., 2000. Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes, 49(12), pp.2201-2207.

Melloul, D. and Cerasi, E., 1994. Transcription of the insulin gene: towards defining the glucosesensitivecis-element and trans-acting factors. Diabetologia, 37(2), pp.S3-S10.

Melloul, D., Marshak, S. and Cerasi, E., 2002. Regulation of insulin gene transcription. Diabetologia, 45(3), pp.309-326.

Miller, C.P., McGehee, R.E. and Habener, J.F., 1994. IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene. The EMBO Journal, 13(5), pp.1145-1156.

Moede, T., Leibiger, B., Pour, H.G., Berggren, P.O. and Leibiger, I.B., 1999. Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1. FEBS letters, 461(3), pp.229-234.

Mosley, A.L., Corbett, J.A. and Ozcan, S., 2004. Glucose regulation of insulin gene expression requires the recruitment of p300 by the β-cell-specific transcription factor Pdx-1. Molecular endocrinology, 18(9), pp.2279-2290.

Natali, A., Galvan, A.Q., Pecori, N., Sanna, G., Toschi, E. and Ferrannini, E., 1998. Vasodilation with sodium nitroprusside does not improve insulin action in essential hypertension. Hypertension, 31(2), pp.632-636.

Nicole Cossrow & Bonita Falkner, 2004. The Journal of Clinical Endocrinology & Metabolism, Volume 89, Issue 6, 1 June 2004, pp. 2590–2594.

Nijpels, G., 1998. Determinants for the progression from impaired glucose tolerance to non-insulin-dependent diabetes mellitus. European journal of clinical investigation, 28, pp.8-13.

Noctor, E. and Dunne, F.P., 2015. Type 2 diabetes after gestational diabetes: the influence of changing diagnostic criteria. World journal of diabetes, 6(2), p.234.

Ogurtsova, K., da Rocha Fernandes, J.D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N.H., Cavan, D., Shaw, J.E. and Makaroff, L.E., 2017. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice, 128, pp.40-50.

Oldroyd, J., Banerjee, M., Heald, A. and Cruickshank, K., 2005. Diabetes and ethnic minorities. Postgraduate medical journal, 81(958), pp.486-490.

Ozougwu, J.C., Obimba, K.C., Belonwu, C.D. and Unakalamba, C.B., 2013. The pathogenesis and pathophysiology of type 1 and type 2 diabetes mellitus. Journal of Physiology and Pathophysiology, 4(4), pp.46-57.

Pamidi, S., Wroblewski, K., Broussard, J., Day, A., Hanlon, E.C., Abraham, V. and Tasali, E., 2012. Obstructive sleep apnea in young lean men: impact on insulin sensitivity and secretion. Diabetes care, 35(11), pp.2384-2389.

Pan, X.R., Li, G.W., Hu, Y.H., Wang, J.X., Yang, W.Y., An, Z.X., Hu, Z.X., Xiao, J.Z., Cao, H.B., Liu, P.A. and Jiang, X.G., 1997. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study. Diabetes care, 20(4), pp.537-544.

Parra, E.J., Hoggart, C.J., Bonilla, C., Dios, S., Norris, J.M., Marshall, J.A., Hamman, R.F., Ferrell, R.E., McKeigue, P.M. and Shriver, M.D., 2004. Relation of type 2 diabetes to individual admixture and candidate gene polymorphisms in the Hispanic American population of San Luis Valley, Colorado. Journal of medical genetics, 41(11), pp.e116-e116.

Peng, Y., Sun, L., Jia, Z., Li, L. and Alexov, E., 2017. Predicting protein DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver. Bioinformatics, 34(5), pp.779-786.

Petersen, H.V., Serup, P., Leonard, J., Michelsen, B.K. and Madsen, O.D., 1994. Transcriptional regulation of the human insulin gene is dependent on the homeodomain protein STF1/IPF1 acting through the CT boxes. Proceedings of the National Academy of Sciences, 91(22), pp.10465-10469.

Petretto, E., Liu, E.T. and Aitman, T.J., 2007. A gene harvest revealing the archeology and complexity of human disease.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry, 25(13), pp.1605-1612.

Pillay, T. S., Langlois, W. J., & Olefsky, J. M., 1995. The Genetics of Non-insulin-Dependent Diabetes Mellitus. In Advances in genetics (Vol. 32, pp. 51-98). Academic Press.

Pinhas-Hamiel, O. and Zeitler, P., 2005. The global spread of type 2 diabetes mellitus in children and adolescents. The Journal of pediatrics, 146(5), pp.693-700.

