CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Elucidating the Molecular Mechanism of Prediabetic Insulin Resistance

by

Naveed Iqbal Soomro

A dissertation submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

Elucidating the Molecular Mechanism of Prediabetic Insulin Resistance

By

Naveed Iqbal Soomro

(DBS181004)

Dr. Javed A Bhalli, Senior Director, DMPK

Frontage Laboratories, Concord, Ohio, USA

Foreign Evaluator No. 1

Dr. Sabeela Qasim Ahmed, RCGP/GMC Examiner

University of Leeds, UK

Foreign Evaluator No. 2

Dr. Syeda Marriam Bakhtiar

(Research Supervisor)

Dr. Syeda Marriam Bakhtiar

(Head, Department of Bioinformatics and Biosciences)

Dr. Sahar Fazal

(Dean, Faculty of Health and Life Sciences)

DEPARTMENT OF BIOINFORMATICS AND BIOSCIENCES

CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

ISLAMABAD

2025

I dedicate this dissertation to my parents specially my father and family, whose unconditional love, unwavering support, and endless encouragement have been my pillars of strength throughout this challenging journey. Their belief in me and sacrifices made have been the driving force behind my success. I am grateful for their patience, understanding, and constant motivation. This achievement is a testament to their unwavering faith in my abilities. I dedicate this work to them as a token of my deepest appreciation and love.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY ISLAMABAD

Expressway, Kahuta Road, Zone-V, Islamabad
Phone:+92-51-111-555-666 Fax: +92-51-4486705
Email: info@cust.edu.pk Website: https://www.cust.edu.pk

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in the dissertation, entitled "**Elucidating the Molecular Mechanism of Prediabetic Insulin Resistance**" was conducted under the supervision of **Dr. Syeda Marriam Bakhtiar**. No part of this dissertation has been submitted anywhere else for any other degree. This dissertation is submitted to the **Department of Bioinformatics & Biosciences, Capital University of Science and Technology** in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the field of **Biosciences**. The open defence of the dissertation was conducted on **April 28, 2025.**

Student Name : Naveed Iqbal Soomro (DBS181004)

The Examination Committee unanimously agrees to award PhD degree in the mentioned field.

Examination Committee :

(a) External Examiner 1: Dr. Muhammad Jawad Hassan
Professor
Shifa Int. Hospital, Islamabad

(b) External Examiner 2: Dr. Samiullah Khan
Associate Professor
QAU, Islamabad

(c) Internal Examiner : Dr. Sohail Ahmed Jan
Associate Professor
CUST, Islamabad

Supervisor Name : Dr. Syeda Marriam Bakhtiar
Associate Professor
CUST, Islamabad

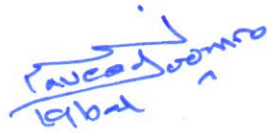
Name of HoD : Syeda Marriam Bakhtiar
Associate Professor
CUST, Islamabad

Name of Dean : Dr. Sahar Fazal
Professor
CUST, Islamabad

AUTHOR'S DECLARATION

I, **Naveed Iqbal Soomro** (Registration No. **DBS181004)**, hereby state that my dissertation titled, '**Elucidating the Molecular Mechanism of Prediabetic Insulin Resistance**' is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/ world.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my PhD Degree.



(**Naveed Iqbal Soomro**)

Dated: 28, April, 2025

Registration No : DBS181004

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the dissertation titled "**Elucidating the Molecular Mechanism of Prediabetic Insulin Resistance**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete dissertation has been written by me.

I understand the zero-tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled dissertation declare that no portion of my dissertation has been plagiarized and any material used as reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled dissertation even after award of PhD Degree, the University reserves the right to withdraw/ revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/ University Website on which names of students are placed who submitted plagiarized dissertation.

(**Naveed Iqbal Soomro**)

Dated: 28, April, 2025

Registration No : DBS181004

List of Publications

It is certified that following publication(s) have been made out of the research work that has been carried out for this dissertation:-

1. **Naveed Iqbal Soomro** and Syeda Marriam Bakhtiar, "Genetic Explication of Impaired Insulin Resistance in Genesis of Metabolic Diseases: Impaired Insulin Resistance", *J Popl Ther Clin Pharmacol*, vol. 31, no. 6, pp. 2655–2669, Jun. 2024, doi:10.53555/jptcp.v31i6.6988..

(Naveed Iqbal Soomro)

Registration No: DBS181004

Acknowledgement

In the name of Allah, the Most Gracious and the Most Merciful.

I am deeply thankful to **Almighty Allah**, whose bounteous blessings have enabled me to complete this research project and write this dissertation. He bestowed upon us the **Holy Quran**, which serves as guidance for the reverent, a cure for diseases, and a blessing for the believers. I am equally grateful to **Prophet Muhammad (Peace be upon Him)**, the most perfect and the best among those ever born on the surface of the earth, who enlightens the hearts of believers throughout their lives. I wish to acknowledge several key figures who have significantly contributed to my research endeavor.

I express my heartfelt gratitude to all those who have supported the successful completion of my PhD dissertation. This journey has been a significant milestone in my academic and personal growth, and I am deeply thankful for the support, guidance, and encouragement I have received throughout the process.

Firstly, I present my heartfelt thanks to **Dr. Sahar Fazal**, Professor/Dean of the Faculty of Health and Life Sciences at the Capital University of Science and Technology, Islamabad. She is a kind and dedicated teacher, serving as a maxim for all students. I find it a perfect opportunity to express my deep thanks and admiration to a person of extreme dedication in the field of Biosciences, my supervisor, Associate Professor/HoD **Dr. Syeda Marriam Bakhtiar**. She provided me with the opportunity to work under her excellent supervision. Her patience, valuable suggestions, and kind interest enabled me to complete my research work. She is an inspiration to me, and I am honored to be under her supervision. May Almighty Allah bless her with perfect health and the potential to carry on her valuable work, Ameen.

I would also like to extend my gratitude towards **Dr. Erum Dilshad**, Associate Professor; **Dr. Arshia Amin Butt**, Assistant Professor; **Dr. Sohail Ahmad Jan**, Assistant Professor; and **Dr. Sami Ullah Jan**, Senior Lecturer; **Dr khalid Mahmmod**, Director Graduate Studies, and all staff of **Capital University of**

Science and Technology, Islamabad. I am highly obliged for their assistance, support, and always being there whenever I needed help.

Completion of research work necessitates a peaceful working environment, proper instrument management, and pleasant company. I wish to express my heartiest thanks to my Ph.D. fellows, Muhammad Qasim Khan and Sobia Khurshid, for their perfect company and cooperation. I would also like to express my sincere thanks to my seniors and juniors for their support and companionship.

I extend my feelings of love and thanks to my friends for their love, support, valuable company, and unlimited moral support during my research work and write-up. No words in any dictionary can acknowledge the sacrifices, love, and moral support given to me by my parents, **Amir Bukhsh Soomro (late)** and **Gulzar Amir Soomro**. My studies and success are wholly and solely attributed to them. Their desires inspire me to excel further in my studies.

Special acknowledgments to my wife, Maira, who has always been there for me during all ups and downs. Without her, I could never have accomplished my Ph.D. degree. I owe my deep admiration to my brothers, Dr. Javed Iqbal, Saeed Iqbal, and Junaid Iqbal, who supported me not only during my research work but throughout my life.

In the end, I want to express my indebtedness to all those who prayed for my betterment and serenity.

(Naveed Iqbal Soomro)

Abstract

Prediabetic insulin resistance represents a crucial stage in the progression towards type 2 diabetes, necessitating the identification of robust biomarkers for early detection and intervention. This study employed a comprehensive approach integrating meta-analysis and advanced in-silico techniques to identify and characterize biomarkers associated with prediabetic insulin resistance.

A rigorous meta-analysis was conducted to systematically review and analyze existing literature, identifying a diverse array of genes and proteins implicated in the dysregulation observed in both diabetes and prediabetes. Key genes such as SGCZ (Sarcoglycan Zeta), HPSE2 (Heparanase 2), ADGRA1 (Adhesion G Protein-Coupled Receptor A1), GLB1L3 (Galactosidase Beta 1 Like 3), PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6), SIRT1 (Sirtuin 1), PPAR (Peroxisome Proliferator-Activated Receptor), PGC1 alpha (Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha), NRF1 (Nuclear Respiratory Factor 1), TG gene (Thyroglobulin), and HBA1c (Hemoglobin A1c) were identified as playing significant roles in these metabolic disorders.

We utilized bioinformatics tools including GeneMania, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gephi for comprehensive gene interaction network analysis, pathway validation, and network construction. This foundational research underscored the intricate genetic and molecular interactions that contribute to the pathophysiology of metabolic diseases, providing a robust starting point for the current investigation.

To expand upon these findings, an initial set of biomarkers (set no. 1) was curated based on the meta-analysis results. Subsequently, an interaction network of set no. 1 genes was generated using the FunCoup server, which facilitated the identification of additional genes (set no. 2) interacting with the initially identified biomarkers. Enrichment analysis was then performed using gene ontology tools to elucidate the biological, molecular, and cellular annotations of these genes, providing deeper insights into their functional roles in metabolic pathways.

Protein sequences and IDs corresponding to the identified genes were retrieved from the UniProt database, and their structural stability was assessed via using BLASTP to ensure homology and ProSA server to validate the modeled proteins with negative Z-scores, indicating structural integrity.

Furthermore, protein-protein docking studies were conducted using the GRAMM-X web server and ClusPro to explore the interactions between set no. 1 biomarkers and set no. 2 novel biomarkers. The results were analyzed using Discovery Studio, revealing ARG13 as the most frequently interacting amino acid residue across all docked complexes. This finding suggested that ARG13 plays a pivotal role in the interaction interfaces between these biomarkers, potentially influencing their functional implications in prediabetic insulin resistance.

In conclusion, the study highlights the complex interplay of genetic, protein, and molecular factors in the pathophysiology of diabetes and prediabetes. The identification of common interacting amino acid residues through molecular docking provides valuable insights into potential novel biomarkers for prediabetic insulin resistance. These findings pave the way for targeted therapeutic strategies aimed at mitigating the progression of metabolic disorders, ultimately contributing to improved patient outcomes and healthcare management.

Keywords: Metabolic Disease, Prediabetes, Insulin Resistance, Inflammation, Pathways, Regulatory Networks, Docking

Contents

List of Figures

List of Tables

Abbreviations

A1C	Glycated Hemoglobin
Ac-CoA	Acetyl-CoA
ADIPOQ	Adiponectin, C1Q, and Collagen Domain-Containing
AMPK	AMP-Activated Protein Kinase
ATP	Adenosine Triphosphate
BMI	Body Mass Index
CAM-KII	Calmodulin Kinase II
CM-R	Chylomicron Remnants
CM-TG	Chylomicron-Triglycerides
CVDs	Cardiovascular Diseases
FAs	Fatty Acids
FFA	Free Fatty Acids
FPG	Fasting Plasma Glucose
FTO	Fat Mass and Obesity-Associated Protein
GCKR	Glucokinase Regulatory Protein
GDM	Gestational Diabetes Mellitus
GIP	Gastric Inhibitory Peptide
GLP-1	Glucagon-like Peptide 1
GLUT4	Glucose Transporter Type 4
GP	Glycogen Phosphorylase
GPCR	G-Protein-Coupled Receptor
GS	Glycogen Synthase
HbA1c	Hemoglobin A1c

HL	Hepatic Lipase
HOMA-IR	Homeostasis Model Assessment of Insulin Resistance
IDF	International Diabetes Federation
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IMCL	Intramyocellular Lipid
IR	Insulin Resistance
IRS1/2	Insulin Receptor Substrate 1/2
IRS1	Insulin Receptor Substrate 1
IRS2	Insulin Receptor Substrate 2
LADA	Late-Onset Autoimmune Diabetes of the Adult
LPA	Lysophosphatidic Acid
MetS	Metabolic Syndrome
MODY	Maturity-Onset Diabetes of the Young
NAFLD	Non-Alcoholic Fatty Liver Disease
OGTT	Oral Glucose Tolerance Test
PCOS	Polycystic Ovary Syndrome
PKCε	Protein Kinase C epsilon
PKCθ	Protein Kinase C theta
PPARG	Peroxisome Proliferator-Activated Receptor Gamma
SIRT1	Sirtuin 1
SNP	Single Nucleotide Polymorphism
T2D	Type 2 Diabetes
TNF	Tumor Necrosis Factor
VLDL	Very Low-Density Lipoprotein
βAR	Beta-Adrenergic Receptor

Symbols

$\geq$	Greater than or equal to
$<$	Less than
$>$	Greater than
$\leq$	Less than or equal to
$mmol/L$	Millimoles per Liter
mg/dL	Milligrams per Deciliter
g	Gram
α	Alpha
β	Beta
γ	Gamma
θ	Theta
λ	Lambda

Chapter 1

Introduction

1.1 Background of the Study

Insulin resistance (IR) refers to an altered response of glucose uptake, typically stimulated by insulin, or can be described as reduced sensitivity of the body to insulin, impacting metabolic processes, growth, and development. This phenomenon results in a diminution of glucose disposal to insulin-sensitive tissues despite comparable plasma levels of circulating insulin, causing hyperglycemia. Such conditions may add up setting for complicated macrovascular and microvascular complications. Insulin resistance in combination with pancreatic β cell dysfunction and inadequate insulin production ultimately progresses to prediabetes, then type 2 diabetes [1].

The pathophysiology of IR is intricate and not fully elucidated. It involves contributions from all organ systems, each playing a predominant role in different physiological states. Skeletal muscles are crucial for insulin sensitivity immediately after meals and during physical activity, while hepatic regulation of glycolysis and gluconeogenesis predominates during fasting [1]. Adipose tissue may contribute to endocrine and inflammatory functions that influence other organs, but it plays relatively little direct role in insulin sensitivity. Immune cells in adipose tissue, including macrophages 1, T, and B lymphocytes also cause insulin resistance. Remember that insulin sensitivity is an uncategorized trait, but relies on different cutoffs as it has been reported

regarding the metabolic features of the study subject population [2]. Challenges in IR determination stems from the absence of a singular parameter, with studies often resorting to indirect indicators based on fasting blood samples, and methods varying in their specificity for hepatic or muscle IR. Additionally, ethnic variations play a role, influenced by compensatory hyperinsulinemia and β cell dysfunction [3].

Prediabetes (PD) is characterized by blood glucose levels that are higher than normal but still below the threshold for diabetes diagnosis. PD is considered a high-risk state for developing type 2 diabetes, with an annual progression rate of 5% to 10%. As such, prediabetes is viewed as an intermediate stage between normal health and diabetes. It often includes impaired fasting glucose (IFG), and impaired glucose tolerance (IGT), or both. The World Health Organization (WHO) and the American Diabetes Association (ADA) have also proposed several terms for this condition, including IFG, IGT, intermediate hyperglycemia, and high-risk state for diabetes development [4].

In 2021, an estimated 464 million people worldwide were living with prediabetes, making up 9.1% of the global adult population. This number is expected to rise significantly, reaching 638 million by 2045. Similarly, in 2021, about 298 million people (5.8% of the global population) had prediabetes, with projections indicating an increase to 414 million (6.5%) by 2045. High-income countries currently have the highest prevalence of prediabetes, but the most rapid growth in cases over the coming decades is expected to occur in low-income countries. These trends highlight the urgent need for effective prevention strategies to slow the rising burden of prediabetes worldwide [6].

Over 38% of American adults are affected by prediabetes [7]. By 2022, prediabetes and diabetes rates in China declined to 15.7% and 7.6%, respectively, from 18.3% and 8.2% in 2013 ($P < 0.05$). The decrease was notable among women, the Han majority, and higher-income, educated groups, while rates remained higher among ethnic minorities and low-income populations. Medication adherence improved significantly (95.9% in 2022 vs. 76.5% in 2013, $P < 0.01$), but fewer individuals managed their diabetes actively (39.7% vs. 53.3%, $P < 0.01$). Self-monitoring rates stayed stable but increased among lower-income groups ($P < 0.05$) [8].

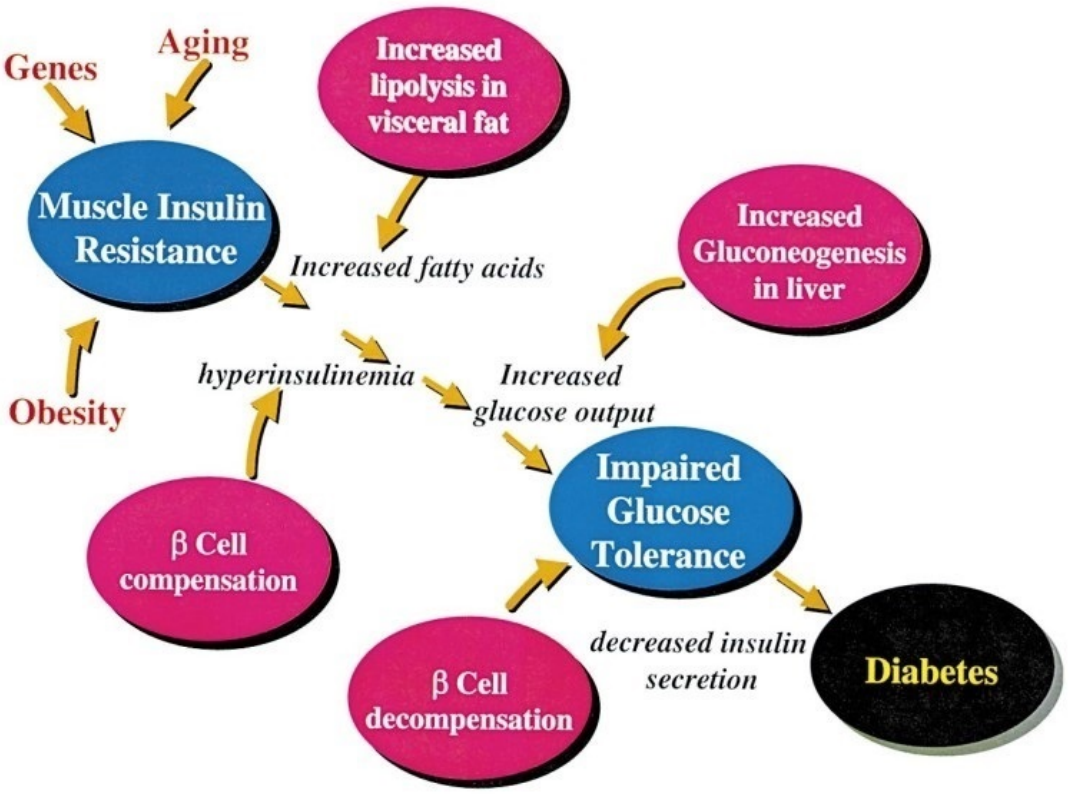


FIGURE 1.1: Insulin resistance to diabetes progression [5].

In Pakistan, according to the 2nd National Diabetes Survey of Pakistan 2016-2017, the prevalence of prediabetes is 14.4% [9]. It is observed that rapid urbanization and changes in diet in developing countries have raised the average Body Mass Index (BMI) as well as prediabetes. There is a significant role of genetics that is observed in various populations in the progression of prediabetes to diabetes [10].

Prediabetes has been observed as a major risk factor for the future development of type 2 diabetes mellitus. Along with type 2 diabetes mellitus other comorbidities (FIG. 1.1) that are associated with prediabetes are polycystic ovary syndrome (PCOS), non-alcoholic fatty liver disease (NAFLD) obesity, cardiovascular diseases, and metabolic syndrome [6].

Diagnosis of prediabetes involves a blood glucose level equal to or above a specific value, which is Fasting Plasma Glucose Test (FPG) ($\geq$ 100mg/dL or 5.5mmol/L), Oral

Glucose Tolerance Test (OGTT) ($\geq$ 140mg/dL or 7.8 mmol/L), Glycated hemoglobin or hemoglobin bounded to glucose (A1C) ($\geq$ 5.7% or 39mmol/mol) [6]. There are several limitations that can affect their accuracy and reliability. The FPG test primarily reflects the liver's glucose production but may fail to detect impaired glucose tolerance (IGT). For instance, a person may have normal fasting glucose levels but still have IGT, which would only be revealed through an OGTT. Additionally, FPG levels can vary due to stress, minor illnesses, or recent physical activity, leading to inconsistent results. For example, an individual might have a normal fasting glucose level one day (99 mg/dL) but test in the prediabetes range (102 mg/dL) another day, causing diagnostic uncertainty. Similarly, fasting longer than required (e.g., 14+ hours instead of 8-12) can result in falsely low glucose readings [4].

The OGTT, though considered a more sensitive test, has its own drawbacks. It is inconvenient, requiring fasting, consuming a glucose solution, and undergoing multiple blood draws over two hours. Patient compliance can also be an issue, as some individuals experience nausea or vomiting after consuming the glucose drink, making the test invalid. Furthermore, OGTT results can be influenced by various external factors such as sleep quality, stress, or recent physical activity. A person with normal glucose regulation might show a false-positive result ($\geq$140 mg/dL) if they had a poor night's sleep before the test. Additionally, OGTT results lack reproducibility, meaning the same person might test within the normal range one day (138 mg/dL) but show prediabetes (142 mg/dL) another day [4].

The HbA1c test, while convenient because it does not require fasting, is not always reliable in certain populations. It can be affected by hemoglobinopathies (e.g., sickle cell disease, thalassemia) and conditions like anemia, leading to falsely high or low readings. For instance, a person with sickle cell disease may have an artificially low HbA1c level, resulting in a missed prediabetes diagnosis. Ethnic and genetic variations also play a role; some populations, such as South Asians and Africans, may naturally have slightly different HbA1c levels, potentially leading to misclassification.

Moreover, HbA1c reflects average blood glucose over two to three months but does not capture short-term fluctuations. A person with frequent post-meal glucose spikes

(e.g., 180 mg/dL) but a normal average glucose may have an HbA1c in the normal range, despite having underlying glucose metabolism issues [6].

Overall, these diagnostic tests can yield inconsistent results, leading to diagnostic confusion. The same person might test positive for prediabetes with one method but negative with another. For example, an individual could have an FPG of 98 mg/dL (normal), an OGTT of 144 mg/dL (prediabetes), and an HbA1c of 5.6% (normal), making it unclear whether they truly have prediabetes. Additionally, the cut-off values for these tests are somewhat arbitrary. A slight variation in test results (e.g., OGTT of 139 mg/dL vs. 140 mg/dL) can mean the difference between a normal diagnosis and prediabetes, even though the actual risk difference may be minimal. Due to these limitations, healthcare providers often recommend a combination of tests rather than relying on a single measure to diagnose prediabetes or diabetes accurately [6].

High fasting insulin levels in prediabetic patients suggest that their bodies are becoming more resistant to insulin. This insulin resistance (IR) happens in various tissues like the muscles, liver, and fat cells. Normally, after eating, insulin should help the liver reduce glucose production, but in people with isolated impaired fasting glucose (IFG), the liver doesn't respond well, even though insulin levels are high. This points to a significant problem with how the liver reacts to insulin [11].

Insulin is supposed to slow down the breakdown of fats, but in people with impaired glucose tolerance (IGT), we see that their fasting levels of free fatty acids (FFAs) are higher, meaning that fat breakdown is happening more than it should. Despite high insulin levels, these people have more FFAs in their blood, which shows that their fat cells aren't responding to insulin as they should [8]. Both IFG and IGT are associated with this fat cell resistance to insulin. However, people with IFG mostly have severe insulin resistance in the liver while their muscles respond relatively normally. On the other hand, those with IGT experience more resistance in their muscles and a milder resistance in the liver [73]. In a healthy body, glucose is absorbed through the gut and triggers the release of hormones like GLP-1 and GIP, which help the pancreas release insulin. But in IFG and IGT, this process is disrupted. Studies that measure C-peptide levels during glucose tolerance tests have shown that the ability

of the pancreas to release insulin is significantly reduced in these conditions. This combination of insulin resistance and reduced insulin production from the pancreas is what puts people at a higher risk of developing prediabetes [12, 13].

Metabolic diseases are a group of disorders that affect the body's ability to process and use nutrients (such as carbohydrates, proteins, and fats) for energy and other essential functions [13]. These disorders can be caused by genetic mutations, environmental factors, or a combination of both. Some examples of metabolic diseases include prediabetes that leads to diabetes type 2, obesity, hypertension, and cardiovascular diseases (CVDs) [14].

Prediabetes refers to a medical condition in which blood sugar levels are elevated, yet not to the point of being classified as type 2 diabetes [15]. It acts as a warning signal that an individual might develop type 2 diabetes, a chronic disease with severe complications like kidney and heart damage [16],[17]. Prediabetes is a condition affecting approximately one in three adults in the United States and one in five adults worldwide, with higher prevalence rates observed in low- and middle-income countries [18].

Diabetes is a chronic metabolic disease associated with insulin abnormalities due to inadequate insulin secretion or ineffective use of insulin. According to the International Diabetes Federation (https://diabetesatlas.org/). 537 million adults are living with diabetes; Diabetes is responsible for 6.7 million deaths in 2021 [18].

Obesity is a condition where there is an excessive accumulation of body fat, which is typically defined by having a BMI of 30 or higher [19]. The prevalence of obesity has been steadily increasing worldwide, with over 650 million adults being classified as obese in 2016. This is a significant public health concern due to the increased risk of chronic diseases associated with obesity [20].

Hypertension, also known as high blood pressure, is a condition where the force of blood against the walls of arteries is persistently high, leading to a strain on the heart and an increased risk of health problems. It is estimated that 1.13 billion people worldwide had hypertension in 2015, and this number is expected to rise to 1.56 bill-

ion by 2025 [21].

Cardiovascular diseases (CVDs) are a collection of medical conditions that impact the heart and blood vessels [22]. They are responsible for a significant proportion of global deaths, accounting for around one-third of all fatalities each year. Cardiovascular diseases (CVDs) are the leading cause of death globally, responsible for 32% of all deaths in 2019 [23]. These metabolic diseases may lead to severe conditions such as Metabolic syndrome (MetS).

Metabolic syndrome is a cluster of conditions that increase the risk of developing cardiovascular diseases, type 2 diabetes, and other health problems. A diagnosis is typically made when a person has at least three of the following conditions: elevated fasting glucose, elevated blood pressure, elevated triglycerides, reduced HDL cholesterol, and abdominal obesity [24]. The global prevalence of metabolic syndrome is estimated to be around 25%, with higher rates in urban areas and among older adults [25].

Insulin resistance plays a central role in the development of metabolic diseases/syndrome (FIG. 1.2). Insulin is a hormone produced by the pancreas that helps to regulate the amount of glucose (sugar) in the blood. Insulin resistance occurs when the body's cells become resistant to the effects of insulin, leading to high levels of glucose in the blood [26].

When insulin resistance develops, the pancreas responds by producing more insulin to overcome the resistance. This leads to higher-than-normal levels of insulin in the blood, a condition known as hyperinsulinemia. Over time, the pancreas may become exhausted and unable to produce enough insulin, leading to high blood sugar levels and eventually Pre-diabetes leading to diabetes. Insulin resistance also affects the body's ability to use and store energy from food [17]. This can lead to the accumulation of fat in the liver and other organs, which can contribute to the development of non-alcoholic fatty liver disease. Additionally, insulin resistance can increase inflammation in the body, which can contribute to the development of other health problems such as cardiovascular disease [27].

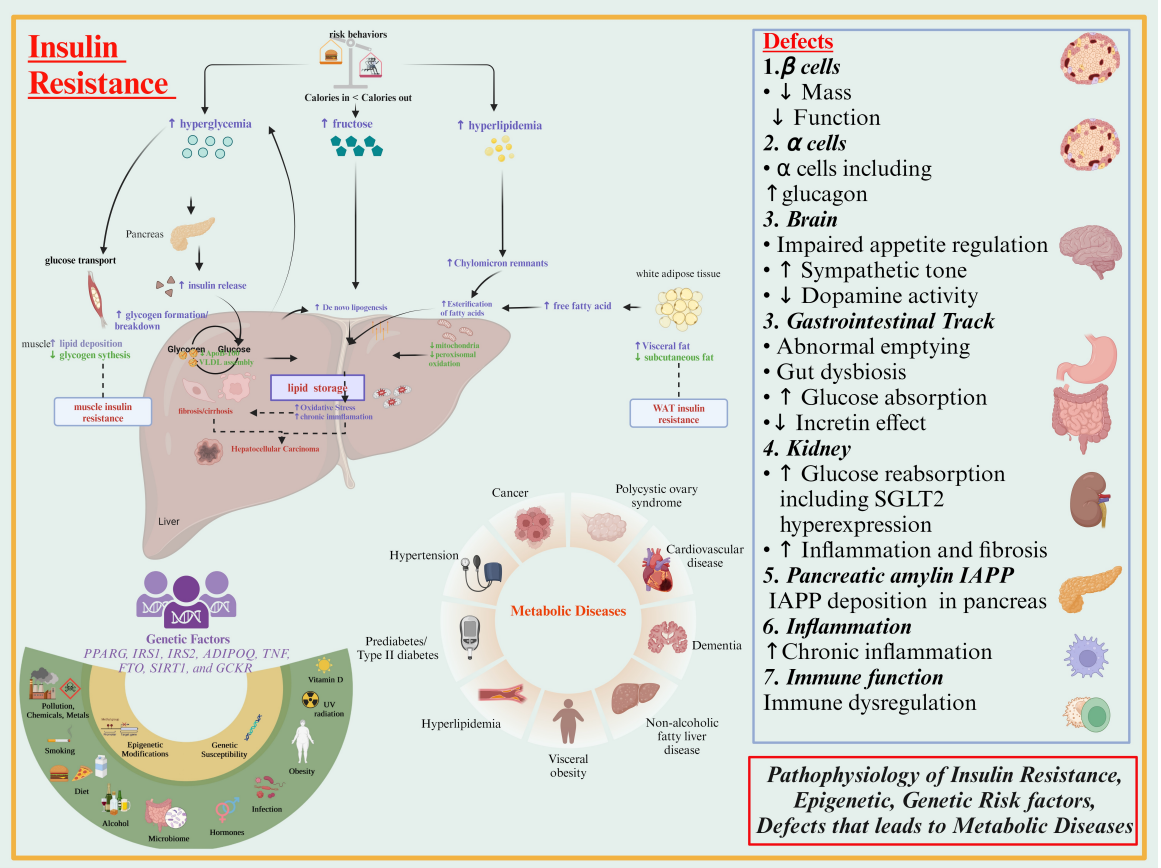


FIGURE 1.2: Role of insulin resistance in the prognosis of metabolic diseases [9, 10]

Insulin resistance and metabolic syndrome are influenced by a complex interplay of genetic and environmental factors. Various genes and genetic variations have been linked to these conditions, including PPARG, IRS1, IRS2, ADIPOQ, TNF, FTO, SIRT1, and GCKR [28]–[29]. These genes are involved in various processes related to energy metabolism, adipocyte differentiation, lipid metabolism, and inflammation, among others. However, the genetic basis of insulin resistance and metabolic syndrome is not fully understood, and additional research is needed to unravel their complex genetic and environmental components.

Omics refers to the comprehensive study of biological molecules or datasets within specific scientific domains. It encompasses the analysis of genes (genomics), proteins (proteomics), metabolites (metabolomics), and other molecular components. In the context of prediabetes, omics methodologies are instrumental in unraveling the molecular intricacies associated with the onset and progression of this condition [30].

Genomics, which explores the structure and function of genes, aids in identifying genetic variations linked to an elevated risk of developing diabetes. By scrutinizing genetic data on a large scale, researchers can pinpoint specific genes or genetic pathways influencing insulin resistance, glucose metabolism, and other prediabetes-related factors [31].

Transcriptomics, focusing on all RNA molecules including mRNA, provides insights into gene expression patterns relevant to prediabetes. This analysis unveils changes in gene activity associated with insulin sensitivity, inflammation, and other pivotal processes, offering potential therapeutic targets or early detection biomarkers [32].

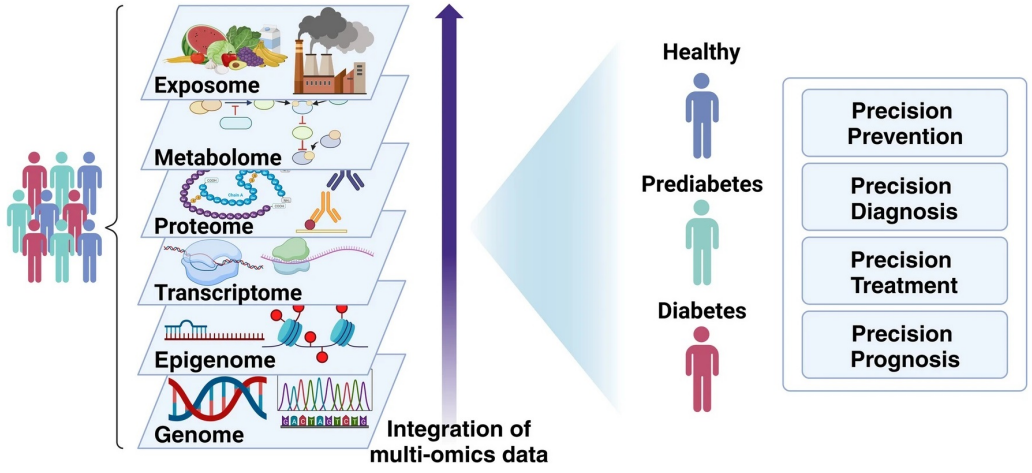


FIGURE 1.3: Integration of multi-omics-data in prediabetic insulin resistance [30]–[33].

Proteomics delves into the study of proteins and their roles within cells. In prediabetes research, proteomic analysis uncovers alterations in protein levels or modifications contributing to insulin resistance, β-cell dysfunction, and metabolic irregularities. Understanding these protein-related mechanisms can guide the development of targeted therapies or diagnostic tools [30].

Metabolomics investigates small molecules (metabolites) in cells, tissues, or biofluids. In prediabetes, metabolomic studies identify metabolic signatures linked to insulin resistance, altered glucose metabolism, and metabolic dysregulation. These metabolic

profiles aid in early detection, risk assessment, and personalized treatment strategies [30]. Integrative omics combines data from multiple omics layers to gain a holistic understanding of biological processes (FIG. 1.3). In prediabetes research, this approach unveils complex interactions between genes, proteins, metabolites, and environmental factors contributing to disease progression. It leads to targeted interventions and precision medicine strategies for managing prediabetes and preventing its transition to diabetes [30]–[33].

1.2 Gap Analysis

The current understanding of prediabetic insulin resistance lacks a comprehensive elucidation of its molecular mechanisms, limiting our ability to conduct a systematic evaluation of health outcomes associated with prediabetes. Furthermore, the absence of formal and validated clinical guidelines for the diagnosis and treatment of prediabetes underscores the need for an in-depth exploration of the genetic and molecular mechanisms underlying prediabetic insulin resistance. This research gap necessitates focused investigation into the genetic and molecular levels to bridge the knowledge gap and facilitate the development of evidence-based guidelines for the early diagnosis and effective treatment of prediabetes.

1.3 Research Objectives

The primary aim of this research is to gain a comprehensive understanding of the molecular underpinnings of prediabetic insulin resistance. By integrating data from various studies and employing advanced computational tools, this study seeks to uncover the critical genetic and metabolic factors that contribute to the early stages of insulin resistance. The specific objectives of the research are outlined below:

Research Objective 1: To conduct a meta-analysis to elucidate the genetic and molecular determinants associated with prediabetic insulin resistance.

Research Objective 2: To identify and validate key genetic and metabolic components implicated in prediabetic insulin resistance through pathway analysis and gene interaction modeling.

Research Objective 3: To computationally validate the functional roles of identified genetic and metabolic components in prediabetic insulin resistance using bioinformatics approaches.

1.4 Significance of Study

Pre-diabetic insulin resistance could result in the progression of severe diseases including diabetes and metabolic syndrome, organ failures, cardiovascular diseases, and even cancer. Therefore, it is necessary to explore mechanisms causing prediabetes onset and progression at the molecular and cellular levels. Understanding these pathways will play a pivotal role in diagnostic and management strategies against Pre-diabetic insulin resistance.

1.5 Problem Statement

Diabetes, metabolic syndrome, and cardiovascular disease once prognoses are difficult to treat and manage. It is necessary to prevent the progression of these diseases by early diagnosis and treatment at the prediabetic stage. Environmental and genetic factors are involved in insulin resistance which is the main cause of prediabetes but what key genetic factors are involved in prediabetic insulin resistance is still not clear. Identifying molecular mechanisms can help detect prediabetes and reduce its burden.

1.6 Research Questions

Elaboration and understanding of genetic and molecular pathways associated with prediabetic insulin resistance.

1. What are the primary genetic variants associated with prediabetic insulin resistance, and how do they influence molecular pathways?
2. What molecular mechanisms underlie the development of insulin resistance in prediabetes, particularly focusing on skeletal muscle, liver, and adipose tissue interactions?
3. How do environmental factors interact with genetic predispositions to contribute to prediabetic insulin resistance?

Identification and validation of key genetic and metabolic components involved in prediabetic insulin resistance based on pathway and gene interaction.

1. Which genetic pathways and metabolic processes play pivotal roles in the development and progression of insulin resistance in prediabetes?
2. Can machine learning or network analysis techniques be applied to identify key genes and metabolic components driving prediabetic insulin resistance?
3. How do gene-gene and gene-environment interactions influence the manifestation of insulin resistance in prediabetes?

Validation of genetic and metabolic components of prediabetic insulin resistance using Bioinformatic approach.

1. Which genetic and metabolic components are critical for prediabetic insulin resistance based on bioinformatic analysis?
2. How do interaction network and pathway analysis reveal key molecular players in prediabetic insulin resistance?
3. Can bioinformatic docking studies validate interactions of proteins involved in prediabetic insulin resistance with high accuracy?

Chapter 2

Literature Review

2.1 Insulin Resistance

Insulin resistance (IR) is defined as a diminished capacity of insulin to promote glucose uptake in the body's cells. It can also be described as the body's decreased responsiveness to insulin, which negatively impacts growth, development, and metabolism.

In IR, insulin–responsive tissues, particularly muscle tissues, fail to absorb glucose efficiently from the bloodstream, resulting in elevated blood glucose levels, or hyperglycemia. This hyperglycemia is a key factor contributing to various adverse macrovascular and microvascular outcomes [3].

When insulin resistance occurs alongside dysfunction of pancreatic beta cells and a relative deficiency of insulin, it can lead to prediabetes and, eventually, type 2 diabetes (T2D). This progression highlights the interplay between IR and insulin secretion in the development of metabolic diseases [34].

In general terms, insulin resistance is recognized as a condition where the normal or elevated levels of insulin do not elicit a strong biological response. This phenomenon is most commonly associated with the decreased sensitivity of insulin–mediated glucose uptake, where insulin's ability to facilitate glucose entry into cells is significantly impaired [35].

2.2 Pathophysiology of Insulin Resistance

The underlying mechanisms of insulin resistance (IR) are intricate and not fully understood. Every organ system plays a role in IR, with different systems taking the lead depending on the body's state. After eating or doing exercise, the skeletal muscles are the heaviest promoter of insulin sensitivity. On the other hand, during periods of fasting the liver comes to the fore in action, with such tasks as glycolysis and gluconeogenesis. Adipose tissue, which has a relatively minor overall effect on insulin sensitivity but serves as a crucial endocrine and inflammatory regulator for other organs. This results in insulin resistance by inflammatory cells, including macrophages; T cells and B cells within fat tissue. Furthermore, the role of the central nervous system in this process is attracting more and more attention. It is important to recognize that insulin sensitivity is not a black-and-white condition but rather exists along a spectrum.

While some studies attempt to classify individuals as "insulin resistant" or "insulin sensitive" using specific cut-off values, this method has its limitations. Firstly, there's no universally accepted way to measure IR—most large studies rely on surrogate markers from fasting blood samples, which vary in how accurately they reflect liver or muscle insulin resistance. Secondly, what is considered clinically significant insulin resistance (IR) can vary across different ethnic groups, influenced by factors such as compensatory hyperinsulinemia and beta cell dysfunction. Thirdly, despite extensive research efforts, including studies like the Framingham Heart Study and the Insulin Resistance Atherosclerosis Study, the full impact of prediabetic IR remains unclear, and many individuals still go undiagnosed [36].

2.3 Molecular Mechanism of Insulin Resistance

Muscles account up to 60 to 70% of the entire body insulin–dependent uptake of glucose. This uptake is done by GLUT4 transporters [37]. If muscular glycogen synthesis is reduced it is called insulin resistance (FIG. 2.1). The resistance to insulin

tends to be reduced with the translocation of intracellular glucose [35]. As per previous studies, no difference has been reported between insulin–resistant type II diabetics and control. Although this was due to hyperinsulinemia in the hyperinsulinemic clamp study [38]

Pathophysiology of IR (Muscles)

- Insulin→IR→GLUT4→Glycolysis or Glycogen Synthesis
- IR → Defected Insulin Receptors, Defected Insulin Signaling or Defected GLUT4 proteins

Insulin Mediated Pathway
Insulin → Tyrosine Kinase Receptor → Phosphorylation → ISR1 → PI3K → PIP2 →PIP3 → Protein Kinase B or AKT → TBC1D1 → RabGDP → RabGTP → Release GLUT4 to surface membrane → Glucose Uptake

Contraction Mediated Pathway (Exercise)
ADK (Adenylate kinase) → 2 ADP → 1 ATP+1 AMP → ATP utilized for energy+AMP →AMPK →TBC1D1 → RabGDP → RabGTP → Release GLUT4 to surface membarne → Glucose Uptake.

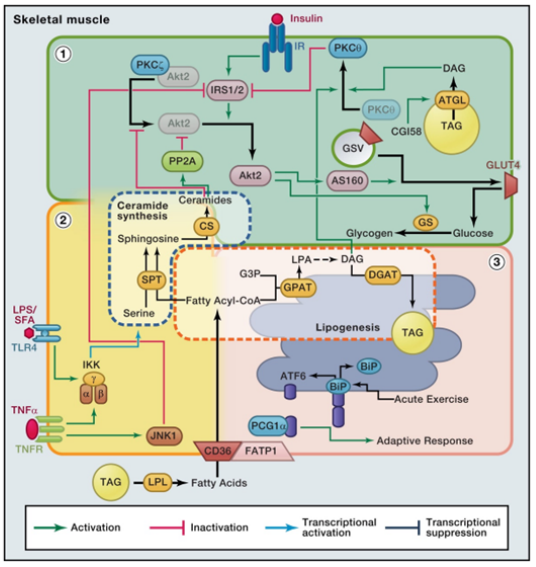


FIGURE 2.1: Insulin resistance mechanism pathway in skeletal muscle cell [38]

Liver's response to insulin resistance and its resulting effects on metabolic processes is illustrated in figure 2.2. Glucose enters liver cells through the GLUT2 transporter, initiating pathways for glycogen storage or glycolysis. In glycolysis, glucose is broken down into pyruvate, which then produces acetyl-CoA in the mitochondria. Acetyl-CoA contributes to ATP generation and free fatty acid (FFA) synthesis, with FFAs ultimately forming triacylglycerol (TAG) in the liver. TAGs may enter the blood-stream, affecting overall lipid metabolism [35].

The insulin signaling pathway is central to this process. When insulin binds to its receptor, a phosphorylation cascade is triggered, starting with activation of receptor substrate proteins (like IRS1) and progressing to PI3K and AKT, which facilitate GLUT4 translocation to the cell surface for glucose uptake. In cases of insulin re-sistance, this signaling pathway is disrupted, reducing glucose uptake and glycogen production in the liver. Insulin resistance also alters lipid metabolism, promoting ceramide and sphingolipid synthesis—molecules involved in cellular stress responses.

Excessive FFAs are converted to TAGs, which can accumulate in the liver and contribute to fatty liver disease [35].

Additionally, insulin resistance triggers inflammatory responses. Cytokines such as IL6 and TNFα activate kinases like JNK1, which interfere with insulin signaling and intensify cellular stress. This leads to a cascade that impairs glucose metabolism further. The diagram (figure 2.2.) also includes the endoplasmic reticulum (ER) stress response, featuring proteins like XBP1 and PERK, which contribute to metabolic and inflammatory stress. As insulin resistance progresses, the liver's decreased ability to take up glucose and synthesize glycogen leads to elevated blood glucose levels, which can eventually result in diabetes. This diagram provides an overview of the intricate network of interactions where insulin resistance disrupts glucose and lipid metabolism, leading to broader metabolic health issues [35].

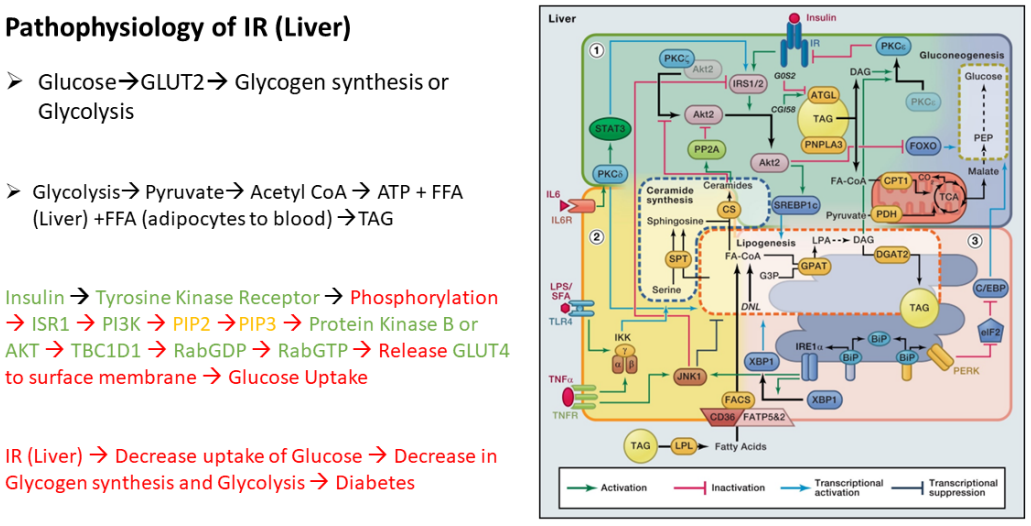


FIGURE 2.2: Insulin resistance mechanism in hepatic cell [38]

In resistance of insulin, the effects are similar on adipose tissues (FIG. 2.3), whereas flux of free fatty acids promotes hepatic VLDL production (very low–density lipoproteins [38] while ketogenesis usually remains repressed by the compensatory hyperinsulinemia. As lipoprotein lipase activity depends on insulin and dysfunctional by insulin resistance, the peripheral uptake of triglycerides from very low–density lipoproteins are also decreased. These biochemical mechanisms account for observed hyper triglyceridemic insulin (FIG. 2.4) [39].

Pathophysiology of IR (Adipocytes AT)

- AT is formed at specific times and locations.
- Once formed, the tissue retains dynamicity, responding to homeostatic and external signals and being capable of a 15-fold expansion

Insulin →INSR-A/B→GLUT4→Glucose uptake→Glycolysis→Acetyl CoA (Citric acid cycle)→ Fatty acid→ Adipocytes→**Lipogenesis**

Insulin→Lipoproteinlipase **LPL**→Blood →Chlymicron+VLDL→Fatty acids→ Adipocytes →**Lipogenesis**

IR or Low Insulin in blood--------->**Lipolysis**

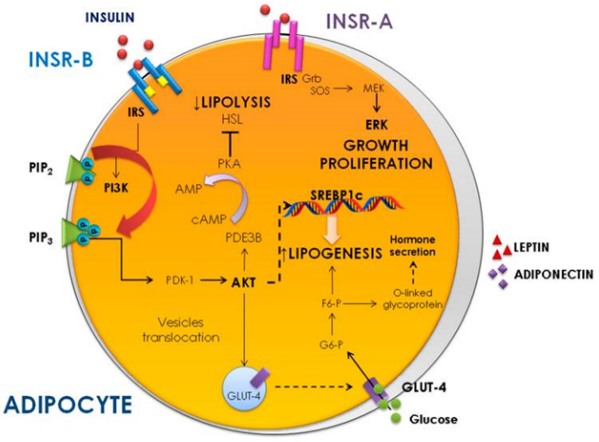


FIGURE 2.3: Insulin resistance mechanism in adipocytes showing lipolysis [40]

Pathophysiology of IR (Adipocytes AT)

Normal
Lipogenensis and Lipolysis--------> Homeostasis

Hyperinsulinemia (Insulin Receptor are Functional)
Glucose uptake Increases ----------> Lypogenesis enhanced----->FFA, TGA increased in blood---> Obesity

Hyperinsulinemia (Insulin Receptor are not Functional)
Glucose in blood (High)------->Lipogenesis decreased Lypolysis enhaced (weight decrease, diabetic case)

Hypoinsulinemia (or insulin insensitivity)
Lypolysis enhanced----->Lypolysis enhanced (diabetes 1)

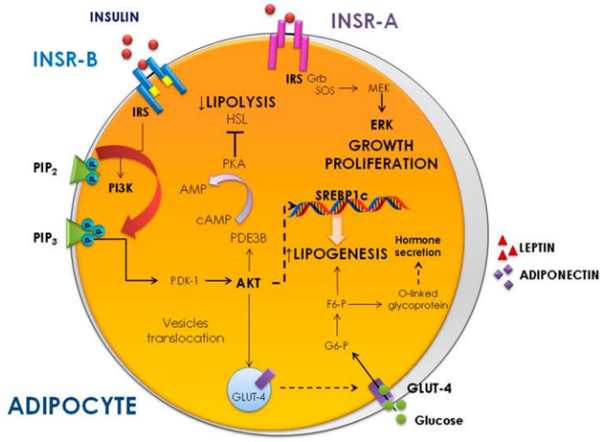


FIGURE 2.4: Insulin resistance in adipocyte (diabetic conditions) [40]

Insulin resistance (IR) is considered a precursor to type II diabetes mellitus (T2DM) and various cardiovascular diseases [34], [40]–[41]. Insulin resistance and compensatory hyperinsulinemia commonly found in many different conditions which includes obesity as well. A syndrome refers to a collection of related abnormalities, as seen with insulin resistance, which often leads to physical complications commonly found in individuals with insulin resistance. Due to differences in tissue response to insulin dependence and sensitivity, insulin resistance syndrome is likely a result of both insulin-dependent effects and varying degrees of resistance to insulin's actions. Metabolic

syndrome, in particular, serves as a clinical marker for identifying individuals at high risk for insulin-resistant cardiovascular diseases [34],[42].

The relationship between obesity, insulin resistance, and type II diabetes has long been valued. a study has established that IMCL (intramyocellular lipid) content is a better forecaster of muscle insulin resistance than fat mass in hale and hearty, young, inactive and lean individuals [43] and nonobese, nondiabetic but insulin–resistant adults [44] and children [45], signifying a potential contributing role for IMCL mediator(s) in this process.

Almost half century ago, Randle and colleagues in their study suggested that insulin resistance of muscles linked with obesity could be accredited to high fatty acid oxidation that bounds insulin–stimulated glucose utilization [45]. Higher delivery and oxidation of fat molecules leads to accumulation of phosphofructokinase inhibitors which are citrates is basic concepts of this model. Phosphofructokinase is very important enzyme which lead to increase in intra myocellular glucose 6 phosphate and glucose. These ultimately impairs glucose uptake and utilization [45].

The activation of PKCε by DAG in the liver disrupts insulin signaling, which hinders the liver's ability to form glycogen in response to insulin. Despite this, the liver continues to synthesize lipids without any reduction. In skeletal muscle, the activation of PKCθ by DAG interferes with insulin signaling as well, preventing glucose from being effectively taken up by the muscles and resulting in more glucose being directed to the liver. However, physical exercise can still enhance glucose uptake by the muscles.

In adipose tissue, the release of cytokines from adipose tissue macrophages (ATMs) stimulates the breakdown of fats, leading to an increased release of fatty acids. These fatty acids further boost lipid production in the liver and stimulate gluconeogenesis through acetyl-CoA-mediated activation of pyruvate carboxylase (PC) and glycerol, which pushes glucose production by providing additional substrates.

This explanation touches on key components like the insulin receptor (IR), glycogen phosphorylase (GP), glycogen synthase (GS), lysophosphatidic acid (LPA), hepatic lipase (HL), chylomicron remnants (CM-R), very low-density lipoprotein (VLDL),

calmodulin kinase II (CAM-KII), insulin receptor substrate 1/2 (IRS1/2), β-adrenergic receptor (βAR), and chylomicron triglycerides (CM-TG). This has been illustrated in figure 2.5 [46].

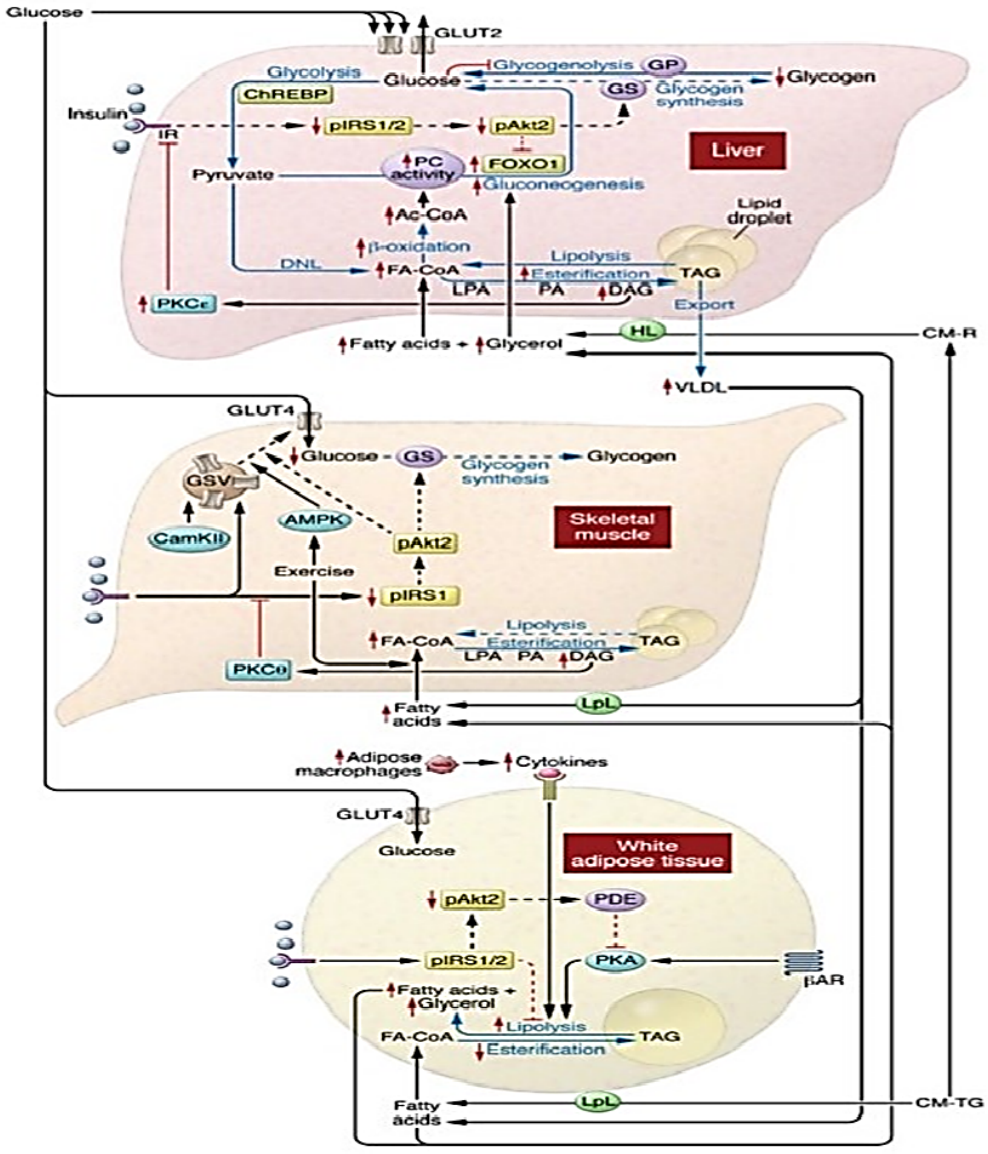


FIGURE 2.5: Overall mechanism of insulin resistance [46]

Family studies suggest that insulin resistance (IR) is about 40%–50% heritable, though the estimates can vary widely. It's thought that up to 44% of this heritability might be due to genetic polymorphisms, many of which remain unidentified and are present in at least 5% of the population. Most of the genetic variants known to be associated with type 2 diabetes (T2D) primarily affect insulin secretion, with only a few influencing

IR directly. Identifying the genetic factors behind IR has been challenging for several reasons [47].

A major limitation is the lack of large scale studies incorporating adequate physiological recording of insulin sensitivity. Among adult participants without diabetes, the highest number of individuals enrolled in a genome-wide association study (GWAS) employed HOMA-IR as the coefficient of fasting plasma glucose. This is undertaken in the absence of HOMA-IR's methodology which does not consider the degree of insulin resistant that occurs in certain tissues or after a meal which may have varying genetic causes. This drawback is further supported by two other studies which observed a very small genetic correlate (-0.53an, -0.57) of HOMA-IR with two accurate methods including euglycemic clamp and minimal model [48].

In the largest genome-wide association studies (GWAS) of physiological measures of insulin resistance (IR) conducted so far, which was carried out by the GENEticS of Insulin Sensitivity, there were no genome-wide associations with inclusion of 5624 participants. Among the findings, a variant in the NAT2 gene (rs1208A – G) was noted in IR studies, and further studies in mice supported the role of NAT2 in insulin response. These observations underscore genetic studies as a useful approach for gaining insight into important clinical aspects of IR, whilst calling for larger studies to demonstrate the meaningful signals. [49].

Further evidence of a contribution to IR having a genetic basis was presented by Dimas whose group delineated five genetic clusters for T2D, one of which was within a cluster containing IR- associated loci such as PPARG, KLF14, IRS1, GCKR, and others. Scott and coworkers developed a 'genetic risk profile' of ten SNPs and claimed to use it in assessing the risk of patients to develop IR with special reference to using clinical insulin resistance by selecting SNPs associated with fasting insulin levels, hypertriglyceridemia, and low levels of HDLcholesterol, which are indicators of IR.