Poulsen, P., Kyvik, K.O., Vaag, A. and Beck-Nielsen, H., 1999. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance–a population-based twin study. Diabetologia, 42(2), pp.139-145.

Reis, A.F., Ye, W.Z., Dubois-Laforgue, D., Bellanne-Chantelot, C., Timsit, J. and Velho, G., 2000. Mutations in the insulin promoter factor-1 gene in late-onset type 2 diabetes mellitus. European journal of endocrinology, 143(4), pp.511-513.

Redondo, M.J., Fain, P.R. and Eisenbarth, G.S., 2001. Genetics of type 1A diabetes. Recent progress in hormone research, 56, pp.69-89.

Robertson, R.P. and Harmon, J.S., 2007. Pancreatic islet β-cell and oxidative stress: The importance of glutathione peroxidase. FEBS letters, 581(19), pp.3743-3748.

Röder, P.V., Wu, B., Liu, Y. and Han, W., 2016. Pancreatic regulation of glucose homeostasis. Experimental & molecular medicine, 48(3), p.e219.

Roy, M.N., Biswas, K.B., Siddiqua, N., Arslan, M.I. and Ali, L., 2007. Determinants of insulin secretion and sensitivity in Bangladeshi type 2 diabetic subjects. Metabolic syndrome and related disorders, 5(3), pp.275-281.

Sanghera, D.K. and Blackett, P.R., 2012. Type 2 diabetes genetics: beyond GWAS. Journal of diabetes & metabolism, 3(198).

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. and Hughes, T.E., 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science, 316(5829), pp.1331-1336.

Sayeed, M.A., Mahtab, H., Khanam, P.A., Latif, Z.A., Banu, A. and Khan, A.K., 2007. Prevalence of diabetes and impaired fasting glucose in urban population of Bangladesh. Bangladesh Medical Research Council Bulletin, 33(1), pp.1-12.

Schober, A.K., Neurath, M.F. and Harsch, I.A., 2011. Prevalence of sleep apnoea in diabetic patients. The clinical respiratory journal, 5(3), pp.165-172.

Shepherd, M., Ellis, I., Ahmad, A.M., Todd, P.J., Bowen-Jones, D., Mannion, G., Ellard, S., Sparkes, A.C. and Hattersley, A.T., 2001. Predictive genetic testing in maturity-onset diabetes of the young (MODY). Diabetic medicine, 18(5), pp.417-421.

Shiau, M.Y., Huang, C.N., Liao, J.H. and Chang, Y.H., 2004. Missense mutations in the human insulin promoter factor-1 gene are not a common cause of type 2 diabetes mellitus in Taiwan. Journal of endocrinological investigation, 27(11), pp.1076-1080.

Shimokata, H., Muller, D.C., Fleg, J.L., Sorkin, J., Ziemba, A.W. and Andres, R., 1991. Age as independent determinant of glucose tolerance. Diabetes, 40(1), pp.44-51.

Skyler, J.S. and Ricordi, C., 2011. Stopping type 1 diabetes: attempts to prevent or cure type 1 diabetes in man. Diabetes, 60(1), pp.1-8.

So, W.Y., Ng, M.C., Lee, S.C., Sanke, T., Lee, H.K. and Chan, J.C., 2000. Genetics of types 2 diabetes mellitus. Hong Kong Medical Journal, 6(1), pp.69-76.

Stride, A. and Hattersley, A.T., 2002. Different genes, different diabetes: lessons from maturity-onset diabetes of the young. Annals of medicine, 34(3), pp.207-216.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. and Balkau, B., 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature, 445(7130), p.881.

Staffers, D.A., Ferrer, J., Clarke, W.L. and Habener, J.F., 1997. Early-onset type-ll diabetes mellitus (MODY4) linked to IPF1. Nature genetics, 17(2), p.138.

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S. and Baker, A., 2007. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nature genetics, 39(6), p.770.

Stoffers, D.A., Chèvre, J.C., Durand, E., Stanojevic, V., Dina, C., Habener, J.F. and Froguel, P., 1999. Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. The Journal of clinical investigation, 104(9), pp.R41-R48.

Sujjitjoon, J., Jungtrakoon, P., Boonyasrisawat, W., Chongjaroen, N., Chukijrungroat, T., Kooptiwut, S., Plengvidhya, N., Banchuin, N. and Yenchitsomanus, P.T., 2012. Molecular genetics of monogenetic beta-cell diabetes. Genomics and Genetics, 1(2), pp.93-108.