This was validated against clamp measures in a cohort of 1899 non-diabetic individuals and has undergone significant development and been used as an intervention in studying genetic Inflammatory Resemblance. And in these studies regarding genetic

factors researchers demonstrated, to the surprise of many, that genetic IR is not related to the degree of obesity, which is also associated with elevated cardiometabolic risk. Also, they have shown that genetic IR correlates inversely to obesity measures such as hip circumference and body fat %age. Increased cardiometabolic risks in patients with genetic IR, it has been speculated, could be due to an inadequate capacity of 'disease' or peripheral fat limiting excessive energy energetics, high ectopic vesicle distribution within the bodies of the patients. This 'adipose tissue expandability' hypothesis explains intestial fat accumulation with higher cardiovascular risk in some rare monogenic forms of lipodystrophy/IR, and some lean people with insulin resistance [50].

Genes that are associated with IR in diabetes type 2 are HNF4A (MODY1), GCK (MODY2), HNF1A(MODY3), PKM, FASN, ACACA, PEPCK, INSR, IRS1, IRS2, GUT4, IGF1, GCKR, SIRT1, GAPDH, PKLR, FASN, SSTR2, FOS, GCK, APOA1, PCK2, G6PC, APOC3, and GLP1R [51]. Few genes that are associated the Diabetes type 2 reported in literature along with their role is summarized in table. 2.1

TABLE 2.1: List of genes associated with insulin resistance in diabetes mellitus type 2.

Gene region	**Function**	**References**
TCF7L2	Transcription factor trans activates proglucagon and insulin genes	[52]
PPARG	Transcription factor involved in adipocyte development	[53]
KCNJ11	Kir6.2 K+ channel risk allele impairs insulin secretion	[49]
WFS1	Endoplasmic reticulum transmembrane protein	[54]

HNF1B	Transcription factor involved in pancreatic development	[55]
SLC30A8	β-Cell zinc transporter ZnT8 insulin storage and secretion	[56]
HHEX	Transcription factor involved in pancreatic development	[57]
CDKAL1	Homologous to CDK5RAP1, CDK5 inhibitor islet glucotoxicity sensor	[58]
IGF2BP2	Growth factor binding protein pancreatic development	[59]
CDKN2A/B	Cyclin-dependent kinase inhibitor and p15 tumor suppressor islet development	[60]
FTO	Alters BMI in general population	[61]
JAZF1	Transcriptional repressor associated with prostate cancer	[62]
CDC123-CAMK1D	Cell cycle/ protein kinase	[8]
TSPAN8-	Cell surface	[36]

LGR5	glycoprotein implicated in gastrointestina cancers	
THADA	Thyroid adenoma associates with PPARG	[63]
ADAMTS9	Secreted metalloprotease expressed in muscle and pancreas	[64]
NOTCH2	Transmembrane receptor implicated in pancreatic organogenesis	[47]
KCNQ1	Pore-forming subunit of voltage-gated K+ channel (KvLQT1) risk allele impairs insulin secretion	[65]

The molecular mechanisms behind insulin resistance (IR) involve the failure of insulin to effectively bind to its receptors, which in turn impairs the activation of PI 3-kinase (PI 3-K) through the insulin receptor substrate (IRS). Normally, the downstream signaling of PI 3-K leads to the activation of protein kinase B (PKB), which plays a critical role in regulating the transcription of target genes via pathways like GSK-3 and Foxo1. Additionally, PKB is essential for promoting glucose uptake by moving GLUT-4 to the cell's plasma membrane.

However, in insulin resistance, this activation and regulation process is disrupted. Beyond this, the insulin signaling pathway branches downstream of the insulin receptor, with PKCς/λ acting as additional effectors. Variations in PKCθ activity, influenced by fatty acids, may interfere with the IRS/PI 3-K pathway and hinder GLUT-4-dependent glucose transport. To clarify, here are the key components: IR

stands for the insulin receptor; IRS refers to the insulin receptor substrate; PDK is phosphoinositide-dependent kinase; PI 3-K is phosphatidylinositol 3-kinase; PKB is protein kinase B; PKC refers to protein kinase C; GSK-3 is glycogen synthase kinase-3; GLUT-4 is glucose transporter 4; Foxo represents forkhead box protein, and FKHR is forkhead in rhabdomyosarcoma (figure 2.6) [66].

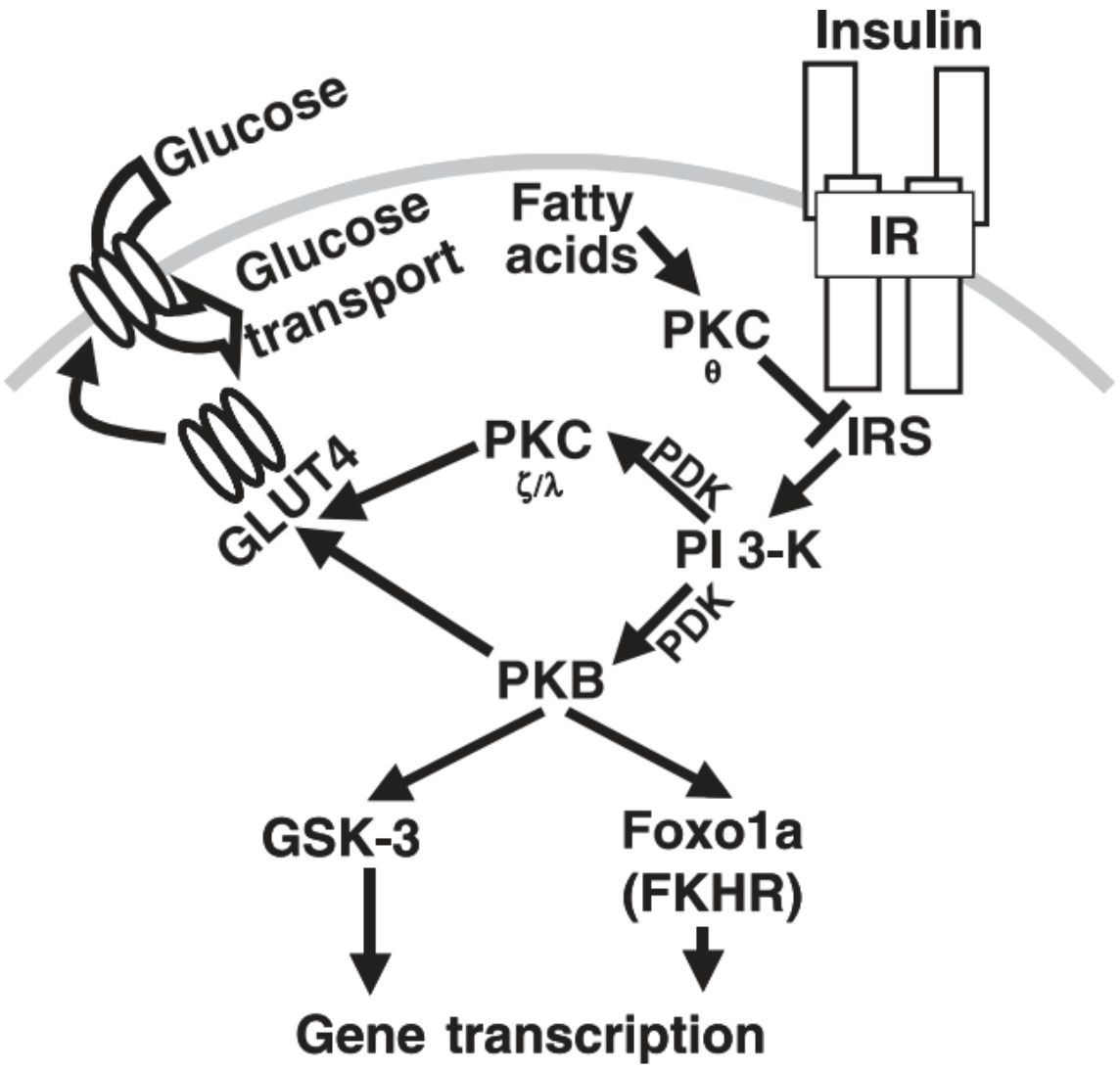


FIGURE 2.6: Schematic representation of insulin resistance [66]

2.4 Measurement of Insulin Resistance

Assessing insulin resistance (IR) is a complex task. Glucose metabolism occurs rapidly and is influenced by factors such as body composition and dietary habits. Moreover,

the physiological contributors to IR vary depending on whether a person is in a fasted or fed state, and whether they are at rest or have exercised. In the fasted state, the liver plays a central role in determining IR, whereas, after eating, skeletal muscle becomes the primary organ responsible for glucose disposal. Although IR can differ significantly between individuals, it remains relatively stable within the same person over time [67].

The gold standard for measuring IR involves assessing how well glucose is taken up in response to insulin. Techniques for this include using radioactively labeled glucose tracers, insulin suppression tests, and clamp studies. Studies II and III specifically utilize the hyperinsulinemic-euglycemic clamp method, where hepatic gluconeogenesis is suppressed by continuously infusing insulin. A feedback loop between a blood glucose analyzer and a glucose infusion helps maintain a steady euglycemic state. The M/I index, which indicates how much glucose is removed from the blood per unit of insulin infused, is used to quantify insulin sensitivity. While clamp studies offer a detailed reflection of physiological processes, they are labor-intensive, requiring several hours of medical monitoring, and are thus impractical for large-scale studies due to the associated personnel, time, and participant fitness requirements [68].

To evaluate insulin resistance, many epidemiological studies usually cite the Homeostasis Model Assessment of Insulin Resistance (HOMA-IR). HOMA-IR operates on the premise of a feedback loop that occurs in fasting conditions, particularly to the interactivity between the amount of glucose released from the liver and the amount of insulin secreted from the pancreas. HOMA-IR is based on routine lab tests and correlates positively but moderately with the more sophisticated but potentially more accurate measures of physiology (the correlation coefficient is approximately from 0.6 to 0.9) and does not truly reflect the physiological status of the individuals. The general HOMA-IR method in the population, however, still has certain predictive power in relation to future cardiovascular disease (CVD) and diabetic complications patterns at the population level. It has been applied in different types of studies such as glycaemia treatment objectives in the UK Prospective Diabetes Study, coronary artery disease (CAD) epidemiology, Mediterranean diet advantages, and genetic

studies related to insulin resistance [69],[33].

The approach selected for the assessment of insulin resistance (IR) depends on the objectives of the study, the size of the sample, and the resources available. For instance, it can be decided here that HOMA-IR is used to determine only hepatic IR whereas the metabolic clamp is reserved for measuring peripheral IR. Whenever fasting periods need to be evaluated, HOMA-IR is the favorite tool whilst the Matsuda index is preferred during stimulated periods. Dynamic measurements may be obtained through minimal model whereas steady-state measurements are predominantly collected using the clamp technique. Last but not least, HOMA-IR is appropriate for aggregate statistics about the entire population, while the clamp approach is more appropriate for single patients [62].

2.5 Prediabetes

Prediabetes is defined as a blood glucose level higher than normal, but lower than the diabetes threshold, Normal blood glucose levels include a fasting plasma glucose (FPG) level below 100 mg/dL (5.6 mmol/L), a 2-hour oral glucose tolerance test (OGTT) result below 140 mg/dL (7.8 mmol/L), and an A1C level below 5.7%. Prediabetes is diagnosed when FPG ranges from 100–125 mg/dL (5.6–6.9 mmol/L), OGTT results fall between 140–199 mg/dL (7.8–11.0 mmol/L), or A1C levels range from 5.7%–6.4%. Diabetes is identified when FPG is 126 mg/dL (7.0 mmol/L) or higher, OGTT results reach 200 mg/dL (11.1 mmol/L) or more, or A1C is 6.5% or above.

Prediabetes is a high–risk state for diabetes development with a yearly 5% to 10%, therefore prediabetes can be defined as an intermediate state between a normal health condition and an diabetic diseased condition. Prediabetic condition generally include presence of Impaired Fasting Glucose (1FG) or Impaired Glucose Tolerance (IGT). It may involve both IFT and IGT in coexistence. The American Diabetic Association (ADA) also includes the HbA1c for diagnosis of prediabetes. According to ADA

guidelines, normal HbA1c levels are below 5.7%, while prediabetes is diagnosed when HbA1c ranges from 5.7% to 6.4%. An HbA1c level of 6.5% or higher indicates diabetes. This test reflects average blood glucose levels over the past two to three months and is widely used for early detection and monitoring of diabetes risk. [70].

In addition to prediabetes, several terms have been suggested such as IFG, IGT, Intermediate hyperglycemia, and high–risk state of developing diabetes by the World Health Organization (WHO) and ADA. Due to confusing and interlinked definitions of prediabetes and diabetes, the diagnosis and treatment are difficult in clinical studies [71], [72].

The term prediabetes denotes to lessened fasting glucose and/or impaired glucose tolerance in subjects who are at higher risk for T2DM (type II diabetes mellitus). Though both types of patients have an increased risk for evolving type 2 diabetes mellitus and cardiovascular diseases, they evident distinguished metabolic irregularities [73].

Glycemic variables that are higher than that of normal values but lower than threshold value is increased risk for onset of diabtese and this is termed as prediabetes (intermediate hyperglycemia) The ratio of prediabetic to convert to diabetics is 5–10% per year. This is same as changing back to normoglycemia. Existence of prediabetes is increasing globally and as per specialists it is projected to be more than 470 million people by end of 2030 [74].

2.6 Epidemiology and Diagnosis of Prediabetes

The incidence and prevalence of prediabetes has been increased in both developed and developing countries. According to International Diabetic Federation (IDF) prevalence of prediabetes worldwide is 280 million in 2011 with projection of 398 million by 2030. In United States more than 38% of adults are suffering from prediabetes. In China, the prediabetes among adults has been reached up to 50%.

In Pakistan according to 2nd National Diabetes Survey of Pakistan 2016–2017 prevalence of prediabetes is 14.4%. It is observed that rapid urbanization and changes in

diet in developing countries have raises the Body Mass Index (BMI) and prediabetes as well. The genetic difference among different population also plays important role in progression of prediabetes to diabetes [75].

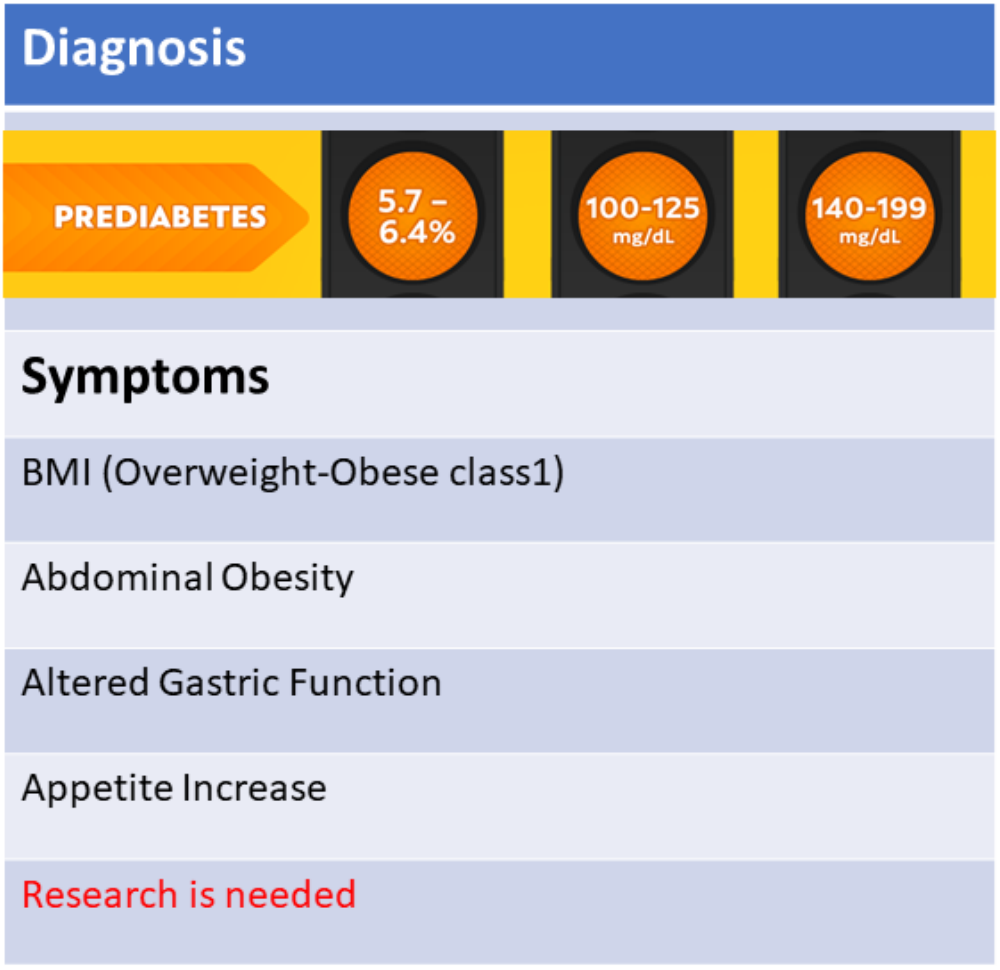


FIGURE 2.7: General diagnostic thresholds for prediabetes

To identify diabetes risk group or prediabetic state individual risk factors such as history of gestational diabetes, or first–degree relative risk or combination of risk factors such as obesity, CVD or metabolic syndrome can be used although their predictive value is not significant enough. In addition to these risk score for incident diabetes is in development phase to categorize individuals that are at high risk of diabetes. These are based on both non–invasive or blood–based risk factors [76]. The diagnosis of prediabetes is based on specific biochemical markers, including an HbA1c level between 5.7% and 6.4%, fasting blood glucose levels ranging from 100 to 125 mg/dL, and postprandial blood glucose levels (measured two hours after a meal) between 140

and 199 mg/dL. Individuals with prediabetes commonly exhibit symptoms such as a body mass index (BMI) in the overweight to obese (Class 1) category, abdominal obesity, altered gastric function, and an increased appetite.

Despite these established diagnostic criteria and associated symptoms, further research is needed to deepen understanding of the mechanisms and long-term impacts of prediabetes, as well as to identify effective intervention strategies (FIG. 2.7).

2.7 Prediabetes to Diabetes Progression

It has been reported that 5 to 10% of population having prediabetes progresses to diabetes annually. Incidence of diabetes for isolated IGT ranges from 4 to 6 percent and isolated IFG ranges from 6 to 9 percent annually as reported in various studies conducted up to 2004. The combined rate of IFG and IGT has been reported up to 15–19 percent.

Recent studies have revealed that the incidence of progression of diabetes from prediabetes is 11 percent annually reported by Diabetes Prevention Program (DPP). US–Multi Ethnic studies reported 6 percent among IFG subjects and in Japanese Population 9 percent IFG and 7 percent A1C incidence in reported among participants. Risk of developing diabetes is broadly close to that raised by A1C on the basis of fasting plasma glucose and 2 hours post load glucose [75].

Among prediabetic patient's 70% individuals will ultimately develop diabetes according to ADA expert group. Chinese diabetes prevention trail reported that the women with gestational diabetes have 20 to 60% chance of developing diabetes 5 to 10 years after pregnancy.

According to 20 meta–analysis studies 13 percent mothers with gestational diabetes developed diabetes after pregnancy compared with 1 percent mothers without gestational diabetes [77]. The variability in estimates is due to difference in parameters used in studies to detect gestational diabetes and diabetes.

2.8 Diabetes

A group of metabolic diseases characterized by high glucose levels i.e. hyperglycemia is called diabetes mellitus (DM). Reasons for this condition are defects in insulin secretion or insulin action or may be both. Long term damage of various organ, non–functioning and total failure of vital organs i.e. kidnies, eyes, nerves and heart are associated conditions of hyperglycemia or DM [78].

Diabtese development consists of various pathogenic pathways. These pathways can range from autoimmune obiletration of the cells of insulin producing organ i.e. pancreas which give rise to insulin deficiency in the body. Due to this deficiency various abnormalities becomes resistant to insulin action. Abnormalities arises in many biochemicals such as carbohydrates, proteins and fatty acids' metabolism due to the deficient action of insulin on target tissues.

Insufficient action of insulin is due to inadequate secretion of insulin or no response from tissues to produced insulin. This can happen at one or more different points in complex hormonal action pathways. Primary cause of hyperglycemia is still unknown as abnormality in either insulin secretion or action may coesxist in a patient or only one type also can be a cause of diabetes mellitus [79].

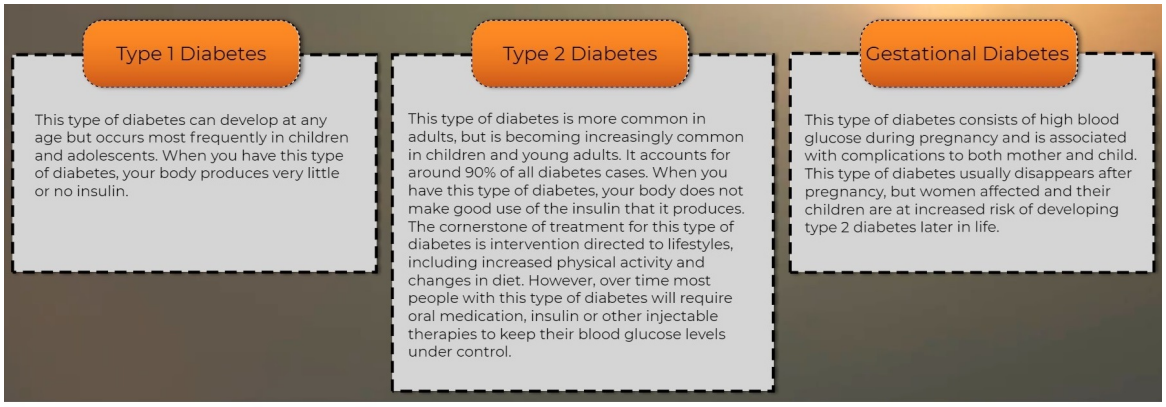


FIGURE 2.8: Types of diabetes [82]

Dibetes mellitus can be divided into different types based on interaction between genetic or environmental factors and is describes as per its etiology. Type 1 Diabetes

Mellitus (T1DM), Type 2 Diabetes Mellitus (T2DM) [80] and Gestational Diabetes Mellitus (GDM) are the most common types of diabetes [81].

There are some other types of diabtes as well such as maturity–onset diabetes of the young (MODY) and late–onset autoimmune diabetes of the adult (LADA), but these are very less common [82]. One of the most common type of DM is T2DM which affects almost 95% of the whole world's population [81].

2.9 Outcomes of Diabetes

Diabetes mellitus (DM) is a very multifaceted metabolic disorder which is characterized by persistent very high glucose levels in the blood [80]. as per the International Diabetes Federation it is one of the biggest global health emergency in twenty first century. [83]. In 2015, worldwide its occurrence was of one in eleven adults (1/11) and the projected occurrence of the impaired glucose tolerance was one in fifteen (1/15) adults. These ratios are expected to increase further, specifically in urban population.

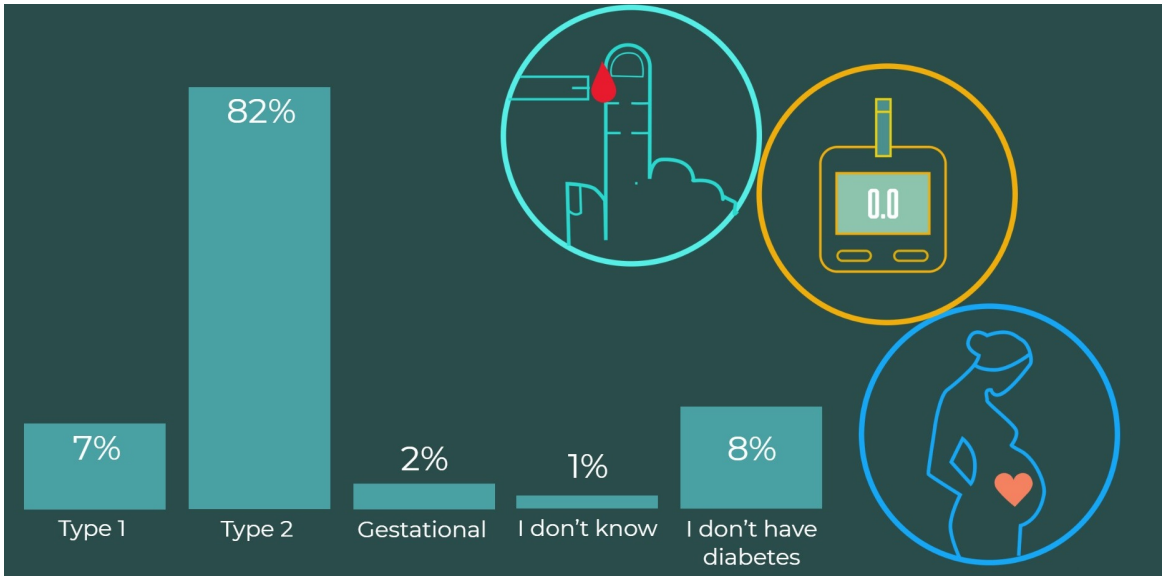


FIGURE 2.9: Public reporting about diabetes [86, 87]

These increasing indices lead to high medical and economic burden. Almost 12% of global health budget is being expended on diabtese worldwide. [83]. Many studies are being conducted on various populations in this regard. A similar study directed

recently in the Romanian population and it showed that diabetes mellitus is one of the main healthcare glitches of their medical system. its occurrence is 11.6% and prediabetes' is 16.5% [84], [85].

It is a foremost international health problem, which is currently affecting almost 246 million people around the world. In the coming 30 years this number is expected to be doubled [88], Majority of the cases are of type 2 diabetes [86], [87] (figure 2.9). In diabetes some structural changes develop in the brain: cerebral atrophy and lacunar infarcts, blood flow changes of both hypo– and hyper perfusion [85]. In diabetic patients, it is found that brain volumes are reduced which restricts to the hippocampus. while it is also observed that there is an inverse association between glycemic control and hippocampal volumes. In some studies and experiments, HbA1C is said to be only and major predictor of hippocampal volumes [85, 89].

Type 1 diabetes is insulin–dependent and people having DM1 need unlike and more complex management of their hyperglycemia as compared to DM2. Frequent monitoring and adjustment of insulin doses, maintenance of diet, and exercise are the prerequisites of Type 1 diabetes management and control. This type of diabetes starts at a very early age [90]. Vascular lesions are also a complication of this disease. These lesions may develop into small and large blood vessels.

This mechanism of forming vascular lesions is multifactorial. Activation of various chemical pathways such as protein kinase C, accumulation of sorbitol depletion of myoinositol is due to derangements in vessels because of high glucose influx. This gives rise to vascular lesions [91], [92].

Microvasular complications also includes coronary heart diseases, peripheral vascular diseases, retinopathy, nephropathy, stroke and erectile dysfunction etc (figure 2.10). there is a continuous relationship between glycemic controls and the incidence and progression of microvascular complications [93]. There are some unmodifiable risk factors for Type II diabetes mellitis that include age, ethnicity, family history (hereditary risk, history of gestational diabetes, birth weight). The risk of the disease is clearly dependent on the demographic factor. In 2005, the CDC reported that 20.6 million

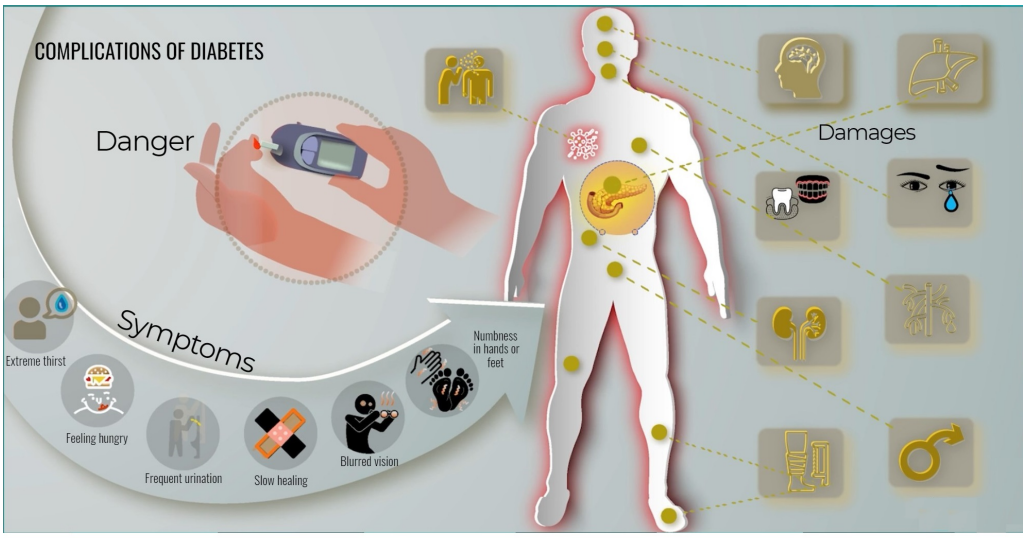


FIGURE 2.10: Complication due to diabetes [93]

residents of the U.S. aged 20 years and older had diabetes, which corresponded to 9.6% of this population. Further, diabetes is reported as a disease whose prevalence increases with increasing age, noting however that out of those when aged sixty years and above approximated 10.3 million or twenty point nine percent of people had the condition [94]. There are many associated risk factors of diabetes and it can disturb many systems in the body with time which is also a cause of serious complications.

These complications (figure 2.11) can be divided into micro and macro vascular, neuropathy (Nervous system damage), nephropathy (renal system damage) and retinopathy (eye damage)grouped into microvascular damage or complications [96]. While cardiovascular diseases, stroke, and peripheral vascular disease lie in macrovascular group. Peripheral vascular disease may cause bruises and injuries that do not heal and lead to gangrene and amputation. NHANES data from 1999 to 2004 shows that occurrence of microvascular complications is much more higher than that of macrovascular complications) [95]. These complications may be episodic (eg, foot ulcers or infections) which can be treated and re occur many times. These usually begin slightly, but with time results in prolonged damage to the organ leading to complete non–functioning of organs. Other complications comprise dental disease, deceased resistance to various infections i.e influenza and pneumonia, and other birth related difficulties in pregnant women having diabetes. Type I and type II diabetes have similar complications to

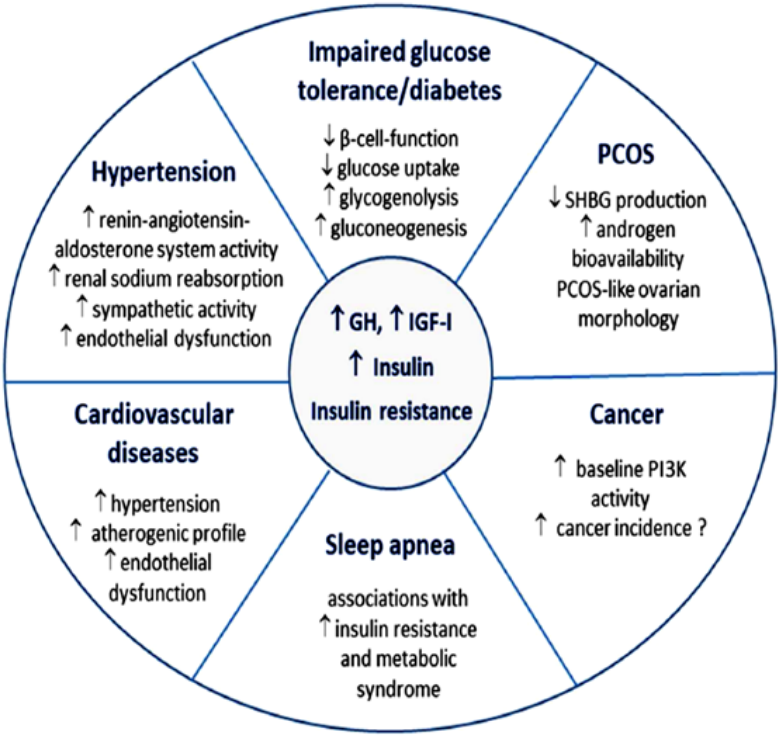


FIGURE 2.11: Outcomes of diabetes [95]

some extent but frequency and age of occurrence may vary accordingly.

2.10 Diagnosis and Treatment of Diabetes

Management of diabetes is to keep levels of blood sugar close to normal as possible. This control should be achieved without causing level of blood glucose to be low. This can be achieved typically by life style changes such as dietary modifications, weight control, exercise and walk and regular use of appropriate medication. To actively participate in management and treatment of disease. Knowledge of the disease is most important thing. As associated complications and risk factors are very uncommon and less severe in patients who have ample knowledge of the disease and manage their blood glucose levels very well [97],[98].

As per the American College of Physicians, 7–8% level of HbA1c is the goal of treatment of diabetes [99]. There are some other risk factors which can accelerate the

negative effects of the disease. These are smoking, hypertension, obesity, metabolic disorders, stress and lack of proper exercise. Special footwear is widely used to reduce the risk of ulcers on diabetic feet although the evidence for its effectiveness remains the same [100].

There are many tests to diagnose diabetes. The current diagnostic methods depend only on blood glucose levels monitoring, and are invasive as they all require blood samples. ADA (The American Diabetes Association)prescribe that people having no symptoms should be tested for T2DM by one of the following tests: OGTT (oral glucose tolerance) test, FPG (plasma glucose)test, and HbA1c (Hemoglobin A1c) test [101], [102] these methods have been recently approved by the ADA (American Diabetes Association) in 2010 and by WHO (World Health Organization) in 2011 (Figure 2.12).

Random blood plasma levels are also used as tests but results are not as much reliable as other mentioned tests [103].

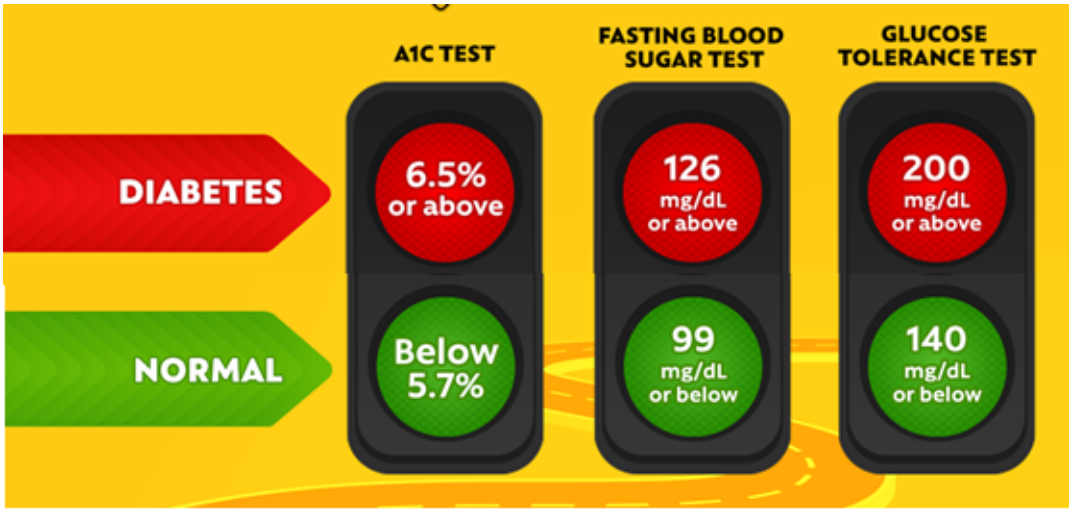


FIGURE 2.12: Diagnosis of diabetes [103]

2.11 Insulin

A very important polypeptide hormone which regulates metabolism of carbohydrates is insulin. It is derived from a latin word “insula” which means “island”. It is because

this hormone is produced in the islets of langerhenns of pancreas [104]. Metabolism of fatty acids, protiens and carbohydrates is regulated by insulin and it promotes the absorption of glucose into muscle especially skeletal muscles, liver and blood [105]. The end product of this absorbed glucose is either glycogen which is formed by glycolysis or triglycerides which is accomplished by lipolysis. Glucose can be converted into both i.e. glycerides or glycogen in liver as well [105]. High levels of insulin control the level of glucose production and secretion from liver by usually inhibiting it [106].

In many tissues protein synthesis also affected by insulin which is circulating in blood. So, it is an anabolic hormone which promotes conversion of small molecules in the cells into large molecules in the blood. On contrary decreased levels of insulin in blood pose an opposite effect and promotes catabolism. Major catabolic process in reversal of body fats. Specialized cells are sensitive to glucose levels in blood. These cells are called Beta cells. These cells respond to high glucose levels in the blood and secrets insulin in response and this effect is reversed when level of glucose is low in blood. Insulin secretion is inhibited by low levels of glucose [35].

Insulin helps in transportation of blood glucose into the various cells of body where it is metabolized into required products to produce energy. Glucose concentration in blood is regulated in this way. When there is high levels of glucose in the blood, insulin acts to increase reuptake of glucose by muscles cells and fat cells [104, 107]. chromosome 11p15 is the locus of INS gene and it is precusros of insulin [108]. In some mammals, such as mice, there are two insulin genes. one of these is a homologous of almost all mammals (Ins2), and the other is a retrospective copy that includes a sequence of stimulants but lacks an intron (Ins1). Both rodent insulin genes are active [109].

2.11.1 Structure and Function of Insulin

The molecule of insulin consists of 51 aminoacids arranged in two polypeptide chains which are linked by disulphide bonds. Structure is shown in fig 2.13. these chains are called alpha and beta chain. The alpha or A chain has 21 residues of amino acids. It

also has additional loop of disulphide bond between A 6 and A11 (Fig. 2.13), whereas the beta chain or B chain consists of 30 residues of amino acids. Primary molecular structure of insulin is known and studies from than 50 species of animals [110],[111].

Although during its working it acts as a single molecule i.e. monomer, while during its storage and biosynthesis it assembles itself into a dimer. It also sometimes assembles itself in hexamers in the presence of zinc. X–ray analysis of insulin insulin molecule has revealed all the three structures i.e. monomeric, dimeric and hexameric [112].

Insulin's crystal structure has been studied well and well documented. It describes the activity and binding affinity of insulin to receptors. It has been studied extensively and further elaborations and research are welcomed for insulin structure and structural activity relationship [113] According to a study and review, downstream signaling of insulin receptors interacts with other signaling pathways of growth factors i.e. IGFI and IGF2 [114]. This approach is very helpful in identifying and demonstrating the importance of insulin receptor ligand agonists. Because this can reveal the potential mimetic of insulins as therapeutic agents for the treatment of diabetes [113].

FIGURE 2.13: Human insulin's primary structure. The black residues are those amino acids which vary among different Animals [110, 111].

The monomeric 3D structure of insulin was first discovered through x–ray crystallography technique in 1926 [115]. As discussed in earlier section Insulin in humans

consists of 51 amino acids and its molecular weight is 5808Da. The biologically active form of insulin which is circulating one has two monomeric chains. These chains are alpha and beta and consists of twenty–one and thirty amino acids respectively. These are linked by strong disulphide bonds at position A7–B7 and A20–B19. At very low concentrations i.e. micromolecular concentrations it gets arranged in a very beautiful symmetric structure with the help of zinc. This structure is hexameric [113]. In insulin synthesis and activity, the alpha cells get cues from neighboring bets cells and secretes glucagon hormone in the blood in exactly opposite fashion [35] This secretion is regulated as per glucose concentration in the blood i.e if it is low, the secretion will be high and vice versa. Glucagon acts to increase the concentration or level of glucose by stimulating biochemical processed in the liver. These processed may be gluconeogenesis or glycogenolysis [105], [35]. Primary mechanism of glucose homeostasis is the secretion of glucagon and insulin in the blood as a result of blood glucose concentrations [35].

2.11.2 Biosynthesis of Insulin

The insulin contains 51 amino acids and have a molecular weight of 5.8kDa. However, the gene responsible for insulin encodes a precursor molecule known as preproinsulin which is mainly a 110 amino acid precursor. Similar to other secretary proteins, this insulin precursor has a hydrophic terminal which is signal peptide of N–terminal. This hydrophobic peptide interacts with signal recognition particles known as cytosolic ribonucleoproteins (SRP))[116]. Insulin is produced by beta cells within pancreas in mammals. Pancreas is primarily and exocrine gland and almost one to three million pancreatic cells (islets) form this endocrine part. This endocrine part is only 2% of the pancreatic mass. While within pancreas these beta cells accounts 65 to 80% of all cells [116].

Insulin is produced from beta cells in pancreas and has a very major role in regulating metabolism of various biochemicals i.e., carbohydrates and fats. It is synthesized as a polypeptide which is known as preproinsulin. This insulin precursor harbors a signal

peptide which consists of 24 residues. This directs the budding of polypeptide from endoplasmic reticulum (ER) in the cell. Further this signal peptide is broken down into polypeptides which is translocated to the ER. There it forms proinsulin precursor. The proinsulin is further packed and folded in the proper structure and confirmation by disulphide bonds. This insulin is then transported across golgi apparatus network. In golgi complex this molecule is converted into active insulin molecule by action of endopeptidases prohormone convertor which is PC1, PC2 and exprotease carboxy peptidases. for the formation of C peptide, this molecule is broken down at two different positions. So, a result mature insulin is formed which contains to chains i.e. A chain and B chain. These chains are connected to each other by disuplphide linkages and A chain also has an interchain bond as well [117],[118].

As discussed in earlier section, initially insulin is formed as a precursor in beta cells of pancreas islets. Short time after its synthesis and assembly in endoplasmic reticulums, this precursor of insulin in transported to complex glogi networsk within the cell. Granules are formed in trans glogi network (TGN) and these are immature in nature. This process of transportation may take upto thirty minutes. The two chain insulin molecules further undergo maturation to active insulin molecule by the action of endopeptidases in the cell. These endopeptidases cut the C peptide from insulin molecule at a position of 64 and 65 having lysine and arginine respectively. The second cite of cleavage is between 31 and 32 amino acid [120]. This mature insulin molecule is packed within mature granules which are waiting for further metabolic signals and nerve stimulation of vegal nerves for exocytosis from the cell in the blood circulation. Many signaling mechanisms involved in the regulation of insulin secretion from the beta cells of pancreas.

Phosphorylation of glucose through glucokinases is a mechanism involved after transportation through transporters (Glut 2). To yield ATP (adenosine triphosphate) Glucose–6 phosphate undergoes a rate limiting step and stepwise metabolism followed by oxidative phosphorylation.

Increased intracellular ATP concentration lowers the output of K1 and ultimately leads to beta cells depolarization. Voltage sensitive Ca^{2+} 1 channels are opened and

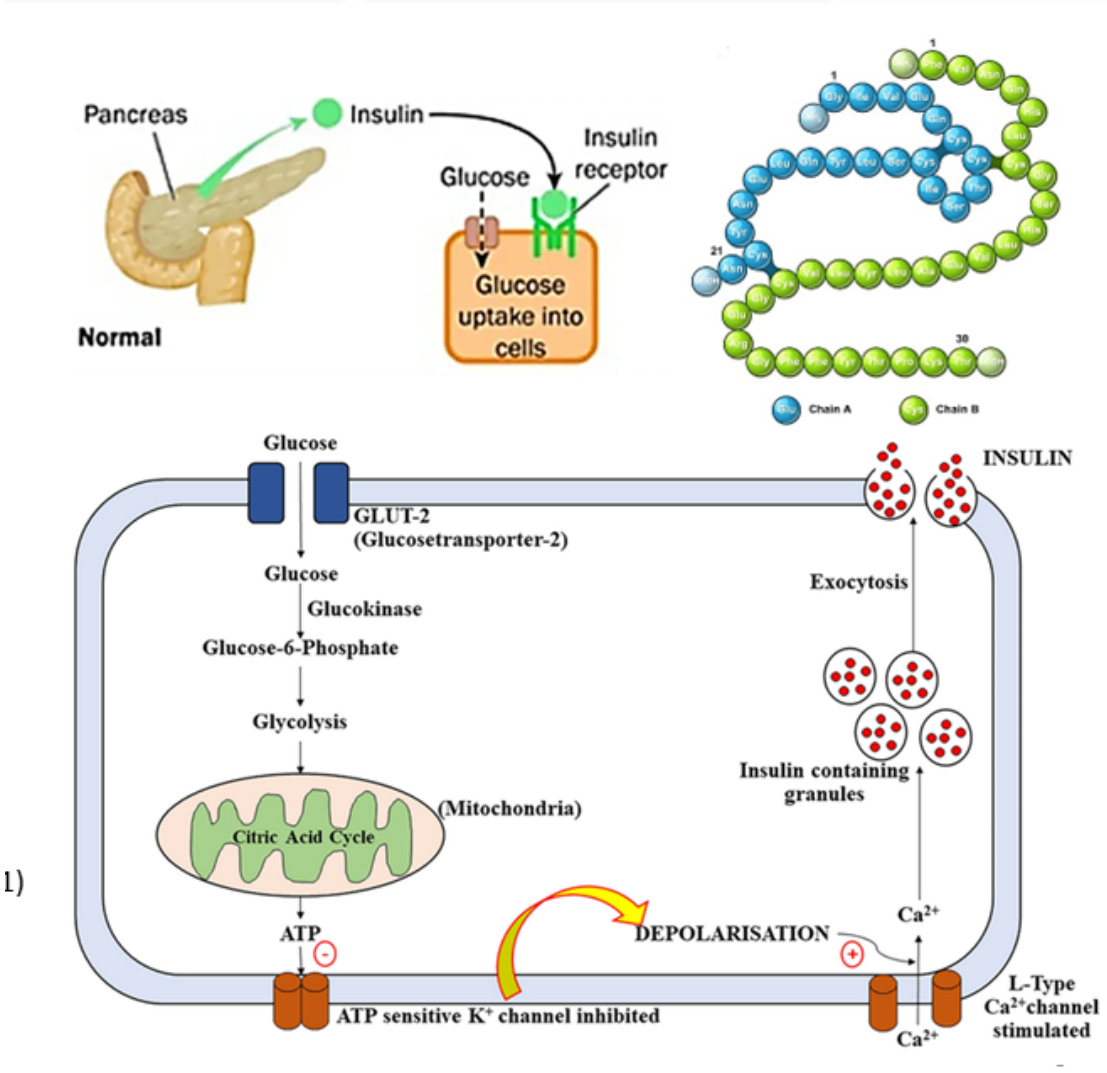


FIGURE 2.14: Mechanisum of synthesis of insulin [119]

this Ca^{2+} is transported inside the cell. When levels of Ca^{2+} are increased inside cell, active kinases gets stimulated and regulates insulin exocytosis and secretion into blood circulation [119].

2.11.3 Insulin Response or Insulin Action

Different types of tissues respond in a different way to insulin. Though tissue sensitivity to insulin associates with the levels of insulin receptors which are articulated on the plasma membrane. Now it has been clear that different components' assembly of insulin chemical signaling pathway is responsible for conferring specificity of insulin

on target cells. So, transport of insulin dependent glucose takes place only in adipose tissues and skeletal muscles. This is because these cells have insulin dependent transporter named as Glut4. Same as inhibition of glucose formation from glycogen with the help of insulin is specific to kidneys and liver.

While on the other hand, effect on ion transport. Synthesis of DNA and synthesis of proteins appears to be universal [119]. The Insulin employs multiple effects in the cell. Action of insulin is mediated by binding of molecule to its receptors. As a result phosphorylation of the receptors and other substrates by tyrosine kinase as shown in fig 2.14

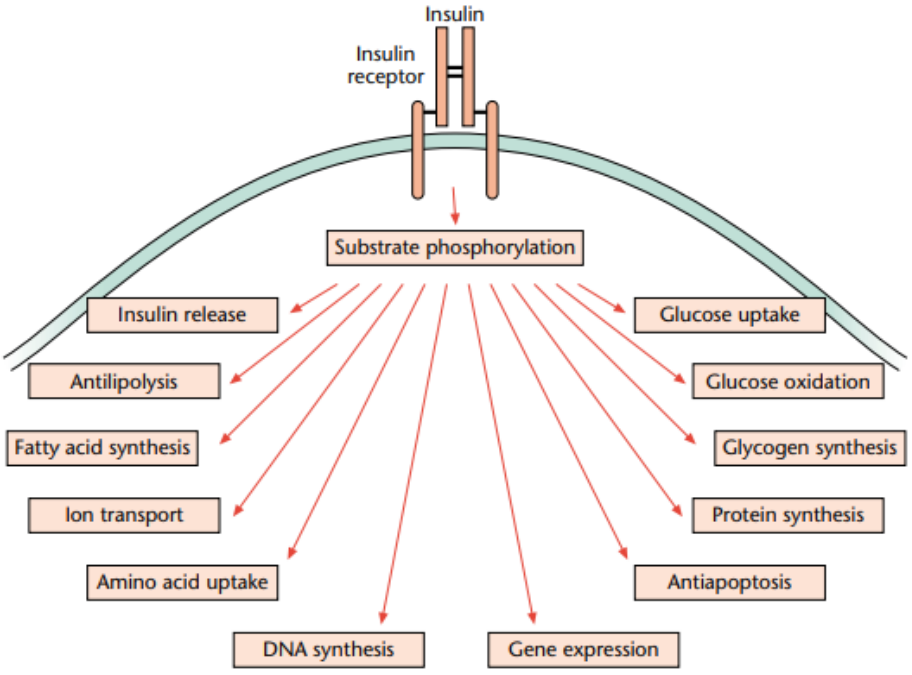


FIGURE 2.15: Action of insulin in cells. Multiple effects of insulin and mediation by receptor binding and tyrosine kinase [119].

2.12 Pathophysiology of Prediabetes

Blood glucose in healthy people is highly regulated. Between 3.9 and 5.6 mmol/L, fasting glucose is maintained and post–meal increase rarely exceed 3 mmol/L The homeostasis of fasting and post–load glucose becomes abnormal during the development of type 2 diabetes. As evidenced by repeated measurements of glucose levels,

insulin sensitivity, and insulin secretion studies, the development of NGT diabetes is an ongoing process [73].

Researchers has confirmed that decrease in insulin sensitivity and insulin secretion from beta cells precedes prediabetes to diabetes as increased glucose levels were observed in people recruited in study of British Whitehall II, 13 years before the diagnosis of diabetes. This pattern of glycemic changes has also been confirmed by other scientists. This suggest that IR begins years before the development of diabetes and that is the stage of prediabetes where decreased beta–cell functioning also exists [12].

Fasting plasma glucose levels are influenced by endogenous glucose production (EGP), which is primarily regulated by the liver. The EGP and fasting insulin product are used as markers of hepatic insulin resistance and show a strong correlation with fasting glucose levels. During the absorption of a glucose-containing meal, changes in glucose levels are influenced by intestinal glucose absorption, suppression of EGP, and total body glucose uptake. In individuals with normal glucose tolerance (NGT), EGP is effectively suppressed after glucose ingestion, but this suppression is less pronounced in those with prediabetes and diabetes. In type 2 diabetes, total body glucose disposal is reduced due to insulin resistance in muscle tissue. If insulin secretion can adequately compensate for insulin resistance, changes in glucose levels may not be noticeable. This suggests that β-cell dysfunction is present during the prediabetic phase by definition. However, β-cell activity cannot be fully assessed without considering the underlying insulin resistance, and cannot be defined solely based on insulin secretion alone [4].

As insulin secretion rises, it can be noted that the level of glucose in the pancreatic β-cell's response depends upon insulin sensitivity of the body. Such dependencies are hyperbolic in nature. The disposition index is a good way of assessing the levels of insulin secretion as a function of insulin resistance. It is usually high in normal individuals and low in persons suffering prediabetes and diabetes. Several medical practitioners have documented abnormal insulin secretion amongst prediabetic people and mostly attributed this to other factors affecting around the β-cells. Studies suggest that there is a fifty percent decrease in β-cell volume among persons with

glucose intolerance. Insulin resistance (IR) was present in both impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) conditions, but the target location of IR was distinct: high hepatic IR was prevalent in IFG while skeletal muscle IR was nearly spared. On the other hand, in subjects with IGT, there is a marked resistance to insulin in the muscle with only minimal improvement in hepatic insulin action. There is an orderly sequential decrease in the total body glucose disposal from normal glucose tolerance (NGT) as type 2 diabetes disorder progresses through any form of Impaired Fatsu Glucose (IFG) to an impaired glucose tolerance(IGT). Impairment of beta (β)-cell responses is present in both impaired fasting glucose and impaired glucose tolerance conditions [12].

2.13 Diagnosis of Prediabetes

Diagnosis of diabetes involve blood glucose level equal to or above specific value. According to ADA there are three methods for the diagnosis of diabetes and some methods can be used for detection of prediabetes in patients.

These methods are Fasting Plasma Glucose Test (FPG), Oral Glucose Tolerance Test (OGTT) and Glycated hemoglobin or hemoglobin bounded to glucose (A1C)

The threshold values of each test are summarized in table 2.2 as diagnostic criteria of prediabetes and diabetes

TABLE 2.2: Diagnostic criteria for prediabetes and diabetes [11]

	FPG	**PG in OGTT**	**A1C**
Normal	<100mg/dL or 5.5mmol/L	<140mg/dL or 7.8mmol/L	<5.7% or 39mmol/mol
Prediabetes	≥ 100mg/dL or 5.5mmol/L	≥ 140mg/dL or 7.8 mmol/L	≥ 5.7% or 39mmol/mol
Diabetes	≥126mg/dL or 7.8 mmol/L	≥200mg/dL or 11.1 mmol/L	≥6.5% or 48mmol/mol

2.14 Limitations of Diagnostic test for Prediabetes

There are several pros and cons in above diagnostic methods. For example, FPG greater than or equal to 140 mg/dL is specific but insensitive test for diagnosis of type 2 diabetes. Alost every subject with FPG $\geq$ 140mg/dL could have plasma glucose in OGTT $\geq$200mg/dL but a many subjects depending on population with plasma glucose in OGTT $\geq$200mg/dL may not have FPG$\geq$ 140mg/dL [73]. A1C test detects the glycation of proteins and indicates chronic hyperglycemia but A1C test does not detect high blood glucose levels directly.

Measuring plasma glucose in OGTT is more accurate than measuring fasting plasma glucose and A1C in diagnosis of type 2 diabetes. However, plasma glucose in OGTT $\geq$200mg/dL also indicates impaired glucose tolerance. It is therefore recommended that more than one test should be used in diagnosis of diabetes [11].

Fasting Plasma Glucose (FPG) level and 2–hour post–glucose load alone did not always detect or predict IGT and IFG, both tests are useful in identifying the altered glycemic conditions. Limitation of A1c test include some specific hemoglobinopathies such as fetal hemoglobin falsely increases and sickle cell hemoglobin (hemoglobin C) lowers A1c levels. Therefore there is possibility of false–negative diagnosis in patients with such conditions [4].

A poor correlation is reported in literature between A1c, IFG and IGT. Due to inability of these blood glucose cut points to capture pathology related to diabetes and probability of developing diabetes in near future, the usefulness of diagnosis of diabetes or prediabetes on the basis of IFG and IGT have been challenged [6].

All three diagnostic tests have their pros and cons. Typically fasting plasma glucose level greater than or equal to 140mg/dL is a very specific but insensitive test for prediction of diabetes mellitus [121]. Almost all individuals with FPG $\geq$ 140 mg/dL will have plasma glucose in oral tolerance test equal to or greater than 200mg/dL . But a substantial portion of individuals (depending on the population) with PG in OGTT $\geq$ 200 mg/dL will not have an FPG $\geq$ 140 mg/dL. Glycation of proteins is

measured by HbA1C test and it indicates chronic hyperglycemia, but this test does not directly prove increased blood glucose levels [122].

As per WHO, if people have impaired fasting glucose or impaired glucose tolerance, they are increased risk of having diabetes. Impaired fasting glucose (IFG) is defined as a fasting plasma glucose levels of ≥6·1 and <7·0 mmol/L, but without reduced glucose tolerance; and impaired glucose tolerance is defined as an fasting plasma glucose concentration of <7·0 mmol/L and a 2 hour post load plasma glucose concentration of ≥7·8 and ≤11·1 mmol/L.

This definition is subject to measuring during a 75 g oral glucose tolerance test (OGTT). Same threshold for impaired glucose tolerance but a lower cutoff for impaired fasting glucose is applied by the American Diabetes Association (ADA). ADA has presented glycated hemoglobin A1c (HbA1c) 5·7–6·4% as a new group for higher diabetes risk [123].

Any pre–diabetes definition limited to IGT and IFG does not include other risk factors for diabetes, such as family history of type 2 diabetes or metabolism. Another reproach of the term pre–diabetes is that many individuals with IFG or IGT will not develop type 2 diabetes. That is why, another name is preferred.

So far the "intermediate hyperglycemia" of the WHO interim group has not been widely accepted [124]. Another name for this is "border line diabetes", but this term is not presently suggested and has no prescribed definition [96].

In 2012 CDC (Centers for Disease Control) had estimated that 86 million people or 1/3 adults had diabetes in the US. However almost 90% of individuals did not know that they are diabetic.

In 2015, the International Diabetes Federation projected that the worldwide occurrence of impaired glucose tolerance (IGT) in adults was 318 million and anticipated to reach 482 million by 2040[125]. The annual development rate to diabetes is 5–10% [6]. Older individuals, having severe insulin resistance (IR), low insulin secretion, and other associated diabetes risk factors are more likely to progress [126].

2.15 Omics

Omics technologies include genomics, proteomics, transcriptomics, and metabolomics, each providing a comprehensive view of molecular interactions within a biological system. These high-throughput techniques enable researchers to analyze complex biological networks, uncover disease mechanisms, and identify potential therapeutic targets. The integration of multi-omics data further enhances precision medicine and personalized treatment strategies. Furthermore, advancements in bioinformatics and computational biology have significantly improved the analysis and interpretation of omics data. These approaches are now widely used in various fields, including drug discovery, agriculture, and environmental sciences, to address complex biological challenges[69].

2.15.1 Genomics

The study of the complete set of DNA, including all of the genes ("the genome"), of an organism is known as genomics. Genome–scale data acquisition has been made far more convenient with the emergence of Next–Generation Sequencing, abbreviated as NGS. This has greatly improved the ability for the analysis of whole genomes, while decreasing the gap that currently exists between phenotype and genotype [69].

Genome–wide association studies, commonly referred to as GWAS, is a research approach usually applied for the identification of genomic variants that are associated statistically with a specific trait or the risk of the development of a disease. Advancements in genomics has turned this research approach into the gold standard method for the identification of complex regions that are associated with complex traits of interest [33]. The Human Genome Project (HGP) has greatly facilitated genomics. The creation of high– resolution genetic and physical maps of every single human chromosome to enhance the ability of localization and identification of genes related to inherited disorders and/or other traits was among the major goals of the HGP. Genetic maps are of high significance, especially for the development of physical maps that depict the order on the chromosome of DNA sequences [34].