Szoke, E. and Gerich, J.E., 2005. Role of impaired insulin secretion and insulin resistance in the pathogenesis of type 2 diabetes mellitus. Comprehensive therapy, 31(2), pp.106-112.

Thomas, K.M., Drevets, W.C., Whalen, P.J., Eccard, C.H., Dahl, R.E., Ryan, N.D. and Casey, B.J., 2001. Amygdala response to facial expressions in children and adults. Biological psychiatry, 49(4), pp.309-316.

Tontonoz, P., Hu, E. and Spiegelman, B.M., 1994. Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor. Cell, 79(7), pp.1147-1156.

Trinder, P., 1969. Determination of glucose in blood using glucose oxidase with an alternative oxygen acceptor. Annals of clinical Biochemistry, 6(1), pp.24-27.

Tripathy, D. and Chavez, A.O., 2010. Defects in insulin secretion and action in the pathogenesis of type 2 diabetes mellitus. Current diabetes reports, 10(3), pp.184-191.

Tuomilehto, J., Lindström, J., Eriksson, J.G., Valle, T.T., Hämäläinen, H., Ilanne-Parikka, P., Keinänen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M. and Salminen, V., 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine, 344(18), pp.1343-1350.

Tuomilehto, J., 2013. The emerging global epidemic of type 1 diabetes. Current diabetes reports, 13(6), pp.795-804.

Turner, R.C., Millns, H., Neil, H.A.W., Stratton, I.M., Manley, S.E., Matthews, D.R. and Holman, R.R., 1998. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23). Bmj, 316(7134), pp.823-828.

Unger, R.H. and Scherer, P.E., 2010. Gluttony, sloth and the metabolic syndrome: a roadmap to lipotoxicity. Trends in Endocrinology & Metabolism, 21(6), pp.345-352.

Vaag, A., Henriksen, J.E., Madsbad, S., Holm, N. and Beck-Nielsen, H., 1995. Insulin secretion, insulin action, and hepatic glucose production in identical twins discordant for non-insulin-dependent diabetes mellitus. The Journal of clinical investigation, 95(2), pp.690-698.

Vaxillaire, M. and Froguel, P., 2008. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. Endocrine Reviews, 29(3), pp.254-264.

Vionnet, N., Stoffel, M., Takeda, J., Yasuda, K., Bell, G.I., Zouali, H., Lesage, S., Velho, G., Iris, F., Passa, P.H. and Froguel, P.H., 1992. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature, 356(6371), p.721.

Waeber, G., Thompson, N., Nicod, P. and Bonny, C., 1996. Transcriptional activation of the GLUT2 gene by the IPF-1/STF-1/IDX-1 homeobox factor. Molecular Endocrinology, 10(11), pp.1327-1334.

Weijnen, C.F., Rich, S.S., Meigs, J.B., Krolewski, A.S. and Warram, J.H., 2002. Risk of diabetes in siblings of index cases with Type 2 diabetes: implications for genetic studies. Diabetic Medicine, 19(1), pp.41-50.

Weng, J., Macfarlane, W.M., Lehto, M., Gu, H.F., Shepherd, L.M., Ivarsson, S.A., Wibell, L., Smith, T. and Groop, L.C., 2001. Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations. Diabetologia, 44(2), pp.249-258.

Wescott, M.P., Rovira, M., Reichert, M., Von Burstin, J., Means, A., Leach, S.D. and Rustgi, A.K., 2009. Pancreatic ductal morphogenesis and the Pdx1 homeodomain transcription factor. Molecular biology of the cell, 20(22), pp.4838-4844.

Wild, S., Roglic, G., Green, A., Sicree, R. and King, H., 2004. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes care, 27(5), pp.1047-1053.

Winter, W.E. and Silverstein, J.H., 2000. Molecular and genetic bases for maturity onset diabetes of youth. Current opinion in pediatrics, 12(4), pp.388-393.

Winter, W.E., 2003. Newly defined genetic diabetes syndromes: maturity onset diabetes of the young. Reviews in Endocrine and Metabolic Disorders, 4(1), pp.43-51.

World Health Organization (WHO), 1999. Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus.

World Health Organization (WHO), 2014. 'Non communicable Diseases Country Profiles', Genève□: WHO Press, 2014, pp.1–210.