2.15.1.1 High–throughput Techniques in Genomics

A variety of techniques and tools exist under genomics that have important applications. In the genomes of individuals of same species, there are slight variations known as single nucleotide polymorphisms, referred to as SNPs that contribute to diversity among them and may also contribute to disease in some cases.

Certain Probe–based chips have been developed consisting of a large number of SNP markers that are spread across the genome. These chips can be employed for the uncovering of associations between genes and the traits of interest. Specie–specific arrays that encompass a vast number of SNPs are available, ensuring that there will be a close link of at least one marker with a quantitative trait loci [35]. This has facilitated GWAS even more. For post–GWAS analysis, gene–based software exist. Many different SNP array data management tools have been developed, with PLINK being the standard [36]. Genomic selection is a form of marker–assisted selection in which genetic markers that cover the whole genome are used for the estimation of animal breeding values. It is applicable in animal breeding done for a trait or traits of interest.

2.15.2 Proteomics

The characterization and quantification of all sets of proteins in a cell, organ, or organism at a specific time is known as a proteome. Proteomics is used for the quantification of proteins in multiple sample types. It uses both shotgun and targeted approaches [37]. Bottom–up proteomics indicates protein characterization by analysis of peptides released from the protein by proteolysis, which, when performed on a protein mixture is known as shotgun proteomics [38]. In top–down proteomics, large protein fragments or protein ions are analysed by mass spectrometry. This is also known as targeted approach [40]. Proteomics facilitates the identification of post–translational modifications and their localization, novel biomarker discovery, as well as study of protein–protein interactions. Proteomics is being adopted by more and more researchers following the establishment of powerful techniques and tools for the identification and

quantification of protein species from complex biological samples [127].

High–throughput Techniques in Proteomics

Mass spectrometry (MS) is the core of proteomics. In case a pre–separation of protein mixtures is needed prior to MS–analysis, one–dimensional or two–dimensional polyacrylamide gel electrophoresis is commonly employed [128]. For the enhancement of automation and creation of a streamed pipeline analysis, liquid chromatography of different types is used to either substitute or accompany gel electrophoresis. Absolute protein quantification is possible through the use of Absolute Quantification of Proteins (AQUA), Artificial Proteins Comprised of Concatenated Peptides (QConCat), Protein Standards for Absolute Quantification (PSAQ) [41]. System–wide screening methods consisting of quantitative steps that utilize label–based techniques such as Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC), Isotope–Coded Affinity Tagging, O Stable Isotope Labelling, Tandem Mass Tags (TMTs), and Isobaric Tagging for Relative and Absolute Quantitation (iTRAQ) are now being implemented as well [42].

2.15.3 Metabolomics

Metabolomics deals with metabolites, which are the end products biochemical processes (metabolism). Metabolites can link genome and proteome to a phenotype, which provides a key to discover the genetic basis of metabolic variation. Metabolomics can utilize a variety of sample types such as cells and cell lines, tissues, biological fluids etc. The metabolome itself consists of the global profiling of all the metabolites in any biological sample [37]. Metabolomics can be used for the detection of molecules such as peptides, amino acids, carbohydrates, nucleic acid, vitamins, alkaloids, organic acids, polyphenols, as well as inorganic species.

High–throughput Techniques in Metabolomics

The small molecule information can be captured in solid, liquid, or gas phase by metabolomics depending on the application and instrumentation. Typically, high

resolution analysis combined with statistical tools are utilized by metabolomics for the derivation of an integrated image of the metabolome. The statistical tools that are used include Principal Component Analysis (PCA), and Partial Least Squares (PLS). Metabolomics commonly utilizes Nuclear Magnetic Resonance (NMR), which can be used for the identification and the simultaneous quantification of a vast range of organic compounds in the micro–molar range. Even mass spectrometry is gaining increased applicability in high through–put metabolomics, usually accompanied by techniques such as electrophoretic techniques or chromatography. Mass spectrometry is gaining preference in metabolomics due to its covering of a wide–range of metabolites as well as its sensitivity [43].

2.16 Systematic Literature Review

Literature reviews contribute significantly to the advancements in science. Even though they are not exactly similar to original research studies, their significance lies in the fact that literature reviews can sum up the existing knowledge and give an idea of how to proceed, as well as the direction to proceed into. Moreover, they can identify potential gaps in the current knowledge, which shows what research could be beneficial to overcome those gaps. There are multiple reasons and advantages of literature reviews. Among the major ones is to assess and sum up thecurrent knowledge and research done. Besides that, important benefit of literature reviews is to identify what is currently unknown. Not only does that highlight the gaps in the current information, it also suggests topics for the future researches. Advancement of theories is also an important reason for literature reviews. Literature reviews can also provide regarding the research findings [44].

2.16.1 Methods and Techniques for Literature Review

A variety of methods are available for carrying out a literature review. The most common method is the Traditional Narrative Review. This method involves verbal

descriptions of research findings that are presented with conclusions as well. Box Score review is a form of review in which there is some quantification involved. It can be used to show the frequency of a particular proposition or for the revelation of an interpretable pattern [44].

2.17 Meta–Analysis

Meta–analysis is a statistical technique that is used for combining the findings from different studies. Meta–analysis usually combines the data from two or more studies. These studies are often addressing the same problem or are about the same topic. The proper coverage of relevant studies and explore the main findings using sensitivity analysis [45].

2.17.1 Methods and Techniques for Meta–Analysis

The methods that are often used for meta–analysis include Mantel–Haenszel methods, that are fixed effect meta–analysis methods that use a different weighing scheme, which depends on the effect measure being employed, for instance, risk ratio, odd ratio, risk difference [46]. These methods work better in cases with few events. Another method is Peto's Odds Ratio Method but it can only be used to combine odd ratios, using an inverse–variance approach but an approximate method for estimation of the log odds ratio [47].

2.18 Prediabetic Insulin Resistance

Both insulin resistance and b–cell dysfunction function have been observed during prediabetes even during early stages and are needed for the majority of prediabetes hyperglycemia. The earliest abnormality appears to be IR (insulin resistance), although in studies which investigate the course of pre–diabetes pathogenicity a substantial he-

terogeneity between individuals and populations exists.

This is no surprise; studies of genome–wide sequencing studies (GWAS) have shown that at least 60 genes confer a risk of type 2 diabetes growth, most of them linked to b–cell biology [124]. Researchers have reported signs of IR and early intermittent hyperglycemia approximately 13 years earlier the diagnosis of diabetes. Blood glucose was close to normal possibly due to compensatory mechanisms to improve pancreatic beta cell insulin development, before they were defeated 2–6 years prior to diagnosis at a time when there was more sustained hyperglycemia that also had to do with declining IR [125].

A few years later, with now clear hyperglycemia, a steeper decrease in the pancreatic b–cell function appeared to lead to a clinical diagnosis of diabetes. But what causes the plasma glucose levels to increase in precisely both fasting and fed condition? The pathophysiology of IFG and IGT was investigated through euglycemic hyperinsulinemia clamping studies. Their data indicate that the isolated IFG (normal post meal blood glucose [126], but not any evidence of hepatic IR, was associated with increased glucose (i.e., glycogenolysis was normally suppressed). Those with a combined IFG/IGT, however, had combined increased gluconeogenesis, insulin–reduced glycogenolysis failure, and extrahepatic IR decreased glucose disposal [129].

2.19 Biomarkers

Biological molecules which could help in the indication of some normal or abnormal body processes or some specific disease are called biomarkers. These are present in body fluids or more specifically biological fluids i.e., saliva, serum and urine etc. and mainly are genes, proteins and some metabolites of biochemicals [130]. Moreover, these biological molecules are also very important for diagnosis and early detection of abnormalities in the body which lead to disease [130]. Extensive research work has been done on identification of many biomarkers but there are some limitations due to which many researches have not extended beyond clinical trials. This is because the specificity, sensitivity and reproductivity are inadequate. So, before clinical trials it is

of great importance that biomarkers should be properly validated. Biomarkers used in clinical settings are of three types which are diagnostic biomarkers, prognostics and threputic biomarkers.

Diagnostic are those which are used in disease identification while prognostic biomarkers are those which are used to predict outcomes of disease [131]. Some common clinically approved biomarkers are hCG (human chorionic gonadotropin) and PSA (prostate specific antigen). hCG is important biomarker of pregnancy and found in urine while PSA helps in the diagnosis of prostate cancer [132]. Despite the numerous biomarker advances, theragnostic biomarkers are still very limited in number [133]. Research has suggested that these biomarkers can be helpful for accurate dosage, prediction of responses for a particular treatment, maximizing efficacy of drugs and minimizing toxicity of drugs for relevant individuals [88]. A good biomarker should be reproducible, sensitive, specific and must be standardized [130]. In recent years several approaches to identifying new biomarkers have been employed[133]. For the discovery and validation of new biomarkers, computational biology has played very important instrumental role in the detection [88].

2.19.1 Applications of Biomarkers

The application of biomarkers can be as under;

- Disease screening, Disease characterization (e.g., trinucleotide repeats)[134].
- To rule out, diagnose, staging, and monitor illnesses
- Prognosis information
- Individualize the therapeutic interferences by response monitoring responses to therapies and predicting consequences in response to them
- Prediction of adverse drug reactions
- To predict and guide drug toxicity treatment (e.g., serum concentration measurements following medication overdose)

- Cell type identification (e.g., histological markers). Use of biomarkers for optimization and individualization of doses of many drugs is not very welcoming approach. For example effects of warfarin on the coagulation process is a complex model and includes two genetic biomarkers that produce very little therapeutic effect [135]. These are also used at numerous stages of drug discovery and development.

- It is also used in screening compounds as a target in drug discovery i.e., cyclooxygenase activity measurement for the identification of potential anti-inflammatory agents.

- Biomarkers are also used as endpoints in pharmacodynamic studies, e.g., for the prevention of heart diseases, serum cholesterol is used as a marker for the action of drugs, and in pharmacokinetic and pharmacodynamics studies [136].

2.20 Biomarkers Associated with Diabetes

Diabetes and prediabetes are metabolic disorders characterized by impaired glucose regulation, and the identification of biomarkers plays a crucial role in their early detection and management. Several biomarkers have been associated with the pathophysiology of these conditions Table 2.3.

Particularly those linked to insulin resistance, inflammation, and beta-cell dysfunction. Key biomarkers include fasting plasma glucose (FPG) and hemoglobin A1c (HbA1c), which reflect long-term glucose control. Additionally, insulin levels, C-peptide, and Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) are commonly used to assess insulin sensitivity. Inflammatory markers such as C-reactive protein (CRP) and interleukin-6 (IL-6) have been implicated in the development of both prediabetes and type 2 diabetes, indicating an underlying inflammatory process. Furthermore, elevated levels of adipokines like leptin and reduced adiponectin are frequently observed in individuals with metabolic disturbances, further contributing to insulin resistance and glucose dysregulation.

TABLE 2.3: The available list of biomarkers used in diagnosis and prognosis of diabetes.

Bio-markers	**Mechanism of Action**	**Association with Deglycation**	**Ref.**
	Traditional Biomarkers		
Hba1c	There is β subunit portion of hemo-globin and attachment of glucose to its amino-terminal part produces Hba1c	Levels of Hba1c is a reflection of continuing glycemia	[137] [138]
FA (Fruct osamine)	It is a keto amine produced by glycosylation of all proteins of serum (primarily albumin)	Increased level of glucose conc. directly affects FA level	[139]
Glycated albumin	It is produced by Albumin glycosy-lation is calculated by the GA fraction to total albumin	Serum concentration of GA is linked with diabetes and prediabetes condition	[140] [141]
1,5 AG	It is a alimentary monosaccharide, and plasma absor-ptions of it is oppositely linked with plasma glucose level.	There are low plasma levels of 1,5 AG in persons with compr-omised glycemic controls in comparison with normoglycemic.	[142]
OGTT	It is oral glucose	Elevated levels	[143]

	tolerance and a direct measure of fasting and 2-hour post plasma glucose levels	are linked with prediabetes and diabetes.	
	Inflammatory Biomarkers		
CRP	It is derived from a primary marker of biosynthesis which is response of acute phase (il-6-dependent hepatic marker)	Directly linked to type II diabetes.	[144] [145]
IL-6	It is an immuno regulator and involved in sustaining glucose balance and metabolism by its action on β cells of pancreas, adipocytes, hepatocytes, and skeletal muscles.	It is related with type II diabetes and IR.	[146] [147] [145]
Wbcs	WBCs number is a immunity and inflammatory marker and it's an indicator of inflammation and vascular (micro and macro) complications in diabetes.	WBCs count is a predictive of impaired insulin action, T2DM, vascular heart diseases.	[148] [149] [150]
Fibrinogen	Fibrinogen is respons-	It is associated with:	[151]

	ible for blood thickness, platelet accumulation, and fibrin for clotting	prediabetes and feebly with diabetes. It is also related to increased 1-hour post levels of glucose and atherosclerosis	[152]
PAI-I	It is indicator of lowered fibrinolysis and its lowered activity leads to substandard coagulation	PAI-I is an independent predicator of diabetes.	[152] [153]
IL-I8	IL-18 levels rise during hyperglycemia.	Directly associated with high risk of type II DM. It is also an indicator of progression from prediabetic state to diabete	[154]
IL-IRA	an anti-inflammatory biomarker rises when pathway of IL-1 is induced by glucose and free fatts via over eating.	level rises in diabetes prediabetic state, decreased sensitivity of insulin	[155] [147]
	Novel Biomarkers		
Adiponectin	High insulin resistance and obesity are associated with lesser levels of adiponectin.	It comes from adipose tissues, anti-inflammatory, anti-therogenic and insulin sensitizing	[156]

		effects	
FetA	Higher risk type II DM with some other complications.	Glycoprotein of liver, encourage lipid-induced IR signaling pathwayof TLR4 production of various inflammatory cytokines	[157] [158]
A-HB	Linked with IR, related with decrease in glycine and serine. upstream of α-KB	Organic derivative, produced through the combination of α-KB. α-KB is a product of amino acid catabolism	[159] [160] [159]
L-GPC	L-GPC is a negative indicator of type II DM progression.	Metabolite formed in the liver by the enzyme phospholipase A2 and by acyltrans-ferase of lecithin cholesterol in the circulation.	[159] [161] [162]
Lp(a)	Reverse relationship of Lp(a) with occurrence of prediabetic state and T2DM.	It is a lipoprotein that lead to atherogenesis.	[163]
HDL	HDL-C encourages insulin secretion and its low conc. may lead to	It is a major lipoprotein.	[164]

	development from prediabetes to diabetes		
Ceramide	It is positively linked with prediabetes and type II DM.	Lipid molecule	[165]
Ferritin and transferrin	Radical formation, β-cell oxidation Damage to DNA hepatic dysfunction, and β-cell apoptosis	Ferritin is an intra cellular protein which stores and control release of iron.	[166] [167]
MBL-associated serine proteases	MASP1 is linked with prediabetes, diabetes, CVD	Enzymes for activation of complement system (lectin pathway) and innate immune responses.	[145] [168]
THBS1	THBS1 is linked with prediabetes, diabetes, CVD	THBS1 initiates formation of multiprotein multiplexes that modulate cellular phenotypes.	[168]
GPLD1	GPLD1 linked with with prediabetes, diabetes, CVD	Role in the pathway of glycosylphos-phatidylinositol	[168]
Acylcarnitine	Increased levels of acyl-carnitine	Acyl-carnitines interact with important inflammatory marker which is NF-Kβ,	[169] [170] [171]

		and promote IR.	
miRNA	Mir-192 and 193b	gene expression	[172]
	found to linked with	promotes insulin	[173]
	IFG, IGT, triglycerides	synthesis.	[174]

Genetic and epigenetic factors further influence insulin resistance, highlighting the complexity of metabolic disease development. Integrating these biomarkers enhances early detection and targeted intervention strategies.

2.21 Identification of Biomarkers

Good understanding with the pathophysiology of the disease is a basic and first step in identification of biomarkers. Factors associated with it also of basic importance. Therefore, a better understanding of any disease makes it workable and easy for identification of the various factors which may be associated with it. Biomarkers associated with mechanism and etiology are best suited for prediction and proper diagnosis of the heart diseases in heart failure. This is also helpful for studying the development and progression of the disease. The second step in biomarker identification is mechanism understanding, whereby intervention can affect the pathophysiology of the disease [136].

The use of biomarkers is very useful for the drug development process as well. It is used in studying different aspects of illness and monitoring the fruitful effects of applied interventions. However, as the events linking pathogenesis to outcomes are usually complex, the easier it is to recognize biomarkers that can diagnose a disease, track the reaction to a medication, and test disease progression are the more aware of the underlying abnormalities of drug conditions and mechanics. For undeveloped and clinical pharmacologists as well as others involved in the identification of biomarkers accumulation of this information is a challenge [136].

2.22 Insilco Identification of Biomarkers

These tools (table 2.4) are reported in the literature for the following functions and after that, we select the best tools for the specific function

TABLE 2.4: Available tools for different analysis

S.No	Tool Name	Function	Reference
1	Go term finder David Go term mapper	Gene annotation	[125]-[129]
2	Blast FASTA MUSCLE	Sequence alignment	[130]-[132]
3	Swiss model Fold X I TASSER Hpred	Homology modeling	[118],[135]
4	Saves Verify 3D	Structure validation	[136],[175]
5	CATH Pock drug p2rank	Active pocket	[137],[138]
6	Autodock vina Arguslab ClusPro	Modeling and simulation	[139],[140],[137]
7	PLIP Struct2net Prise	Protein interactions	[141]-[143]
8	Webgivi David	Enrichment analysis	[131],[126],[144]

TopFun

2.23 Risk factors of Prediabetes

Several factors increase the risk of developing prediabetes (figure 2.16). Being overweight or obese is a significant risk factor, particularly when excess fat is stored around the abdomen, as it contributes to insulin resistance. Age is also a factor, with individuals over 45 being more susceptible. A family history of type 2 diabetes increases the likelihood of developing prediabetes, as does leading a sedentary lifestyle. Certain ethnic groups, including African Americans, Hispanics, Native Americans, and Asian Americans, have a higher predisposition to this condition.

FIGURE 2.16: Risk factors of prediabetes [63]

Women who experienced gestational diabetes during pregnancy face an elevated risk of prediabetes later in life, and those with polycystic ovary syndrome (PCOS) are similarly at increased risk due to hormonal imbalances and associated weight issues. High blood pressure and abnormal cholesterol levels, particularly low HDL cholesterol and high triglycerides, are linked to a greater risk of insulin resistance. Additionally, sleep disorders such as sleep apnea, smoking, and diets high in red and processed meats,

sugary drinks, and highly processed foods further contribute to the development of prediabetes. Addressing these risk factors through lifestyle changes like weight loss, increased physical activity, and healthier eating can significantly reduce the risk of progressing from prediabetes to type 2 diabetes.

2.24 Genetic and Molecular Risk Factors for Prediabetes

Prediabetes is a symptomatic stage that progresses to type 2 diabetes. At this stage, IR is compensated by hyperinsulinemia. Prediabetes is a stage at which active intervention programs such as healthy diet and exercise can delay or stope the progression of the disease to type 2 diabetes [176]. The discovery of early biochemical changes that are associated with this stage of disease is important for individuals who are at high risk of developing type 2 diabetes so that they can get benefit from above intervention programs. These molecular markers may also be used in monitoring the progression of disease as well as understanding the pathogenesis of disease. Considering these facts molecular markers associated with IR, a hallmark of prediabetes, suggests a solid reason for the identification of disease molecular biomarkers of prediabetes.

Several genes and associated SNPs (table 2.5) have been reported in literature that are linked with prediabetes [72].

TABLE 2.5: Genetic biomarkers associated with prediabetes

S.No.	Genes	SNPs	Reference
1	ADIPOQ	rs1501299 rs266729	[176]
2	TCF7L2	rs7903146	[177], [178]
3	IGF2BP2	rs4402960	[177]
4	CDKAL1	rs7754840	[177]
5	HHEX	rs1111875	[177]

6	HNF1A	rs1169288	[177]
7	TNF- α (-238G/A, -308G/A)	rs361525 rs1800629	[179]
8	UCP2	rs659366	[180]
9	MTMR9	rs2293855	[181]
10	TRPM6	rs8042919	[182]
11	CLDN19	rs719676	[182]
12	SLC41A2	rs2463201	[182]
13	CNNM2	rs3740393	[182]
14	FXYD2	rs948100	[182]
15	GCK	rs2908289	[183]
16	YKT6	-	[183]

A biomarker is a naturally occurring molecule such as gene, protein, SNP, miRNA, metabolite or characteristic through which a particular pathological or physiological process, diseases, altered health outcome etc. can be identified. A biomarker is a measurable indicator of presence of disease sate or indicator for representing severity of disease. More generally a biomarker is any molecule that can be used to detect the particular disease state or some physiological state in an organism. For insulin sensitivity, adipokines and more recently myokines and hepatokines are found to be potential biomarkers but still they are not used in clinical settings. Due to which in routine screening many cases of prediabetes are missed, therefore, a better biomarkers are needed for a simple and early detection of abnormalities of glucose metabolism and prediction of type 2 diabetes at prediabetic stage [178].

TABLE 2.6: Molecular biomarkers associated with prediabetes

S.No.	**Molecular Biomarkers**	**References**
1	adiponectin	[176]
2	RBP4	[184]
3	chemerin	[184]

4	A-FABP	[184]
5	FGF21	[184]
6	fetuin-A	[184]
7	TNF-α	[179]
8	Uncoupling Protein 2	[180]
9	Myotubularian	[181]
10	myostatin	[184]
11	IL-6	[179]
12	Serum Magnesium	[182]
13	irisin	[184]
14	C-Reactive Proteins (hsCRP)	[185]
15	Amino acid (valine, leucine, and isoleucine, tyrosine, phenylalanine)	[186]
16	Advanced glycation end product	[187], [188]
17	miRNAs (miR-9, miR-29a, miR- 30d, miR34a, miR-124a, miR146a and miR375)	[59]

When detecting biomarkers, both targeted and untargeted methods can be employed. Targeted methods focus on measuring or detecting a specific subset of molecules that are pre-defined. In contrast, untargeted methods aim to capture all detectable signals, with the goal of identifying and annotating as many molecular species as possible from the experimental data.

Targeted methods offer greater reliability because they can be compared with established reference samples. However, they are limited to detecting only the molecules that were specified in advance. On the other hand, untargeted methods have the

potential to measure a vast range of markers, limited only by the capabilities of the platform technology.

However, these methods face challenges due to the current lack of standardized procedures, the complexity of data processing, and the availability of resources for accurate annotation.

Several biomarkers have been reported in literature (table 2.6) but these are still not in use in clinical setting. It is still not clear that how these molecular biomarkers are involved in progression of prediabetes, diabetes and other comorbidities associated with prediabetes [184].

This project aims to explore mechanism for diagnostic and management strategies for Pre–diabetic insulin resistance by searching genetic and molecular mechanism that are associated with pre–diabetic insulin resistance.

2.25 Elucidation of Omics to understand Prediabetes

The identification of genetic, protein, and metabolite biomarkers is crucial for understanding and predicting the development of prediabetes. These biomarkers can provide valuable insights into the underlying mechanisms of the disease and help in early diagnosis and intervention.

2.25.1 Genetic Biomarkers

Genetic variations can significantly influence an individual's susceptibility to prediabetes. Specific genes associated with insulin resistance, glucose metabolism, and beta-cell function have been identified. For instance, variations in the TCF7L2 gene are strongly linked to an increased risk of developing type 2 diabetes and prediabetes. Other genes, such as PPARG and FTO, also play roles in fat metabolism and obesity,

which are critical risk factors for prediabetes.

2.25.2 Protein Biomarkers

Key proteins in insulin signaling, glucose transport, and inflammation serve as essential biomarkers for prediabetes. Low adiponectin levels correlate with insulin resistance, while elevated C-reactive protein (CRP) indicates increased inflammation. Additionally, insulin-like growth factor-binding protein 1 (IGFBP-1) and fetuin-A are linked to insulin resistance and glucose intolerance.

2.25.3 Metabolite Biomarkers

Metabolomics, the study of small molecules (metabolites) within cells, tissues, or organisms, offers a comprehensive approach to identifying biomarkers for prediabetes. Metabolites involved in glucose and lipid metabolism, such as glucose, lactate, and free fatty acids, are key indicators. Elevated levels of branched-chain amino acids (BCAAs), including leucine, isoleucine, and valine, have been linked to insulin resistance and an increased risk of developing prediabetes.

Additionally, altered levels of acylcarnitines and ceramides have been observed in individuals with insulin resistance, providing further insights into metabolic disturbances associated with prediabetes. The identification and study of genetic, protein, and metabolite biomarkers are essential for advancing our understanding of prediabetes. These biomarkers not only help in early detection and risk assessment but also provide targets for potential therapeutic interventions. Ongoing research in this area holds the promise of more effective prevention and management strategies for prediabetes and related metabolic disorders.

Together, these metabolomic insights underscore the complexity of prediabetic insulin resistance and highlight the potential of integrated omics approaches in uncovering novel diagnostic and therapeutic targets. Advancing our understanding in this domain

is crucial for developing personalized strategies to prevent the progression from pre-diabetes to type 2 diabetes.

2.26 Meta-analysis, Systematic and Bioinformtaics Approch to Elucidate Molecular Mechanisum of Prediabetes

The elucidation of the molecular mechanisms underlying prediabetes often employs meta-analysis, systematic review, and bioinformatics approaches, which collectively offer comprehensive insights into the disease. Meta-analysis integrates data from multiple studies to synthesize findings and identify consistent patterns across diverse populations and methodologies. This method allows researchers to pool results from various studies investigating genetic, protein, and metabolite biomarkers associated with prediabetes, thereby enhancing statistical power and generalizability.

Systematic reviews systematically gather and assess relevant literature to provide a comprehensive overview of existing knowledge. By critically analyzing studies on genetic variations (such as those in genes like TCF7L2, PPARG, and FTO), protein biomarkers (including adiponectin, C-reactive protein (CRP), IGFBP-1, and fetuin-A), and metabolite biomarkers (like glucose, BCAAs, acylcarnitines, and ceramides), systematic reviews highlight key findings and identify gaps in current understanding.

Bioinformatics approaches play a pivotal role in integrating and analyzing largescale omics data sets. These methods utilize computational tools to interpret genomic, proteomic, and metabolomic data, facilitating the identification of molecular pathways and networks involved in prediabetes.

For instance, bioinformatics tools such as pathway analysis (using databases like KEGG) and network analysis (using platforms like GeneMania and GEPHI) help uncover interactions among genes, proteins, and metabolites implicated in insulin resistance and glucose metabolism. Together, meta-analysis, systematic review, and

bioinformatics approaches synergistically contribute to elucidating the complex molecular mechanisms of prediabetes. By integrating evidence across diverse studies and leveraging computational tools, researchers can uncover novel biomarkers, pathways, and therapeutic targets, ultimately advancing our understanding and management of prediabetes and related metabolic disorders.

Chapter 3

Research Methodology

3.1 Elaboration of Pathways

Genetic and molecular components and pathways associated with prediabetic insulin resistance were be investigated through meta-analysis. Meta-analysis refers to the process of systematically assessing the results of previous studies to drive statistically proven conclusions. In the current study, international databases such as PubMed Scopus and local repositories were be searched for research articles and reviews. Original studies were retrieved to potential prognostic role of genes and molecules that are associated with prediabetic insulin resistance, meta-analysis and review articles were also be investigated along with their primary references [189]. The search syntax was designed and filters were used to include human studies. Studies were not excluded based on the ethnicity of the studied population [48], [189]. A standard protocol of meta-analysis was adopted [47].

3.2 Network and Pathway Analysis

Genes identified from Meta-analysis were analyzed through specific tools such as Gene mania or any other to predict the functions of genes and to construct the interaction

networks. Networks associated with the pathophysiology of the disease was identified and metabolic pathways were be retrieved from KEGG [190]. Gene interaction data was retrieved and was used to make networks using software such as Metascape or Gephi; closeness centrality, harmonic centrality, betweenness centrality, and eigncentrality was used by adjusting the threshold depending on the values to find the hub genes. After the identification of the hub genes, enrichment analysis of hub genes was performed through specific tools such as ToppGene. All the hub genes were uploaded on the selected tool for pathway enrichment based on gene ontology (GO) for biological processes, molecular functions, diseases, drugs, pathways, and cellular components [191]. Molecular factors identified from the Meta-analysis were statistically analyzed and filtered for validation. For validation of molecular factors various tools can be used such as Go term finder, David, GO term mapper for gene annotation; Blast, FASTA, MUSCLE for sequence alignment; Swiss model, Fold X, I TASSER, Hpred for homology modeling; Saves, Verify 3D for structure validation; CATH, Pock drug, p2rank for active pocket; Autodock vina, Arguslab, ClusPro for modeling and simulation; PLIP, Struct2net, Prise for protein interactions; Webgivi, David, TopFun for enrichment analysis. The metabolic pathway for prediabetic insulin resistance was predicted, Selection of tools for network and pathways analysis depends upon the nature of the molecular component selected [192].

3.3 Validation of Genetic, Metabolic Components

A comprehensive methodology was employed to validate and analyze predicted protein structures and interactions implicated in prediabetic insulin resistance. Initially, protein sequences from two sets—set 1, identified manually through literature review, and set 2, interacting genes identified through network analysis—were retrieved from the UniProt Knowledgebase in FASTA format. These sequences were subjected to BLASTp analysis against the Protein Data Bank (PDB) to identify structurally similar proteins, ensuring a robust starting point for structural modeling. Following sequence alignment, the predicted protein structures were validated using multiple tools.

The ProSA server evaluated the energy profiles of the modeled proteins to assess their structural integrity and identify potential discrepancies compared to experimentally determined structures. Concurrently, MetaMQAPII and PHYRE2 servers provided additional validation by predicting 3D protein structures and highlighting any structural variations. These validation steps were crucial for ensuring the accuracy and reliability of the modeled protein structures.

Next, protein-protein interactions between set 1 (receptors) and set 2 (ligands) genes were investigated through docking simulations using the GRAMM-X and Cluspro servers. GRAMM-X facilitated the docking process by predicting how proteins interact spatially and assessing binding affinities, while Cluspro validated these interactions through detailed analysis of surface complementarity and model scores.

Finally, Discovery Studio software was utilized for in-depth analysis of the docked protein complexes. This tool allowed for visualization of interaction interfaces, identification of interacting amino acids, and assessment of structural stability through surface analysis. The methodology integrated these computational and analytical approaches to elucidate the molecular mechanisms underlying prediabetic insulin resistance, potentially identifying novel biomarkers and therapeutic targets for further study and clinical application.

3.4 Statistical Analysis

Statistical analysis was performed via SPSS for the identification of novel genetic and metabolic components associated with prediabetes for use in clinical practice [176]. A student t-test and Z-test was used to examine the statistically significant differences in genetic and molecular biomarkers such as serum glucose, total cholesterol, triglycerides, etc. Pearson's chi-square test was used to examine the association of mutations detected with prediabetes. Multivariate logistic regression analysis was performed to determine the association of molecular and genetic biomarkers with prediabetes [176]. Overall overview of methodology adopted to achieve objective 1 to objective 3 is presented in figure 3.1.

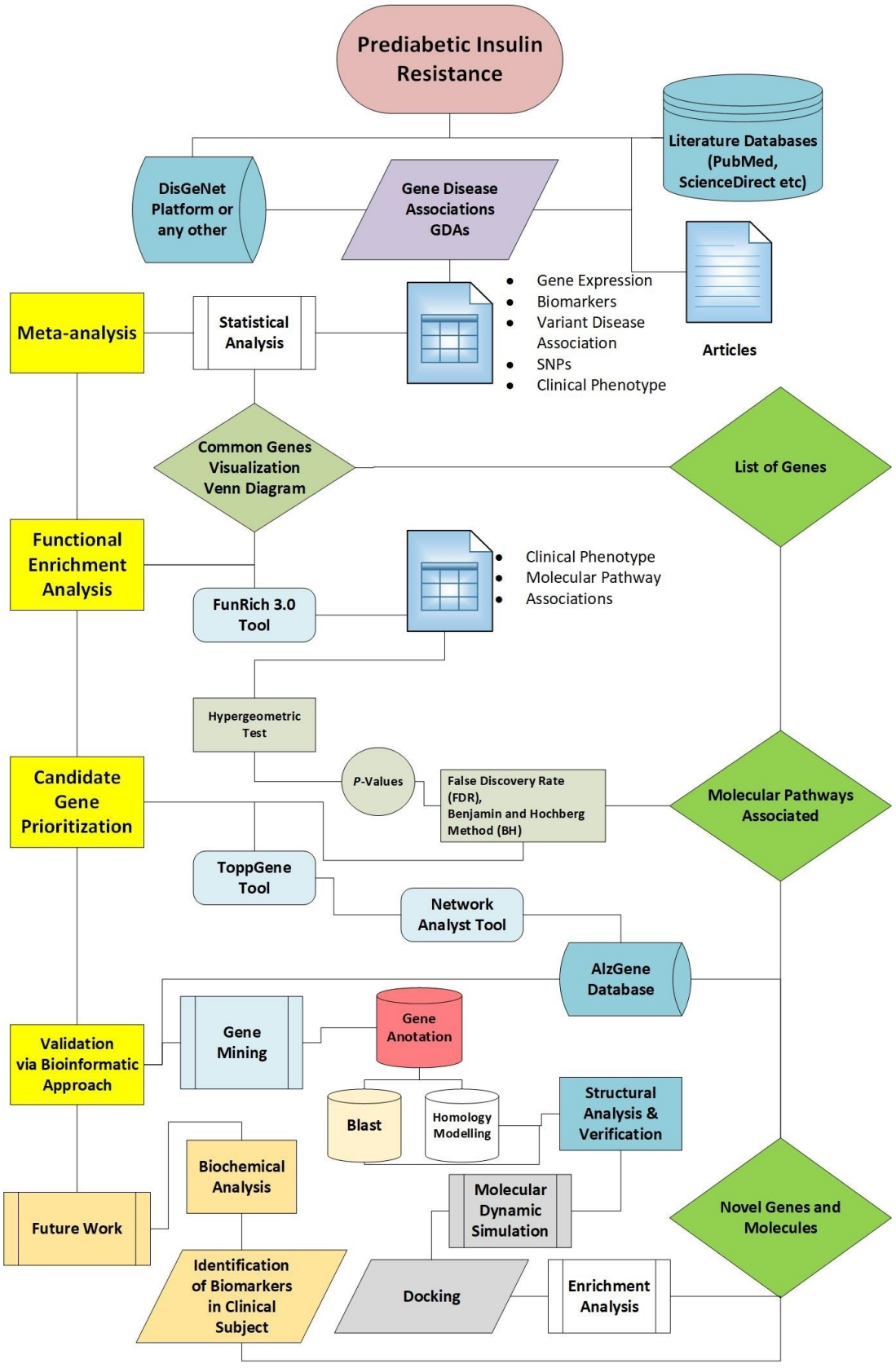


FIGURE 3.1: Overview of methodology for objective 1 to objective 3

3.5 Methodology Breakdown to Achieve each Objective

3.5.1 Objective 1 (RQ1-RQ3)

3.5.1.1 Research Question

Determining the research question is the first and most important step of starting a meta-analysis. It provides a proper target to conduct the meta-analysis on and helps in searching relevant material. The question should not very broad, as the resulting articles and workload would be vast and unmanageable. It is preferred for the question to be more narrowed down to ensure more precise and accurate results, as well as relevant data and manageable workload.

The research questions in this study were:

1. Genomic, proteomic and metabolomic biomarkers associated with prediabetes assessed through high throughput omic techniques.
2. Identification of key genomic, proteomic and metabolomic biomarkers associated with prediabetes.
3. Interaction of key omic biomarkers in disease progression based on pathway and gene interaction.

3.5.1.2 Literature Search

Independent searching was performed through scientific databases. The major data -bases include PubMed and Google Scholar that was employed.

The main searching strategy employed was keyword searching, keywords such as "Omics in Prediabetes", "Prediabetes", "Omics", "Genomics of Prediabetes", "Metabolomics of Prediabetes", "Meta-analysis of Prediabetes" was searched. Searching

based on the key words is known as grey literature. Our studies will include data of last 10 years, from 2012 to 2022.

3.5.1.3 Selection Criteria

The following parameters was considered for selection criteria.

Inclusion Criteria

- Original research articles
- Time period: 2012-2022 (last decade)
- Articles reporting Prediabetes with high-throughput –omic techniques Research articles with valid clinical data
- Research articles published in reputed science journals

Exclusion Criteria

- Review articles
- Articles published before 2012
- Articles reporting other diseases
- Research articles published in non-reputed science journals

3.5.1.4 Research Articles

A comprehensive review of numerous research articles was undertaken, resulting in the acquisition of a wide array of diverse data. Rigorous selection criteria were meticulously applied to all the articles evaluated to ensure the relevance and quality of the studies included in the analysis. Based on these criteria, certain articles were systematically included or excluded from the final results, thereby refining the scope of the investigation.

To organize the gathered information, all pertinent data from the selected articles were initially compiled into an Excel spreadsheet, designated as the Source Table. This Source Table encompassed heterogeneous data types, primarily in qualitative form, capturing various dimensions such as study methodologies, participant demographics, biomarkers assessed, and key findings.

The compilation process involved careful extraction and categorization of relevant information to maintain consistency and facilitate subsequent analysis. Once the data were thoroughly aggregated, appropriate statistical methods were employed to interpret the findings, enabling the identification of significant trends and correlations within the dataset.

This structured approach ensured a robust synthesis of the existing literature, providing a solid foundation for drawing informed conclusions about the biomarkers associated with diabetes and prediabetes. The methodological rigor applied throughout the process enhanced the reliability and validity of the study's outcomes, contributing valuable insights to the field of metabolic disorder research.

3.5.1.5 Source Table

After studying and reviewing research articles, an excel sheet was prepared containing all the important information from all the articles. The information in this sheet will include the title, author(s), publication year, journal, and abstract. It was called the Source Table, or Master Table. For further proceedings, this table was of utmost importance.

3.5.1.6 Qualitative Data

Qualitative data analysis does not include quantification of data. Data is in the form of a text, which is used for data analysis and its explanation [48]. The initial data that was obtained from the Source Table constructed after literature searching and the research articles obtained was qualitative in nature. This qualitative data was su-

bjected to a variety of statistical methods for proper statistical analysis.

3.5.1.7 Statistical Analysis

To obtain quantitative data from the qualitative data obtained from the Source Table, a variety of statistical methods need to be applied for statistical analysis. The selection of appropriate methods depends on the type of data that is initially obtained. Since different methods work for different types of data in different researches, there is no one-size-fits-all method that can be used. The selection of an appropriate method was done once the qualitative data has been obtained.

3.5.1.8 Quantitative Data

Quantitative data analysis involves quantification of data. It includes statistical analysis of data shows results in form of relative frequency. Once the qualitative data has undergone statistical analysis, the resulting data that is obtained is the quantitative data. Quantitative data provides reliable information and is more definite and accurate as compared to qualitative data. Quantitative data is important as further resulting plots and charts such as PRISMA charts and Forest plots was made based on this data.

3.5.1.9 Heterogeneity

Heterogeneity of data in meta-analyses refers to the variation in all the study outcomes among different studies. Heterogeneous data essentially means data with high variability. Heterogeneity in meta-analyses is defined by two major concepts: The Fixed-Effect Model and The Random-Effects Model. Despite the fact that both models utilize similar formula sets, and sometimes even yield similar estimates does not mean that these models are interchangeable. The Fixed-Effect Model suggests that there is a common effect size of all the studies. According to this model, all the studies included in the meta-analysis have enough uniformity for a common conclusion to be

drawn from all the results [49]. On the other hand, the Random-Effects Model suggests that there is a different effect size of all the studies. According to this model, the outcomes may show variation from study to study. A conclusion can still be drawn despite it [50].

3.5.1.10 PRISMA Chart

For compilation of the results, the PRISMA statement was utilized. The abbreviation PRISMA stands for "Preferred Reporting Items for Systematic Reviews and Meta-Analyses". The PRISMA chart consists of a 27-item checklist as well as a four-phase flowchart. The items that are included in the checklist are those that are considered significant for transparent meta-analyses reporting. PRISMA can provide an accurate and reliable dataset [51].

3.5.1.11 Forest Plot

For proper graphical representation of data, Forest plots are used often in meta-analyses. Forest plots are significant as they are useful for quick and efficient scanning and interpretation of all the data. Forest plots are commonly referred to as the "total summary effect of the meta-analyses." Forest plots have been proven to be quite helpful in exploring the possible causes of heterogeneity. Forest plots can be modified for the study type as well, such as for diagnostic accuracy studies [52]. Graphical presentation of methodology for objective 1 is illustrated in 3.2

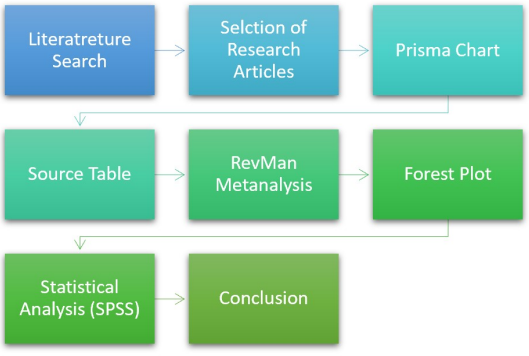


FIGURE 3.2: Graphical presentation of methodology for objective 1

3.5.2 Objective 2 (RQ4-RQ6).

3.5.2.1 Data Collection

The top-down bioinformatics approach was used, core genes that are associated with insulin resistance and metabolic syndrome were primarily retrieved from literature i.e., PubMed (https://pubmed.ncbi.nlm.nih.gov/) as well as from pathways that are available in databases such as KEGG (https://www.genome.jp/ kegg/pathway .html).

3.5.2.2 Gene Interaction Networks

Gene interaction network with other genes and the function of genes were predicted through the online tool GeneMANIA (http://genemania.org/) [193]. Nodes and edges of interacted genes were further processed in Excel to remove duplication.

3.5.2.3 GGI Network Construction and Identification of Hub Genes

The Excel file consisting of nodes and edges was uploaded to Gephi 0.9.2 (https: //gephi .org/) [194] and gene regulatory networks and interaction data were retrieved to find hub genes associated with insulin resistance. Data obtained from Gephi was analyzed values of closeness centrality [195], harmonic centrality [196], betweenness centrality [197], and eigncentrality were compared by adjusting threshold values to identify hub genes of every network.

3.5.2.4 Functional Annotation and Pathway Enrichment Analysis

Enrichment analysis of hub genes was performed through the ToppGene tool (https:// toppgene. cchmc.org/) [198]. Hub genes were uploaded to the ToppGene tool for pathway enrichment based on gene ontology (GO) for biological processes, molecular function, and cellular components. This analysis helps identify key biological pathways and molecular mechanisms associated with the hub genes, providing insights into their functional roles [199].

3.5.2.5 Validation of Hub Genes

Finally, hub genes identified were further validated through previously published literature to investigate whether these genes are involved in inflammation directly or through any metabolic disease that causes inflammation. Methodology overview for objective 2 is presented in figure 3.3

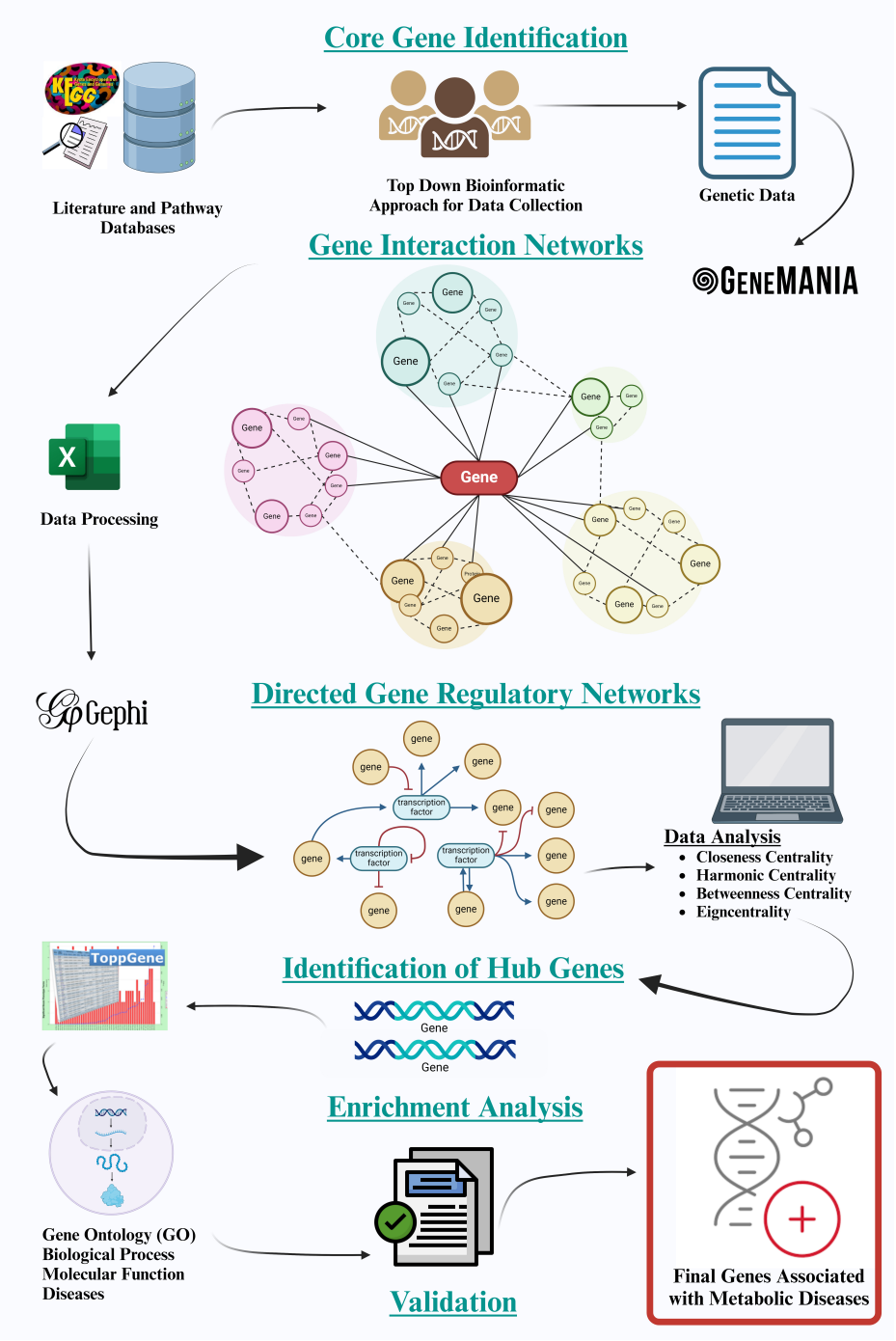


FIGURE 3.3: Methodology overview for objective 2

3.5.3 Objective 3 (RQ6-RQ9)

3.5.3.1 Identification of Pre-Diabetic Insulin Resistance Genes via Text Mining

In a comprehensive manual literature review, 25 genes implicated in prediabetic insulin resistance were identified. These genes, which were found to play crucial roles in the pathophysiology of insulin resistance in humans, were extracted through text mining methodologies and designated as set 1.

3.5.3.2 Identification of Novel Biomarkers Linked to Known Genes via FunCoup 2.0 Gene Interaction Networks

Gene interaction networks were constructed for the genes identified through literature review using the FunCoup 2.0 database (http://FunCoup.sbc.su.se). The analysis revealed numerous additional genes interacting with those initially selected. FunCoup, a bioinformatics tool for generating global gene and protein interaction networks, facilitated this process.

The newly identified genes, including ENO2, PRDX1, ALDH2, PYGB, MDH2, ACTB, TUBB, EEF2, PGK1, RACK1, PGD, PGM1, VCP, HSPA8, EEF1A1, HSP90AB1, TUBB4B, ENO1, TALDO1, PSMD2, MDH2, ACSS2, VDAC2, HSPAL, SLC25A5, HSPD1, ATTP51B, and GPI, were noted as novel genes. Consequently, two sets of genes were established: the first set obtained through manual literature review and the second set comprising genes identified through interaction with the initial list. These biomarkers identified through gene interaction networks were designated as set 2.

3.5.3.3 Enrichment Analysis of Identified Gene Sets

Enrichment analysis was performed on the genes in set 1 and set 2. This analysis provided various outcomes, including molecular, cellular, and biological annotations

of the gene sets. Additionally, tissue specificity of the selected genes was assessed through enrichment analysis. The analysis involved measuring functional relationships between experimentally derived genes or proteins and a database of gene-protein sets within large-scale genomic data. An over-representation-based enrichment analysis was employed for its numerous advantages:

1. it tests the functional overlap of gene/protein combinations,
2. it accounts for genes with missing annotations,
3. it considers the physical interaction network structure among the gene/protein sets, and
4. it can establish tissue-specific gene/protein interactions [200].

3.5.3.4 Retrieval of Protein Sequences for Gene Sets 1 and 2

Protein sequences for genes in set 1 (selected manually) and set 2 (interacting with set 1 genes) were retrieved from the UniProt Knowledgebase in FASTA (canonical) format. These sequences were then compiled into a Word document. UniProt, or Universal Protein Resource, provides a comprehensive, stable, and easily accessible resource for protein sequences and functional annotations [201]. The UniProt database is produced by three leading bioinformatics institutions and is composed of four distinct components: the UniProt Archive, the UniProt Knowledgebase, the UniProt Reference Clusters, and the UniProt Metagenomic and Environmental Sequence Database [202].

3.5.3.5 Gene Annotation Using STRAP Tool

Both set 1 and set 2 genes were annotated using the Software Tool for Researching Annotation of Proteins (STRAP). UniProt identifiers were utilized to conduct enrichment analysis, generating results within approximately 2-3 hours, depending on the number of genes processed. The UniProt IDs corresponding to all proteins in set 1

and set 2 were recorded. These UniProt IDs were then inputted into the STRAP tool to obtain protein annotations presented in table format, along with Gene Ontology (GO) charts depicting proteomic data [203].

3.5.3.6 Identification of Similar Protein Structures in PDB Database Using BLAST Sequence Alignment

BLASTp was employed to compare protein sequences obtained from UniProt against the Protein Data Bank (PDB) database. The protein sequences were inputted in FASTA format into the BLASTp query sequence box to identify similarities with existing protein sequences from various species in the database [204]. BLAST (Basic Local Alignment Search Tool) is a program used for comparing biological sequences such as proteins, amino acids, and nucleotides, and it accesses a comprehensive database containing a vast number of protein and nucleotide sequences. This comparison allows for the identification and characterization of the subject sequence based on similarities with known sequences [205].

3.5.3.7 Prediction of Protein Domains and Structures

Proteins identified from the BLASTP results were further analyzed in the Protein Data Bank to identify comparable three-dimensional structures. Subsequently, the Robetta server was utilized to predict potential domains within each selected protein sequence. The Robetta server provides predictions of protein domains using both ab initio modeling and comparative modeling approaches, offering insights into the structural organization of the identified protein domains [206].

3.5.3.8 Validation and Analysis of Predicted Protein Structures

Validation of Predicted Domains

To ensure the reliability of the modeled protein structures, the ProSA server was employed. ProSA evaluates the energy profiles of protein models, providing Z scores

that indicate their deviation from experimentally determined structures. This server is instrumental in identifying potential errors in theoretical and experimental protein structures by comparing them against a database of known protein structures. For this study, protein sequences retrieved in PDB format were individually subjected to ProSA analysis to validate their structural models and assess their conformity to expected energy profiles [207].

Cross-Checking Model Validation

Further validation of the protein structures was conducted using MetaMQAPII and PHYRE2 servers. These tools assess protein 3D structures by analyzing PDB files submitted through their respective interfaces. MetaMQAPII and PHYRE2 utilize advanced algorithms to predict and verify protein structures, providing validation results typically within a few minutes to hours, depending on the complexity of the protein and server workload. Results are delivered via email, detailing structural predictions and potential discrepancies between predicted and experimental structures.

Protein-Protein Docking

To investigate potential interactions between proteins encoded by set 1 (used as receptors) and set 2 (used as ligands), protein-protein docking simulations were performed using the GRAMM-X web server [208]. Additionally, the Cluspro server was utilized to validate docking results by assessing the surface complementarity and model scores of the predicted protein complexes. Cluspro provides multiple models of docked protein structures, allowing for comprehensive analysis of potential interaction sites and confirmation of docking accuracy [209], [210].

3.5.3.9 Analysis of Docking Results and Identification of Interacting Amino Acids

For detailed analysis of docking results and identification of interacting amino acids within docked complexes, Discovery Studio software was utilized. Developed by Dassault Systèmes BIOVIA, Discovery Studio offers a suite of tools for molecular modeling

and simulation, including visualization of protein structures and analysis of molecular interactions. In this study, Discovery Studio was employed to scrutinize the spatial arrangement of docked protein complexes, analyze the interaction interfaces, and validate the accuracy of predicted protein-protein interactions through surface analysis and structural comparison [211]. This comprehensive analysis provides insights into the molecular mechanisms underlying protein interactions implicated in prediabetic insulin resistance, potentially identifying novel biomarkers and therapeutic targets [211].

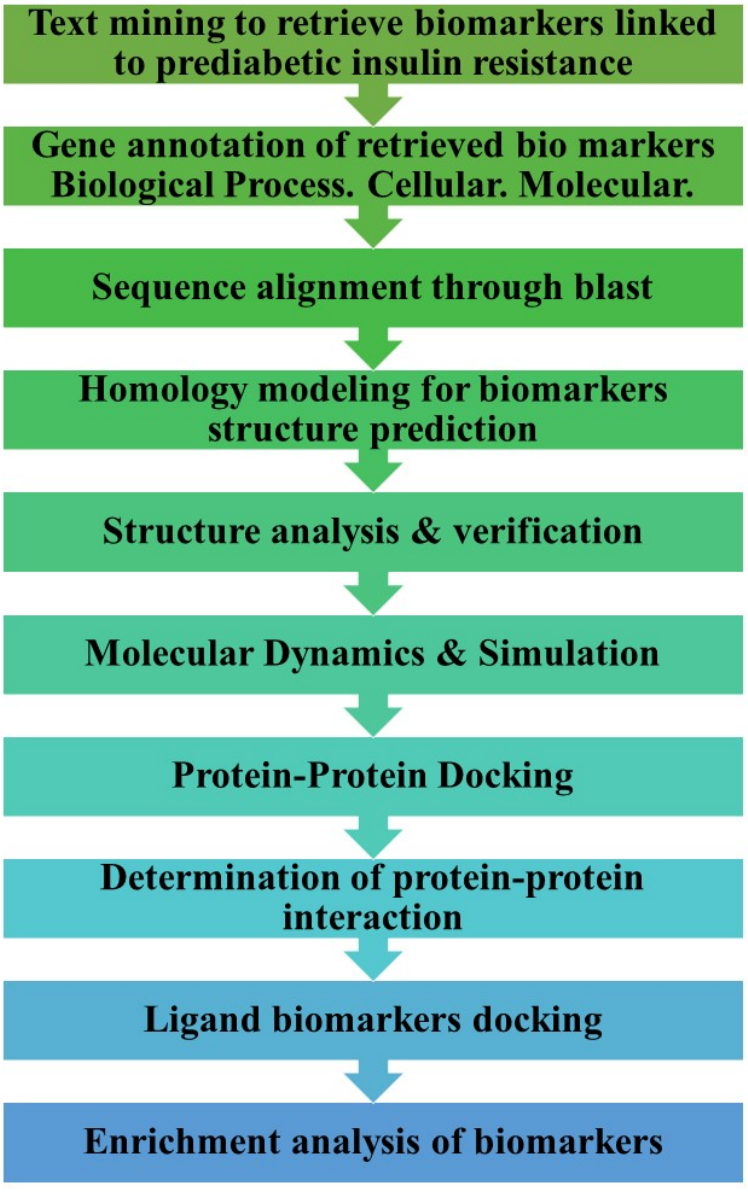


FIGURE 3.4: Methodology overview for objective 3

3.6 Limitations, Potential Biases, and Data Verification

While this study integrated a robust meta-analysis and advanced in-silico methodologies to identify potential biomarkers for prediabetic insulin resistance, certain limitations and biases needed to be considered. The effectiveness of the meta-analysis depended on the quality and consistency of the included studies. Variability in study populations, methodologies, and reporting standards across different sources may have introduced selection bias and affected the generalizability of the findings. Furthermore, computational approaches, including gene interaction analysis, enrichment studies, and molecular docking, relied on existing biological databases, which had inherent limitations due to incomplete or evolving datasets. The accuracy of protein structure predictions and interaction models was also influenced by the resolution and reliability of available structural data, potentially impacting the interpretation of biomolecular interactions.

To ensure the reliability of results, multiple verification steps were implemented throughout the study. Homology assessments were performed using BLASTP, while structural validation of modeled proteins was conducted through the ProSA server, ensuring the integrity of predicted structures. Additionally, docking analyses were cross-validated using independent platforms such as GRAMM-X and ClusPro to minimize computational biases. Despite these rigorous verification measures, experimental validation remained essential to confirm the clinical relevance of the identified biomarkers. Future research should focus on functional studies and clinical trials to bridge the gap between computational predictions and practical medical applications, ensuring these findings contribute effectively to early disease detection and management.

Chapter 4

Results

4.1 Meta-analysis (RQ1-RQ3)

4.1.1 PRISMA Chart

For the compilation of the results, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement was employed to ensure a methodical and transparent reporting process. PRISMA is a widely accepted framework designed to enhance the quality of systematic reviews and meta-analyses.

The framework includes a 27-item checklist and a four-phase flowchart, both of which are essential for transparent and comprehensive reporting. The checklist encompasses various items deemed crucial for the thorough reporting of systematic reviews and meta-analyses. These items ensure that the methodology, results, and discussions are presented with utmost clarity and completeness, thus facilitating reproducibility and critical appraisal by other researchers.

The flowchart visually represents the selection process of studies included in the review, encompassing identification, screening, eligibility, and inclusion phases. PRISMA structured approach is integral in providing an accurate and reliable dataset for the systematic review. Initially, 23 studies were selected and incorporated into the source

table. These studies were obtained from various databases and additional sources, ensuring a comprehensive collection of relevant literature (figure 4.1).