World Health Organization (WHO), 2016. Global report on diabetes. World Health Organization. Geneva, Switzerland. Retrived from https://www.who.int

World Health Organization (WHO), 2018. Diabetes. Retrived on oct, 2018. Retrived from https://www.who.int/news-room/fact-sheets/detail/diabetes.

Yoshiuchi, I., Yamagata, K., Yang, Q., Iwahashi, H., Okita, K., Yamamoto, K., Oue, T., Imagawa, A., Hamaguchi, T., Yamasaki, T. and Horikawa, Y., 1999. Three new mutations in the hepatocyte nuclear factor-1α gene in Japanese subjects with diabetes mellitus: clinical features and functional characterization. Diabetologia, 42(5), pp.621-626.

Zeggini, E., 2007. A new era for type 2 diabetes genetics. Diabetic Medicine, 24(11), pp.1181-1186.

Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G. and Ardlie, K., 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature genetics, 40(5), p.638.

Zimmet, P., Alberti, K.G.M.M. and Shaw, J., 2001. Global and societal implications of the diabetes epidemic. Nature, 414(6865), p.782.

APPENDIX

TBE buffer preparation	TBE buffer is a buffer solution made up of Tris base, boric acid and EDTA (or Tris-borate-EDTA altogether). Making concentrated (5x) stock solution of TBE by weighing 54 g Tris base (FW = 121.14) and 27.5 g boric acid (FW = 61.83) and dissolving both in approximately 700 milliliters of deionized water. Then add 20 milliliters of 0.5 M EDTA (pH 8.0). Adjust the pH at 8.0. And then adjust the volume up to 1 Liter. Making working solution from (5x) stock TBE: dilute the stock buffer at 1:4 (stock buffer: double distil water) ratio.
NCBI	The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). It's an online database that contains GenBank, alignment tool (BLAST), books, and scientific publications etc.
MEGA X	Molecular Evolutionary Genetics Analysis (MEGA) X is computer software for aligning multiple sequences also it is use for conducting statistical analysis of molecular evolution and for constructing phylogenetic trees. It includes many sophisticated methods and tools for phylogenomics and phylomedicine. It is licensed as proprietary freeware.
DNA/RAN counter 2.0	It is an online based software that can count number of nucleotide of any sequence.
RevTrans 2.0b Server	The RevTrans 2.0b is a server that is a part of the DTU bioinformatics under department of Bio and Health Informatics. It can convert any query nucleotide sequence into its reveres sequence virtually. Also can translate them, aligns the peptide sequences, and uses this as a scaffold for constructing the corresponding DNA multiple alignment.
ProtParam	ProtParam (References / Documentation) is a tool which allows the

	computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. This tool slog with ExPAsy (Expert Protein Analysis System) server.
PSIPRED	PSI-blast based secondary structure PREDiction (PSIPRED) is a method used to investigate protein structure. It is a server-side program, featuring a website serving as a front-end interface, which can predict a protein's secondary structure (beta sheets, alpha helixes and coils) from the primary sequence. It is available as a web service and as software.
PDBsum	PDBsum is a database that provides an overview of the contents of each 3D macromolecular structure deposited in the Protein Data Bank.
PDB	The Protein Data Bank (PDB) is a database for the three-dimensional structural data of large biological molecules, such as proteinsand nucleic acids. The data, typically obtained by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, and submitted by biologists and biochemists from around the world, are freely accessible on the Internet via the websites of its member organizations.
UniPortKB	UniProt Knowledgebase (UniProtKB) is a protein database partially curated by experts, consisting of two sections: UniProtKB/Swiss-Prot (containing reviewed, manually annotated entries) and UniProtKB/TrEMBL (containing unreviewed, automatically annotated entries).
CDD	Conserved Domain Database (CDD) is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins.
Swiss-Model	SWISS-MODEL is a structural bioinformatics web-server dedicated to homology modeling of 3D protein structures.
UCSF Chimera 1.11.2	UCSF Chimera 1.11.2 is an extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories,

	and conformational ensembles.
SAMPDI	SAMPDI Web Server provides fast and accurate predictions for the effects of single amino acid substitution on the binding free energy of protein-DNA complex. SAMPDI is running on Clemson University's Palmetto Supercomputer Cluster.
PROCheck	PROCheck checks the stereochemical quality of a protein structure, producing several PostScript plots analyzing its overall and residue-by-residue geometry.
Cn3D	Cn3D is a Windows that acts as a helper application for web browsers to view three-dimensional structures.
MAFFT-method	It is an algorithm that is used in RevTrans 2.06 server which shows the transcribed amino acid sequence with corresponding nucleotides.