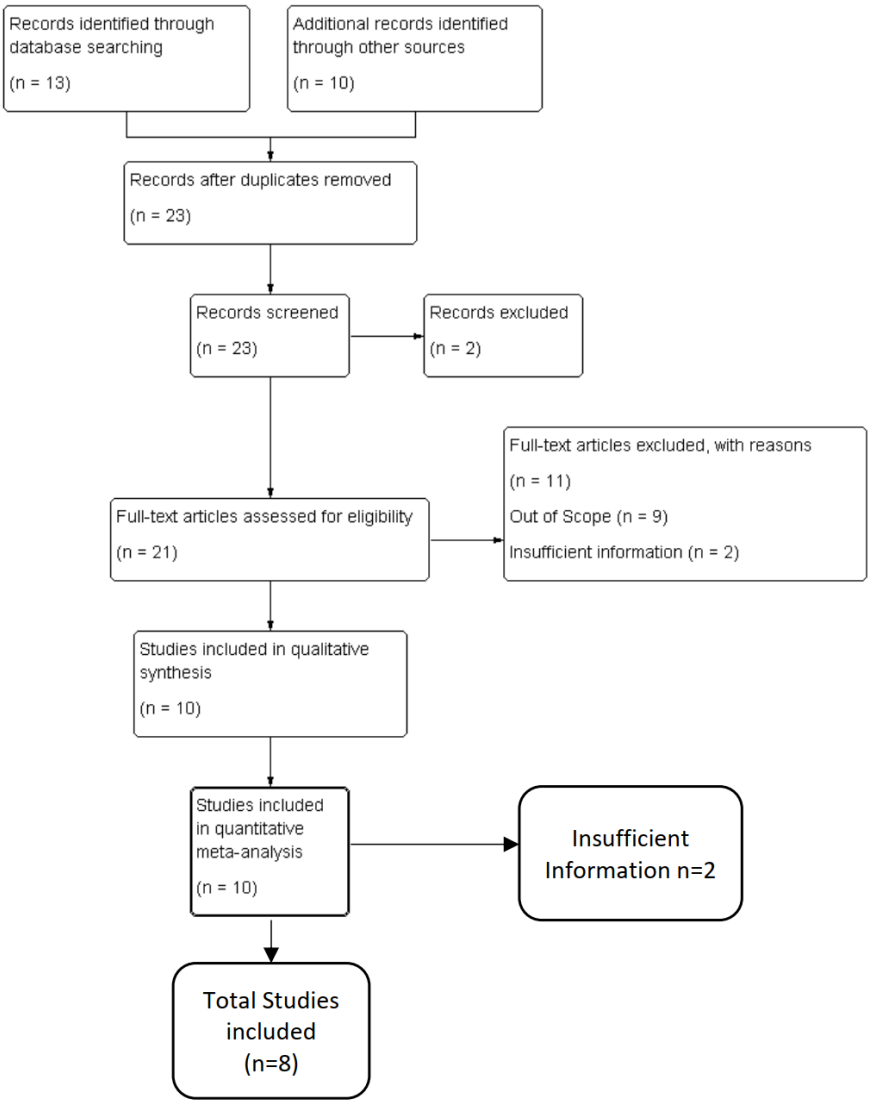


FIGURE 4.1: PRISMA chart

However, 2 articles were excluded from the analysis due to their publication dates falling outside the predefined timeline. Consequently, 21 studies remained and were subjected to a rigorous screening process to determine their eligibility. During the screening phase, 9 studies were excluded because they did not align with the scope of the review. Specifically, these studies did not adequately address the focus of the research, which was centered on genomic, proteomic, and metabolomic factors related

to prediabetes (figure 4.1). Furthermore, 2 additional studies were excluded due to insufficient information regarding these specific biological mechanisms underlying prediabetes. As a result of this meticulous screening process, the final dataset comprised studies that met all inclusion criteria and provided substantial information on the genomic, proteomic, and metabolomic aspects of prediabetes. This ensured that the review was grounded in robust and relevant data, thereby enhancing the reliability and validity of the findings.

4.1.2 Study Characteristics and Source Table

The source table 4.1 for the meta-analysis presents a comprehensive overview of several pivotal studies focused on prediabetes, summarizing essential information for each study including the title, first author and year of publication, ethnicity of participants (figure 4.2, number of participants, number of prediabetic and normal individuals, and the specific variants investigated. This detailed compilation facilitates a robust meta-analysis by ensuring all relevant studies are systematically compared and contrasted.

TABLE 4.1: Number of participants from each ethencity included in meta-anlysis

Ethencity		**Health Status**			**Total**
		Predi-abetic	**Normal**	**Left Study**	
White USA	Count	2057	1289	5	3351
	% within Ethencity	61.4%	38.5%	0.1%	100.0%
Spain	Count	12	21	31	64
	% within Ethencity	18.8%	32.8%	48.4%	100.0%
Germany	Count	132	208	0	340
	% within Ethencity	38.8%	61.2%	0.0%	100.0%

Australia	Count	105	242	0	347
	% within Ethencity	30.3%	69.7%	0.0%	100.0%
Korea	Count	924	799	0	1723
	% within Ethencity	53.6%	46.4%	0.0%	100.0%
USA	Count	51	46	9	106
	% within Ethencity	48.1%	43.4%	8.5%	100.0%
AA/EA	Count	158	65	0	223
	% within Ethencity	70.9%	29.1%	0.0%	100.0%
Chinese	Count	15017	11001	0	26018
	% within Ethencity	57.7%	42.3%	0.0%	100.0%
Total	Count	18456	13671	45	32172
	% within Ethencity	57.4%	42.5%	0.1%	100.0%

The first study, "A Genome-Wide Association Study of Prediabetes Status Change" by Liu *et al.* (2022) (table 4.2), explores genetic variants associated with prediabetes status changes among a cohort of 3,351 White participants from the USA. Of these participants, 2,057 were classified as prediabetic, while 1,020 were normal. This study employed a genome-wide association study (GWAS) to identify genes and proteins that may play a role in the progression of prediabetes.

In the second study (table 4.2), "Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes," Tulipani *et al.* (2016) conducted metabolomic profiling on a smaller Spanish cohort of 64 participants. The study identified 12 prediabetic and 21 normal individuals, focusing on biomarkers and metabolites linked to morbid obesity and prediabetes. This research highlights the

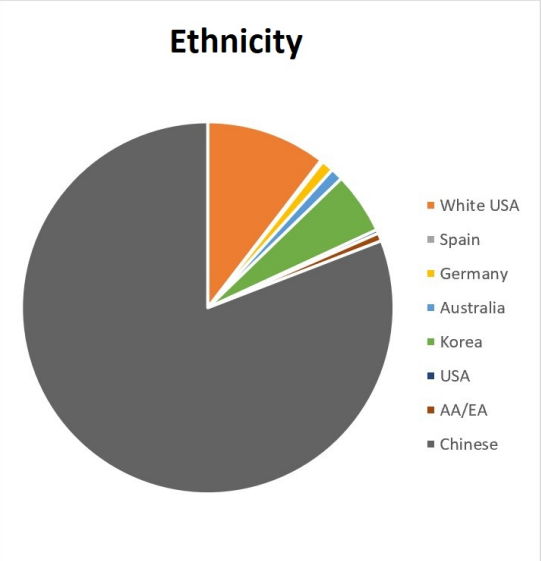


FIGURE 4.2: Pie chart representing total number of participants from each study

metabolic differences between individuals with discordant phenotypes, contributing valuable insights into the metabolic underpinnings of prediabetes.

Li *et al.* (2022), in their study titled "Diagnostic Performance of Sex-Specific Modified Metabolite Patterns in Urine for Screening of Prediabetes," investigated the efficacy of using urine metabolite patterns to screen for prediabetes among 340 German participants. The study included 132 prediabetic and 208 normal individuals, examining how sex-specific metabolite modifications could serve as diagnostic markers for prediabetes (table 4.2). This study underscores the potential for non-invasive diagnostic methods in identifying prediabetes.

The fourth study, "Human SHC-transforming protein 1 and its isoforms p66shc: A novel marker for prediabetes" by Jelinek *et al.* (2021), focused on an Australian cohort of 346 participants, with 105 prediabetic and 242 normal individuals. The researchers identified SHC-transforming protein 1 and its isoforms as novel protein markers for prediabetes, suggesting new avenues for early detection and intervention (table 4.2).

Lee *et al.* (2020), in their study "Identification of metabolic markers predictive of prediabetes in a Korean population," examined a large Korean cohort of 1,723 participants. With 924 prediabetic and 799 normal individuals, this research identified

proteins and metabolites predictive of prediabetes, providing a comprehensive analysis of metabolic markers that could forecast the onset of the condition in this population (table 4.2).

The study titled "Longitudinal multi-omics of host–microbe dynamics in prediabetes" by Zhou *et al.* (2019) analyzed the dynamic interactions between host and microbial factors in prediabetes over time. The study included 106 participants from the USA, with 51 prediabetic and 46 normal individuals. This longitudinal study integrated multi-omics approaches, examining genes, proteins, and metabolites to understand the complex biological interactions contributing to prediabetes (table 4.2).

Dagogo-Jack *et al.* (2020), in their "Pathobiology and Reversibility of Prediabetes in a Biracial Cohort (PROP-ABC) Study: design of lifestyle intervention," explored the impact of lifestyle interventions on the reversibility of prediabetes in a biracial cohort. The study included 223 participants, comprising 158 prediabetic and 65 normal individuals from African-American and European-American backgrounds. This research emphasized the potential for lifestyle modifications to reverse prediabetes and provided insights into the pathobiology of the condition across different ethnic groups (table 4.2).

Finally, Sun *et al.* (2022), in "The Association Between the Triglyceride-to-High-Density Lipoprotein Cholesterol Ratio and the Risk of Progression to Diabetes From Prediabetes: A 5-year Cohort Study in Chinese Adults," examined a large Chinese cohort of 26,018 participants. The study identified 15,017 prediabetic and 11,001 normal individuals, focusing on the ratio of triglycerides to high-density lipoprotein cholesterol as a predictive marker for the progression from prediabetes to diabetes over five years. This study highlighted significant metabolic predictors of diabetes progression in a large and diverse population (table 4.2). Overall, the source table encapsulates a wide range of studies that collectively enhance our understanding of the genetic, proteomic, and metabolic factors associated with prediabetes (figures 4.3, 4.4). By systematically comparing these studies, the meta-analysis aims to identify consistent patterns and markers that could improve early detection and intervention strategies for prediabetes.

Table 4.2: Source table and characteristics of studies included

Title	**First Author, Year**	**Ethnicity**	**Part.**	**Prediabetic**	**Normal**	**Variants**
A Genome-Wide Association Study of Prediabetes Status Change Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes	Liu *et al.*, 2022	White USA	3351	2057	1020	Genes/ Proteins
	Tulipani *et al.*, 2016	Spain	64	12	21	Biomarkers/ Metabolites

Diagnostic Performance of Sex-Specific Modified Metabolite Patterns in Urine for Screening of Prediabetes	Li *et al.*, 2022	Germany	340	132	208	Biomarkers/ Metabolites
Human SHC-transforming protein 1 and its isoforms p66shc: A novel marker for prediabetes	Jelinek *et al.*, 2021	Australia	346	105	242	Proteins
Identification of metabolic markers predictive of prediabetes in a Korean population	Lee *et al.*, 2020	Korea	1723	924	799	Proteins/ Metabolites

Longitudinal multi-omics of host–microbe dynamics in prediabetes	Zhou *et al.*, 2019	USA	106	51	46	Genes/ Proteins/ Metabolites
Pathobiology and Reversibility of Prediabetes in a Biracial Cohort (PROP-ABC) Study: design of lifestyle intervention	Dagogo-Jack *et al.*, 2020	African-Americans and European - Americans	223	158	65	Lifstyle interventions/ Others

The Association Between the Triglyceride-to-High-Density Lipoprotein Cholesterol Ratio and the Risk of Progression to Diabetes From Prediabetes: A 5-year Cohort Study in Chinese Adults	Sun *et al.*, 2022	Chinese	26018	15017	11001	Metabolites

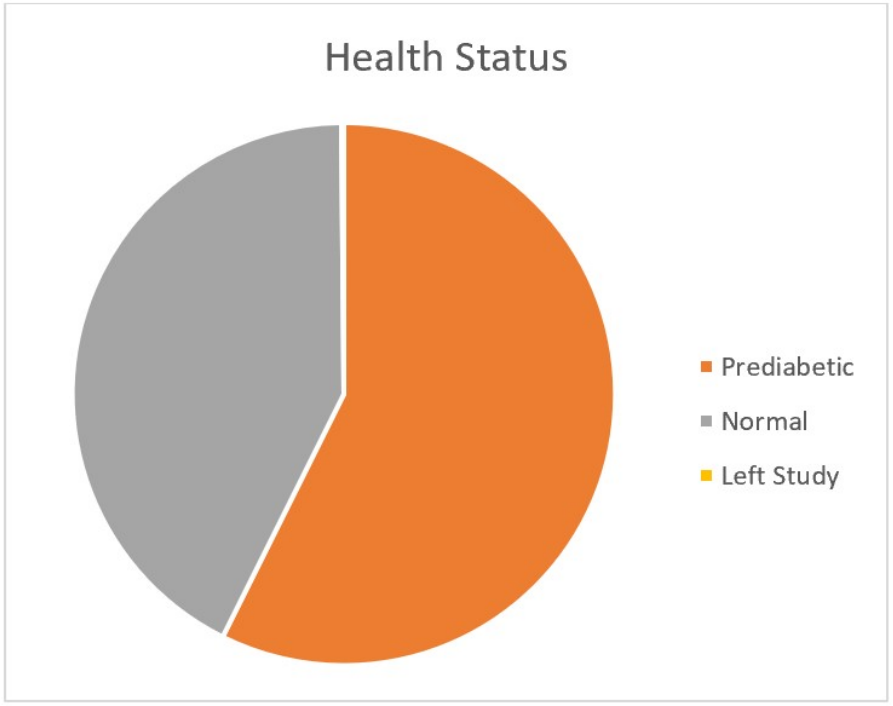


FIGURE 4.3: Health status of participants

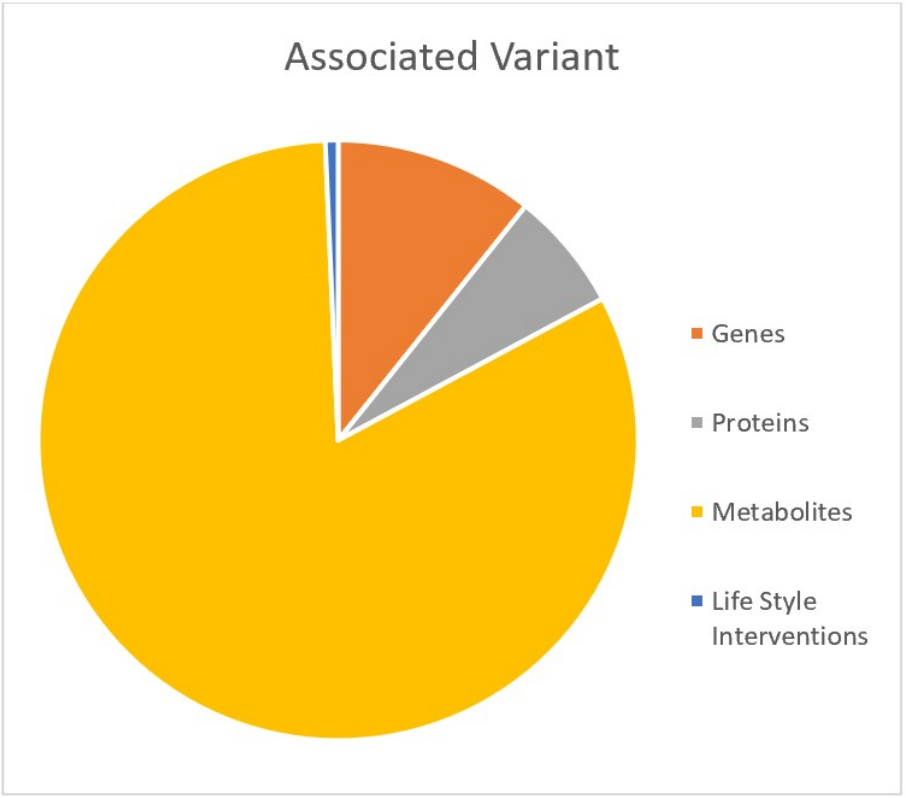


FIGURE 4.4: Association of prediabetes with genes, proteins and metabolites

4.1.3 Statistical Analysis and Forest Plot

The statistical analysis for this meta-analysis was conducted following a meticulous selection process, as outlined by the PRISMA flow diagram (figure 4.1). Initially, ten studies were identified manually through the PRISMA process. However, two studies were excluded because they did not meet the inclusion criteria, specifically as they were review articles rather than original research studies. This left a total of eight studies for the final analysis.

Data extraction from the selected studies included crucial information such as the study author's name, year of publication, total number of participants, the number of prediabetic cases, and the number of normal individuals without the disease. The odds ratio (OR) and the natural logarithm of the odds ratio (log OR) are essential for conducting a meta-analysis using REVMAN 5 software. However, the odds ratios could not be directly calculated from the filtered data extracted from the research articles.

To address this issue, we converted the type of variables from quantitative to qualitative, specifically from discrete to nominal categories. This conversion allowed us to apply appropriate statistical tests. Binary Logistic Regression was performed using SPSS to calculate the odds ratio, p-values, and standard error for all ethnicities and for each study included in the meta-analysis.

For further accuracy and verification, the probabilities and log odds ratio values were calculated using an online tool available at http://vassarstats.net/tabs˘odds.html. This online tool facilitated the accurate calculation of these statistical parameters, which were then included in the meta-analysis.

Subsequently, the studies were imported into REVMAN 5 software for the meta-analysis. REVMAN 5 is specifically designed for conducting systematic reviews and meta-analyses, providing robust tools for statistical analysis and visualization. The software was used to compile the extracted data, calculate the combined odds ratios, and generate the forest plot.

The forest plot obtained from REVMAN 5 visually represents the odds ratios from each study, along with their confidence intervals. This plot provides a clear overview of the variability and overall effect size across the included studies. The calculated odds ratios for each study are shown alongside a summary measure that indicates the combined effect size across all studies.

Overall, this detailed statistical analysis, conducted through a combination of SPSS for logistic regression and REVMAN 5 for meta-analysis, ensures a rigorous and comprehensive evaluation of the data. The forest plot (figure 4.5) effectively summarizes the findings, highlighting the consistency and magnitude of the association between the identified variables and prediabetes across different studies and ethnicities.

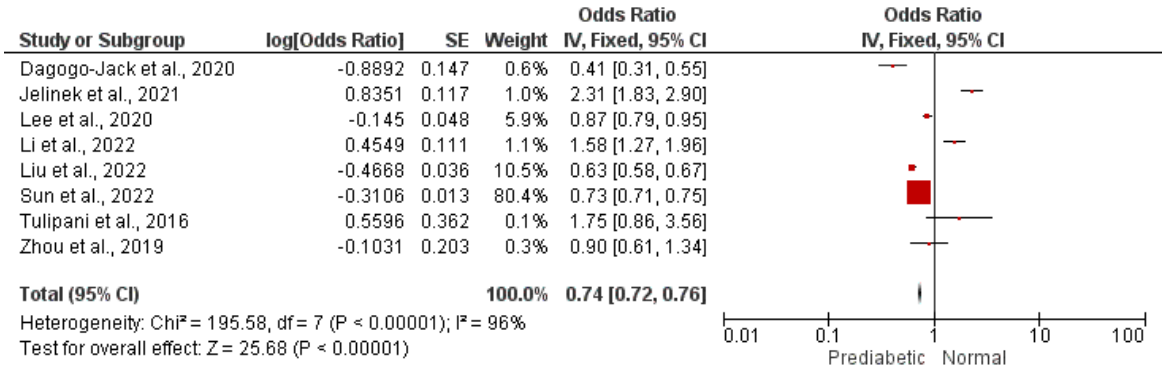


FIGURE 4.5: Forest plot

4.1.4 Odds Ratio and Standard Error Calculation Using Binary Logistic Regression in SPSS for Each Study

To understand the association between various factors and the likelihood of prediabetes across different ethnic groups, binary logistic regression was applied to each study's data using SPSS. This method enabled the calculation of odds ratios (ORs) and their standard errors (SEs), providing a measure of the strength and precision of these associations (table 4.3).

The data from each study were meticulously compiled, including the number of prediabetic and normal participants. Variables were classified into dependent (prediabetes status: prediabetic or normal) and independent (various risk factors) categories.

Where necessary, quantitative variables were converted to qualitative (nominal) variables to facilitate logistic regression analysis. In SPSS, the Analyze menu was used to navigate to Regression and then Binary Logistic Regression, with the prediabetes status set as the dependent variable and independent variables including demographic and clinical factors pertinent to each study.

Logistic regression was run separately for each study to calculate the odds ratio and its standard error. The Enter method was employed to include all selected independent variables in the model. The output provided the coefficients (B), standard errors (SE), Wald statistics, significance levels (p-values), and odds ratios (exp(B)). The odds ratio indicates the likelihood of prediabetes given the presence of specific risk factors compared to their absence, while the standard error provides an estimate of the variability or uncertainty of the odds ratio.

For Liu *et al.* (2022), focusing on White USA participants, odds ratios were calculated for various genetic markers associated with prediabetes status change, with standard errors indicating the precision of these estimates. In the study by Tulipani *et al.* (2016) involving Spanish participants, odds ratios were determined for biomarkers and metabolites indicative of prediabetes, with quantified standard errors reflecting the variability in the estimates. Li *et al.* (2022) assessed odds ratios for sex-specific urine metabolite patterns linked to prediabetes in a German cohort, with standard errors calculated to evaluate the reliability of these associations (table 4.3).

Jelinek *et al.* (2021) established odds ratios for SHC-transforming protein 1 and its isoforms as markers for prediabetes in an Australian population, with computed standard errors showing the degree of uncertainty. Lee *et al.* (2020) evaluated odds ratios for metabolic markers predictive of prediabetes in a Korean cohort, including standard errors to demonstrate the extent of variability. Zhou *et al.* (2019) calculated odds ratios for multi-omic interactions affecting prediabetes in a USA cohort, with provided standard errors indicating precision levels (table 4.3).

In the study by Dagogo-Jack *et al.* (2020), odds ratios were determined for lifestyle interventions impacting prediabetes in African-Americans and European-Americans,

with quantified standard errors reflecting variability. Finally, Sun *et al.* (2022) calculated odds ratios for the triglyceride-to-HDL cholesterol ratio predicting diabetes progression from prediabetes in a large Chinese cohort, with standard errors provided alongside odds ratios to indicate the precision of the estimates (table 4.3).

The application of binary logistic regression in SPSS enabled the calculation of odds ratios and their standard errors for each study, allowing for a robust analysis of the association between various risk factors and prediabetes across diverse ethnic groups. The odds ratios provide insight into the likelihood of developing prediabetes given specific conditions, while the standard errors indicate the reliability and variability of these estimates. This detailed statistical analysis forms the basis for the meta-analysis conducted using REVMAN 5 and the subsequent generation of the forest plot (figure 4.3), offering a comprehensive understanding of prediabetes across different populations (table 4.3).

TABLE 4.3: Odds ratio and standard error calculation using binary logistic regression in SPSS for each study

First Author, Year	**Ethnicity**	**Participants**	**Prediabetic**	**Normal**	**OR**	**Log of Odds**	**S.E**	**P-Va.**
Dagogo-Jack *et al.*, 2020	African-Americans & European-Americans	223	158	65	0.411	-0.892	0.147	0
Jelinek *et al.*, 2021	Australia	346	105	242	2.305	0.835	0.117	0
Lee *et al.*, 2020	Korea	1723	924	799	0.865	-0.145	0.048	0.003

Li *et al.*, 2022	Germany	340	132	208	1.576	0.4549	0.111	0
Liu *et al.*, 2022	White USA	3351	2057	1020	0.627	-0.4668	0.036	0
Sun *et al.*, 2022	Chinese	260 18	150 17	110 01	0.733	-0.3106	0.013	0
Tulipani *et al.*, 2016	Spain	64	12	21	1.75	0.5596	0.362	0.122
Zhou et al., 2019	USA	106	51	46	0.902	-0.1031	0.203	0.612

4.1.5 Overall Results Obtained from Meta-analysis

A detailed meta-analysis has highlighted a broad spectrum of genes and proteins that exhibit dysregulation in both diabetes and prediabetes.

This analysis revealed that the genes SGCZ, HPSE2, ADGRA1, GLB1L3, PCSK6, SIRT1, PPAR, PGC1 alpha, NRF1, TG gene, and HBA1c play significant roles in these metabolic disorders.

4.1.6 Genes and their Roles

1. SGCZ (Sarcoglycan Zeta): Involved in muscle integrity and function, its dysregulation is linked to muscular dystrophies and may impact insulin signaling pathways.

2. HPSE2 (Heparanase 2): Associated with extracellular matrix remodeling, its alteration can affect cellular metabolism and glucose homeostasis.

3. ADGRA1 (Adhesion G Protein-Coupled Receptor A1): Plays a role in cell adhesion and signaling, impacting inflammatory responses in diabetes.

4. GLB1L3 (Galactosidase Beta 1 Like 3): Involved in carbohydrate metabolism, its dysregulation can disrupt glucose breakdown.

5. PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6): Regulates the activation of precursor proteins, influencing metabolic pathways.

6. SIRT1 (Sirtuin 1): A critical regulator of metabolic homeostasis, its impaired function is closely linked to insulin resistance.

7. PPAR (Peroxisome Proliferator-Activated Receptor): Central to lipid metabolism and glucose regulation, its dysregulation can lead to metabolic syndrome.

8. PGC1 Alpha (Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha): Involved in mitochondrial biogenesis and oxidative metabolism.

9. NRF1 (Nuclear Respiratory Factor 1): Regulates mitochondrial function and energy metabolism.

10. TG Gene (Thyroglobulin): Implicated in thyroid hormone synthesis, affecting metabolic rate.

11. HBA1c (Hemoglobin A1c): A marker for long-term glucose levels, its elevated levels indicate poor glucose control in diabetes.

4.1.7 Proteins and their Functions

4.1.7.1 Amino Acids

1. Leucine, Valine, Tyrosine, Phenylalanine, Isoleucine: Branched-chain and aromatic amino acids involved in protein synthesis and metabolic signaling.

2. Fluorodeoxyglucose: A glucose analog used in PET scans to monitor glucose uptake in tissues.

3. Nonesterified Fatty Acid: Reflects lipid metabolism and is associated with insulin resistance.

4. p66Shc: A protein involved in oxidative stress responses and aging.
5. Monocyte Chemoattractant Protein-1 (MCP-1): Plays a role in inflammatory processes contributing to insulin resistance.
6. C-Reactive Proteins: Inflammatory markers linked to cardiovascular risk in diabetes.
7. GROalpha (CXCL1): A chemokine involved in inflammatory responses.
8. Zeta-Sarcoglycan Protein: Involved in maintaining muscle integrity.
9. Heparinase 2: Regulates extracellular matrix dynamics and affects cell signaling.
10. Adhesion G Protein-Coupled Receptor A1 (ADGRA1): Influences inflammatory responses.
11. Galactosidase Beta 1 Like 3 (GLB1L3): Participates in glycosylation processes.
12. Proprotein Convertase Subtilisin/Kexin Type 6 (PCSK6): Activates other proteins through proteolytic cleavage.

4.1.8 Molecules Identified

1. Heparan Sulfate Proteoglycans: Key components of the extracellular matrix affecting cellular communication.
2. ALKP (Alkaline Phosphatase): Enzyme linked to bone and liver health.
3. IL-1RA (Interleukin-1 Receptor Antagonist): Modulates inflammatory responses.
4. VCAM1 (Vascular Cell Adhesion Molecule 1): Plays a role in leukocyte adhesion and inflammation.
5. Pentosidine-Glucuronide: An advanced glycation end product indicating oxidative stress.
6. Glutamyl-Lysine-Sulfate: A marker for protein modification.

7. Indoxyl Sulfate: A uremic toxin associated with kidney dysfunction.
8. Dihydroxyphenyl-Gamma-Valerolactone-Glucuronide: A metabolite derived from dietary polyphenols.
9. Fluoro-6-Thia-Heptadecanoic Acid: An analog of fatty acids used in metabolic studies.
10. Kynurenine: A metabolite in the tryptophan degradation pathway.
11. LysoPC (Lysophosphatidylcholine): A lipid involved in cell membrane dynamics.
12. Phosphatidylcholine Acyl-Alkyl and Diacyl: Major components of cell membranes influencing lipid signaling.
13. Sphingomyelin: A type of sphingolipid involved in cell signaling and membrane structure.
14. Carbohydrates (Glucose and Fructose): Primary energy sources whose dysregulation is central to diabetes.
15. Lipids (Phospholipids, Sphingomyelins, Triglycerides): Essential for energy storage and cellular structure.
16. Amino Acids (Branched-Chain, Aromatic, Glycine, Glutamine): Building blocks of proteins and key metabolic intermediates.

4.1.8.1 Metabolic Ratios

1. Triglyceride-to-High-Density Lipoprotein Cholesterol Ratio: An important marker for assessing cardiovascular risk and metabolic health in diabetic conditions.

This extensive analysis underscores the complex interplay between genetic, protein, and molecular factors in the pathophysiology of diabetes and prediabetes. Understanding these interactions is crucial for developing targeted therapies and managing these metabolic disorders effectively.

4.2 Candidate Gene Prioritization (RQ4-RQ6)

4.2.1 Identification of Core Genes

Seven core genes of insulin resistance were identified from literature and pathways of several organs such as adipose tissue, liver, and skeletal muscles. These genes are IRS1, IRS2 AKT2 functional in adipose tissue, liver, muscles, and heart, FOXO1 functions in the liver and heart, TNF-α, DAG, and IKK-β functions in kidney, liver, and muscles (table 4.4). The selection of core genes associated with metabolic diseases involved a thorough process. We began with a comprehensive literature review to identify consistently reported genes and prioritized those from Genome-Wide Association Studies (GWAS) with statistically significant associations.

Table 4.4: Core genes selected from published literature and pathways

Sr No.	**Genes**	**Organ**
1.	IRS1	Adipose tissue, liver, muscles, heart
2.	AKT2	Adipose tissue, liver, muscles, heart
3.	FOXO1	Liver, heart
4.	TNF-α	Kidney, liver, muscle
5.	DAG	Liver, muscle, kidney
6.	IKK-β	Liver, muscle, kidney
7.	IRS2	Adipose tissue, liver, muscles, heart

Our focus on functional relevance considered genes actively participating in metabolism and insulin signaling pathways. Transcriptomic studies and protein-protein interaction networks provided valuable insights. The rigorous evaluation included genetic validity through replication and consideration of clinical relevance. Meta-analyses and consensus among studies strengthened our selection. Notably, we prioritized genes experimentally validated through functional assays and animal studies, ensuring the identification of robust core genes associated with metabolic diseases.

4.2.2 Interaction Networks

After conducting a thorough literature and pathway analysis to verify the genes, we identified interactions among the genes using GeneMania (figure 4.6) and downloaded the resulting interaction data for genes.

While we were able to obtain interaction data for five of the seven genes, TNF-α and IKK-β genes were not recognized by GeneMania. In addition, we investigated networks that involved multiple pathways and retrieved the relevant pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes) (table 4.5).

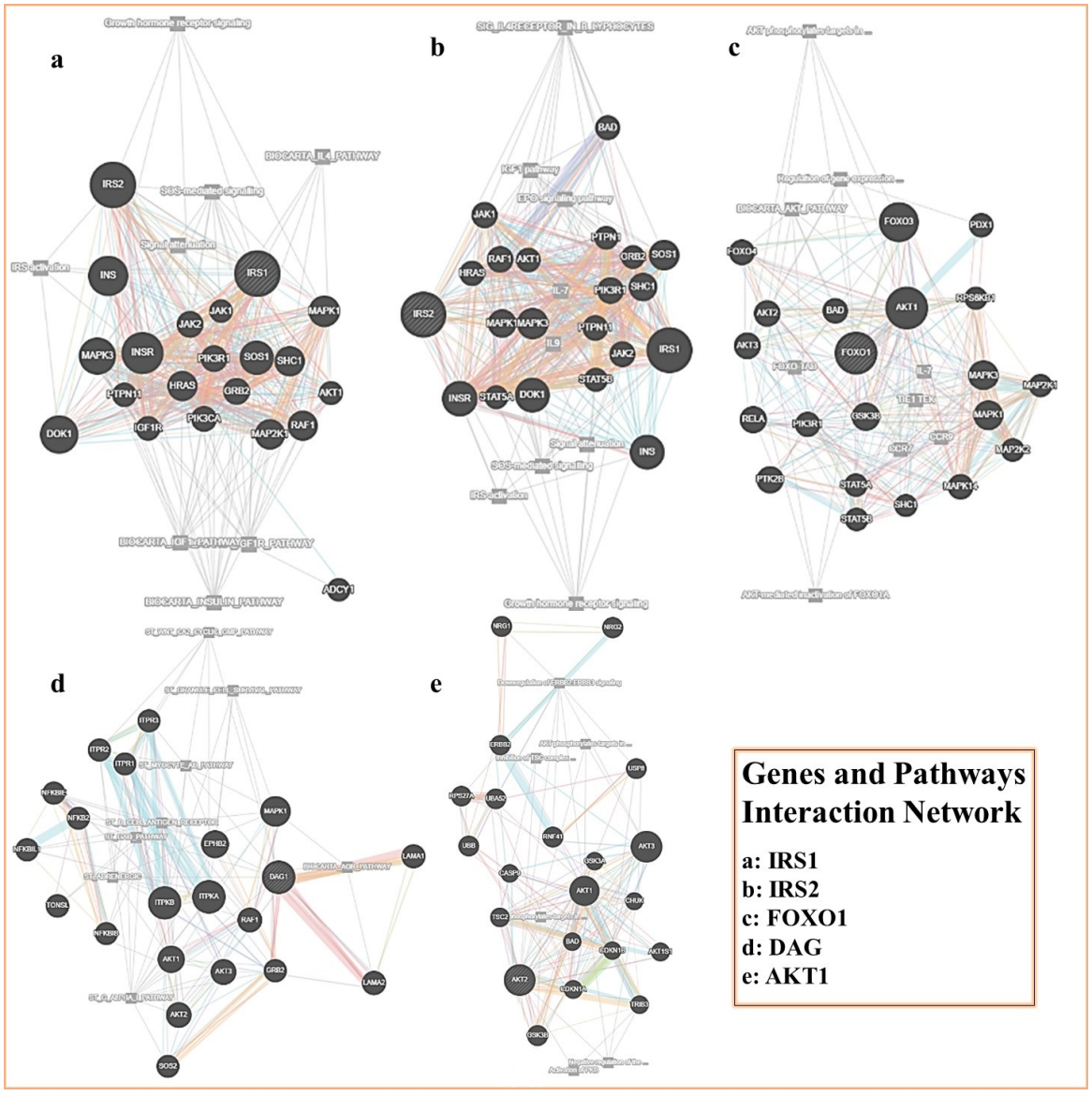


FIGURE 4.6: Genes and pathway interaction networks

TABLE 4.5: Pathways associated with genes.

Sr No.	**Gene name**	**Gene ID**	**Pathways name**	**Pathway with KEGG**	**Pathway ID**
1	IRS1	ENSG 00000 169047	Growth Harmone Signaling	JAK-STAT signaling	map04630
				PI3K signalling	map04151
				thyroid hormone	map04919
				Growth hormone synthesis, secretion, and action	map04935
			IRS activation	Oocyte meiosis	map04114
				Yersinia infection	map05135
				Insulin resistance	map04931
				Colorectal cancer	map05210
				Adipocy-tokine signaling pathway	map04920
			Signal attenuation	Insulin resistance	map04931

				Human T-cell leukemia virus 1 infection	map05166
				Viral carcinogenesis	map05203
2	IRS2	ENSG00000185950	EPO signaling pathway	Calcium signaling pathway	map04020
				Epithelial cell signaling in Helicobacter pylori infection	map05120
				Peroxisome	map04146
			IGF1 PATHWAY	Longevity regulating pathway	map04211
				AMPK signaling pathway	map04152
				Ras signaling pathway	map04014
			IL 7	Bile secretion	map04976
				Penicillin and cephalosporin biosynthesis	map00311

				Human papilloma virus infection	ma051565
				Natural killer cell-mediated cytotoxicity	map04650
3	FOXO1	ENSG00000150907	AKT pathway	PI3K-Akt signaling pathway	map04151
				Insulin signaling pathway	map04910
				ErbB signaling pathway	map04012
				Estrogen signaling pathway	map04915
			Regulation of gene expression	Regulation of actin cytoskeleton	map04810
				MicroRNAs in cancer	map05206
				Type II diabetes mellitus	map04930
				Alcoholism	map05034

			FOXO TAD	Longevity regulating pathway – worm	map04212
				Human papilloma virus infection	map05165
4	DAG	ENSG 00000 173402	ST ADRE NERGIC	beta-Adrenergic receptor agonists/ antagonists	map07214
				Adrenergic signaling in cardiomyocytes	map04261
				Dilated cardiomyopathy	map05414
				Antiarrhythmic drugs	map07037
5	AKT2	ENSG 00000 105221	Down regulation of ERBB3	MicroRNAs in cancer	map05206
				EGFR tyrosine kinase inhibitor resistance	map01521
			down regulation of ERBB2	Platinum drug resistance	map01524

These KEGG pathway insights, in combination with gene interaction data from GeneMania, provide a foundational framework for understanding the complex molecular mechanisms underlying prediabetic insulin resistance.

4.2.3 Identification of Hub Genes

We utilized a tool called Gephi 0.9.2 to develop and analyze Directed Gene Regulatory networks (figure 4.7). We applied various statistical tests, including degrees and average weighted degrees, to determine the number of interactions between the genes. The centralities of the nodes in the graph were also evaluated, such as betweenness, which indicates the significance of a node in connecting different parts of the network, closeness, which measures the shortest distance between nodes, and harmonic, which assesses the influence of neighboring nodes. Through this analysis, we aimed to identify hub genes in the network, which are nodes with the highest centralities. These hub genes may play crucial roles in regulating the activity of other genes and pathways within the network, making them potential targets for further investigation and therapeutic intervention.

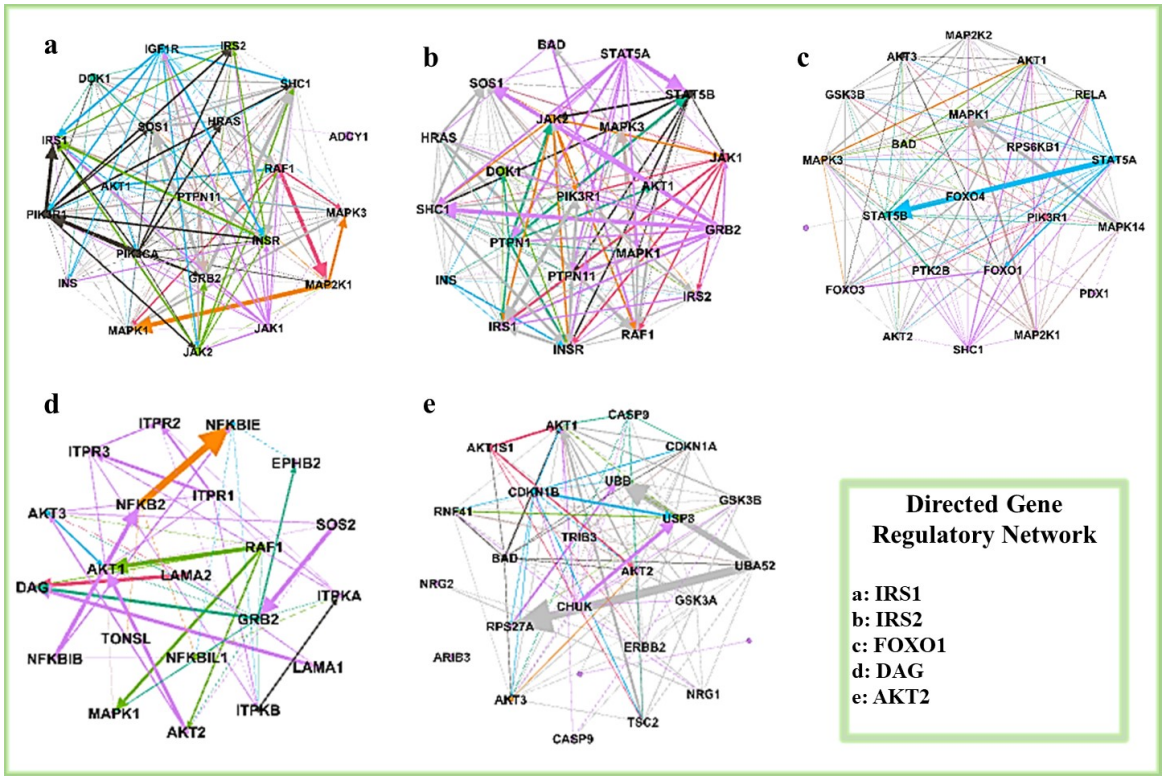


FIGURE 4.7: Directed gene regulatory networks of genes

Several genes were observed to be interacting with 5 key genes selected from published literature showing association with insulin resistance and metabolic diseases. Through Directed Gene Regulatory Network analysis and application of statistical test results,

we were able to filter 18 Hub genes UBB, UBA52, NRG1, AKT3, NFKBIL1, AKT1, RELA, MAPK1, MAPK14, PTK2B, GSK3B, MAPK3, DOK1, SOS1, RAF1, SHC1, INSR and PIK3R1.

These genes were filtered based on statistical test results values Closeness Centrality (CC) > 0.5); Harmonic closeness Centrality (HC) > 0.5); Eigen Centrality (EC) > 0.1); and Betweenness Centrality (BC) > 1) (Table 4.6).

TABLE 4.6: Identified hub genes through gene regulatory network

S. No.	**Genes**	**CC value**	**HC value**	**EC value**	**BC value**	**Hub**
1	UBB	0.7	0.785714	0.310124	18.08333	0
2	UBA52	0.714286	0.8	0.147258	22.75	0
3	NRG1	1	1	0.284012	8	0
4	AKT3	0.666667	0.75	0.356075	4	0.11174
5	NFKBIL1	1	1	0.126848	2.5	0.07556
6	AKT1	1	1	0.816136	7	0.04572
7	RELA	0.8	0.875	0.198619	11.12349	0.19884
8	MAPK1	0.666667	0.791667	0.364812	20.10043	0.15782
9	MAPK14	0.727273	0.8125	0.165843	16.45758	0.16889
10	PTK2B	0.727273	0.854167	0.166394	3.783009	0.19484
11	GSK3B	0.571429	0.760417	0.520458	18.7	0.16461
12	MAPK3	0.833333	0.9	0.428249	2.490079	0.10664
13	DOK1	1	1	0.392475	1.392857	0.11116
14	SOS1	1	1	0.350811	4.123413	0.15544
15	RAF1	0.833333	0.9	0.102	4.393687	0.21599
16	SHC1	1	1	0.331766	18.94607	0.21408
17	INSR	0.615385	0.6875	0.68485	16.50043	0.08371
18	PIK3R1	1	1	0.133489	2.946068	0.24045

The identification of these 18 hub genes highlights their central role in the regulatory network of insulin resistance. Their high centrality values suggest significant influence

in signal transduction and metabolic regulation, making them promising candidates for further experimental validation and potential targets for therapeutic intervention in prediabetes.

4.2.4 Functional Annotation and Pathway Cross-talk Network

The enrichment analysis method was used to identify over-represented classes of genes or proteins within a large group of samples, to reveal any existing associations with disease phenotypes. One approach to this was to use GO enrichment analysis, which can help to identify functionally enriched pathways of hub genes by examining their involvement in biological processes, molecular functions, and cellular components. Functional enrichment analysis is another approach, which applies statistical tests to match genes of interest with certain biological functions.

By identifying these functional associations, enrichment analysis can help to shed light on the underlying mechanisms driving disease phenotypes and provide potential targets for therapeutic intervention. To identify functionally enriched pathways associated with hub genes, GO enrichment analysis was performed for biological process, molecular function, and cellular components. The enrichment analysis was validated using TOPPGENE and literature. Only statistically significant pathways (as determined by the FDR B&H test) and had positive values greater than 1 were selected. These significant pathways and hub genes were then used to generate a network of functionally significant pathways with cross-talk between them (Figure 4.8).

4.2.5 Validation of hub genes

An enrichment analysis was performed on a set of 18 genes, which were found to be enriched in functionally diverse pathways. Among these 18 genes, 8 hub genes were validated through literature and all of the hub genes were found to have a direct or indirect role in inflammation.

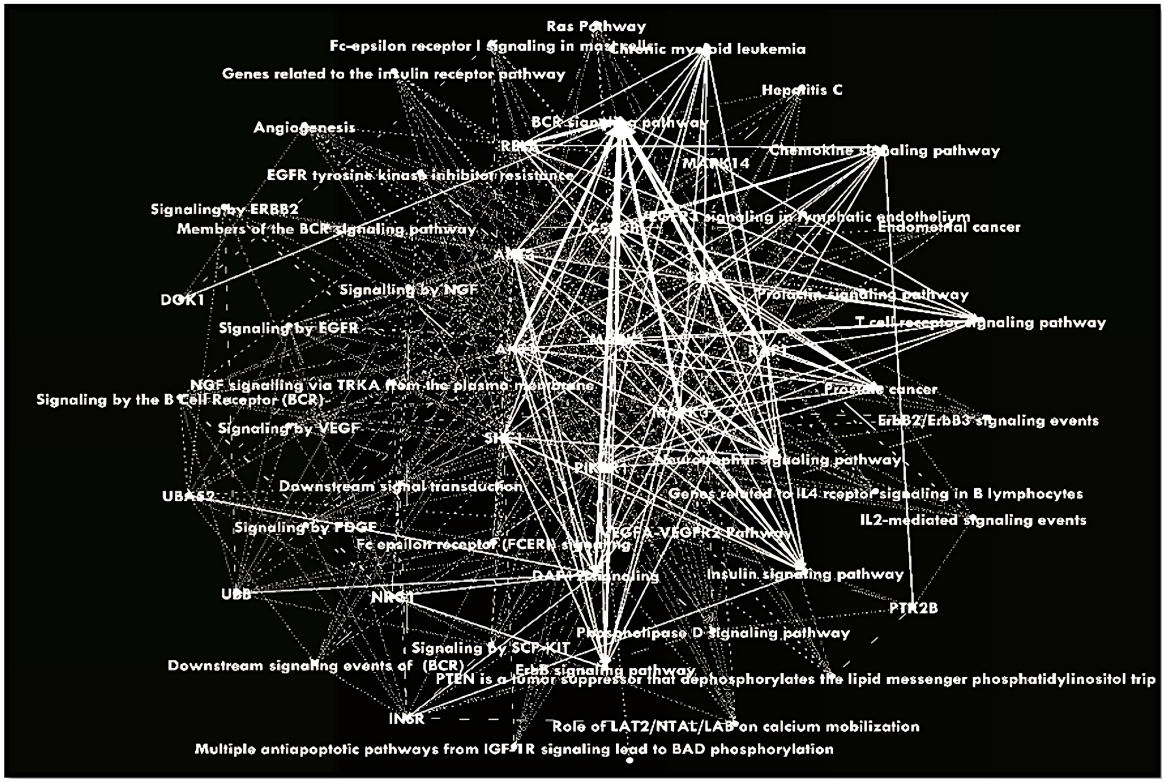


FIGURE 4.8: Pathway cross-talk generation of functionally enriched pathways of hub genes

4.2.5.1 Gene 1: SHC1

p66Shc is an adaptor protein that connects surface receptors to intracellular signaling pathways, influencing cellular response to nutrient availability. It is resistant to diabetes and obesity but inhibits glucose metabolism, promoting a shift towards anabolic metabolism. By counteracting insulin and mTOR signaling, p66Shc limits glucose uptake and metabolism, differentiating it from other Shc1 isoforms [212].

4.2.5.2 Gene 2: AKT1

AKT1 is a key mediator of insulin, IGF1, and glucose responses and regulates muscle hypertrophy and atrophy. We examined the role of AKT1 variants in metabolic syndrome endophenotypes and found that insulin activated all Akt isoforms in lean muscles, while only Akt-1 was activated in obese muscles. Reduced IRS-1 expression and PI3K activity were observed in obese muscles, which are essential for insulin indu-

ced Akt activation [23], [213].

4.2.5.3 Gene 3: PIK3R1

PIK3R1 regulatory subunits are vital for insulin signaling, with mutations causing severe insulin resistance in PI3K-dependent pathways. These mutations contribute to syndromes affecting growth, vision, fat distribution, diabetes, and ovarian cysts, highlighting PIK3R1's role in development and disease. [214].

4.2.5.4 Gene 4: GSK3B

GSK-3β is an enzyme that is suppressed by insulin. Elevated levels of GSK-3β result in insulin resistance and diabetes. Its role in beta cell development and function is poorly understood. GSK-3β deficiency in beta cells may control their growth through feedback inhibition of the insulin receptor/PI3K/Akt signaling pathway [215].

4.2.5.5 Gene 5: AKT3

Resistin plays a role in insulin resistance by reducing insulin-stimulated glucose oxidation and decreasing IRS-1 and Akt1 function, leading to reduced glucose uptake and GLUT4 translocation. This may contribute to the pathophysiology of type 2 diabetes in obesity [216].

4.2.5.6 Gene 6: RELA

TRB1 functions as a transcriptional coactivator of RELA in adipocytes, enhancing the expression of proinflammatory cytokines. This dual role of TRB1 likely contributes to the amplification of inflammatory responses in white adipose tissue, a characteristic feature observed in conditions such as sepsis, insulin resistance, and obesity-associated type 2 diabetes. The upregulation of TRB1 may thus serve as a molecular link between chronic low-grade inflammation and metabolic dysregulation. Understanding

the regulatory mechanisms of TRB1 could provide novel insights into the pathogenesis of metabolic diseases and uncover potential therapeutic targets aimed at modulating adipose tissue inflammation [217].

4.2.5.7 Gene 7: MAPK1

The role of p38 MAPK in regulating glucose transport in insulin-sensitive skeletal muscle, potentially through increased GLUT-4 activity, is not universally accepted. The specific role of p38 MAPK in regulating insulin-stimulated glucose transport in cultured myocytes is unclear [218], [219].

4.2.5.8 Gene 8: INSR

Obesity is linked to insulin resistance and inflammation, and low-dose insulin infusion triggers a response in circulating human mononuclear cells (MNC). MNCs from obese individuals have reduced levels of phosphorylated insulin receptor beta subunit (p-INSR-beta), increased levels of pro-inflammatory mediators (IKBKB, SOCS, and PRKCB2) that are associated with p-INSR-beta, and reduced insulin sensitivity [220].

4.3 Results for Objective 3 (RQ6-RQ9)

We employed an in-silico approach to identify biomarkers associated with prediabetic insulin resistance. Initially, biomarkers were selected from literature review and designated as set no 1. Subsequently, an interaction network was constructed using set no 1 genes, revealing additional genes interacting with these initial biomarkers, which were termed set no 2. Enrichment analysis was then conducted on all identified genes to annotate their biological, molecular, and cellular functions. UniProt was used to obtain UniProt IDs and protein sequences for all genes. The genes identified through interaction network and literature mining were subjected to BLASTP analysis against the Protein Data Bank (PDB) to validate their three-dimensional protein structures.

Energy profiles of the modeled proteins were evaluated using the ProSA server, which indicated negative Z scores across all protein structures, confirming their structural integrity.

Protein-protein docking simulations were performed using the GRAMM-X web server and Cluspro to predict interactions between proteins from set no 1 (known biomarkers) and set no 2 (novel biomarkers). Docked protein complexes were further analyzed using Discovery Studio, revealing ARG13 as the most frequently occurring amino acid residue among all docked proteins.

Molecular docking of set no 1 biomarkers and set no 2 novel biomarkers was conducted to identify common interacting amino acid residues, providing insights into potential molecular interactions relevant to prediabetic insulin resistance

4.4 Prediabetic Insulin Resistance Genes Through Text Mining

By literature review through manual process, 25 different genes were selected that are placed in table 4.7. there were many articles on pub med and google scholar that were reviewed and then selected these 25 genes that play role in prediabetes and insulin resistance.

Table 4.7: Genes list from text mining

Sr.no	Gene Name	Reference
1	HNF4A(MODY1)	[158]
2	GCK(MODY2)	[159]
3	HNF1A(MODY3)	[160]
4	PKM	[161]
5	FASN	[161]
6	ACACA	[161]
7	PEPCK	[161]

Table 4.7 continued from previous page

Sr.no	Gene Name	Reference
8	INSR	[162]
9	IRS1	[163]
10	IRS2	[163]
11	GUT4	[162]
12	IGF1	[162]
13	GCKR	[162]
14	SIRT1	[164]
15	GAPDH	[164]
16	PKLR	[163]
17	SSTR2	[164]
18	FOS	[161]
19	GCK	[161]
20	APOA1	[161]
21	PCK2	[161]
22	G6PC	[161]
23	APOC3	[162]
24	GLP1R	[162]

Through a systematic literature review, 24 distinct genes were identified and compiled in table 4.7. These genes were selected based on their documented roles in prediabetes and insulin resistance, drawing from comprehensive reviews of articles available on PubMed and Google Scholar.

4.4.1 Identification of Novel Biomarkers Associated with Established Genes via Gene Interaction Networks

The FunCoup database (http://FunCoup.sbc.su.se) was utilized as a bioinformatics tool to analyze network interactions among genes and their products, facilitating

the creation of comprehensive global interaction networks [221]. Through a systematic literature review using the FunCoup server, genes interacting with those initially queried were identified. These newly discovered genes, which interacted with the queried genes, were documented and compiled in table 4.8 as novel biomarkers.

The interaction network derived from literature analysis was visually depicted in figure 4.9. This network illustrated the extensive interactions of genes from set no 1 with various other genes, highlighting their interconnectedness and potential roles in biological processes related to prediabetes and insulin resistance. This Figure represents a network graph illustrating gene interactions pertinent to prediabetes. Here is a scientific interpretation of the results:

Node Characteristics:

Each node represents a gene, with larger nodes possibly indicating greater significance or stronger associations within the network. The colors, ranging from blue to black, may denote varying levels of expression, significance, or types of interactions.

Gene Clusters:

The graph exhibits several clusters of tightly interconnected genes, suggesting functional cooperation or involvement in the same biological pathways. The central cluster with larger nodes likely represents key genes involved in prediabetic conditions.

Key Genes:

- HSP90AB1, GAPDH, ENO1, ACTB, TUBB, etc.: These genes are central and exhibit numerous connections, indicating their critical roles within the network. They may serve as key regulators or biomarkers in prediabetes.
- INSR, IRS1, IGF1: Genes located on the periphery yet connected to the central cluster are likely involved in insulin signaling, a crucial aspect of glucose metabolism often disrupted in prediabetes.

Sub-networks:

- Genes such as APOA1, APOC3 form a distinct sub-network, implicating their roles in lipid metabolism, an important factor in metabolic syndromes and prediabetes.
- SIRT1, IRS2: Another sub-network, suggesting involvement in insulin resistance and metabolic regulation.

Gene Interactions:

The edges, representing interactions or relationships between genes, form a dense network, particularly in the central region, suggesting a complex interplay where these genes may influence each other's expression or function.

Potential Biomarkers:

Genes with numerous interactions and central positions in the network, such as HSP90AB1 and GAPDH, could serve as potential biomarkers or therapeutic targets for prediabetes. In summary, this network graph highlights critical genes and their interactions in the context of prediabetes, indicating a complex interplay among genes involved in glucose metabolism, insulin signaling, and lipid metabolism. The central and highly connected genes might be crucial in elucidating the molecular mechanisms underlying prediabetes and could be targets for early diagnosis or therapeutic interventions.

Table 4.8: Genes list through interaction network

Sr.no	Gene Name	Sr.no	Gene Name
1	ENO2	16	YWHAE
2	PRDX1	17	HSP90AB1
3	ALDH2	18	TUBB4B
4	PYGB	19	ENO1
5	MDH2	20	NPEEPS
6	ACTB	21	TALDO1
7	TUBB	22	PSMD2

Table 4.8 continued from previous page

Sr.no	Gene Name	Sr.no	Gene Name
8	EEF2	23	MDH2
9	PGK1	24	ACSS2
10	RACK1	25	VDAC2
11	PGD	26	HSPAL
12	PGM1	27	SLC25A5
13	VCP	28	HSPD1
14	HSPA8	29	ATTP51B
15	EEF1A1	30	GPI

The genes that formed an interaction network with the Set 1 genes are detailed in table 4.8. Genes from Set 1 that were not present in the interaction network were excluded from the analysis.

4.4.2 Enrichment analysis

Enrichment analysis evaluates the functional relationship between experimentally derived genes or proteins and a database of known gene/protein sets involved in common biological tasks. A commonly used method for this purpose is over-representation analysis (ORA) [200]. Through enrichment analysis, gene annotations were performed at the cellular, molecular, and biological levels for both Set 1 and Set 2 genes. These annotations helped identify the diverse functions of the genes at different levels.

TABLE 4.9: Molecular Annotation through Enrichment Analysis

Annotation (pathway/ process)	**Significance of network) distance distrib.**	**Signifi cance of overlap (Fisher test,**	**Dataset size (uploaded gene set)**	**Dataset size (pathway gene set)**	**Dataset size (overlap)**

	(XD-Score)	**q-value)**			
Insulin-like growth factor receptor binding	2.0538	0.00041	20	13	IRS1 INSR IGF1
Phosphatidy linositol 3-kinase binding	1.3269	0.00081	20	20	IRS2 IRS1 INSR
Lipopoly-saccharide binding	0.802	1	19	11	HSPD1
ATPase activity, coupled	0.802	1	19	11	HSPA8
AMP binding	0.7339	1	19	12	ACSS2
Enzyme regulator activity	0.5839	1	19	15	PSMD2
Cell surface binding	0.5133	1	19	17	HSPD1
Oxidoreductase activity, acting on the aldehyde or oxo group , of donors NAD or NADP as acceptor	0.5133	1	19	17	ALDH2
Kinesin binding	0.4575	1	19	19	ACTB

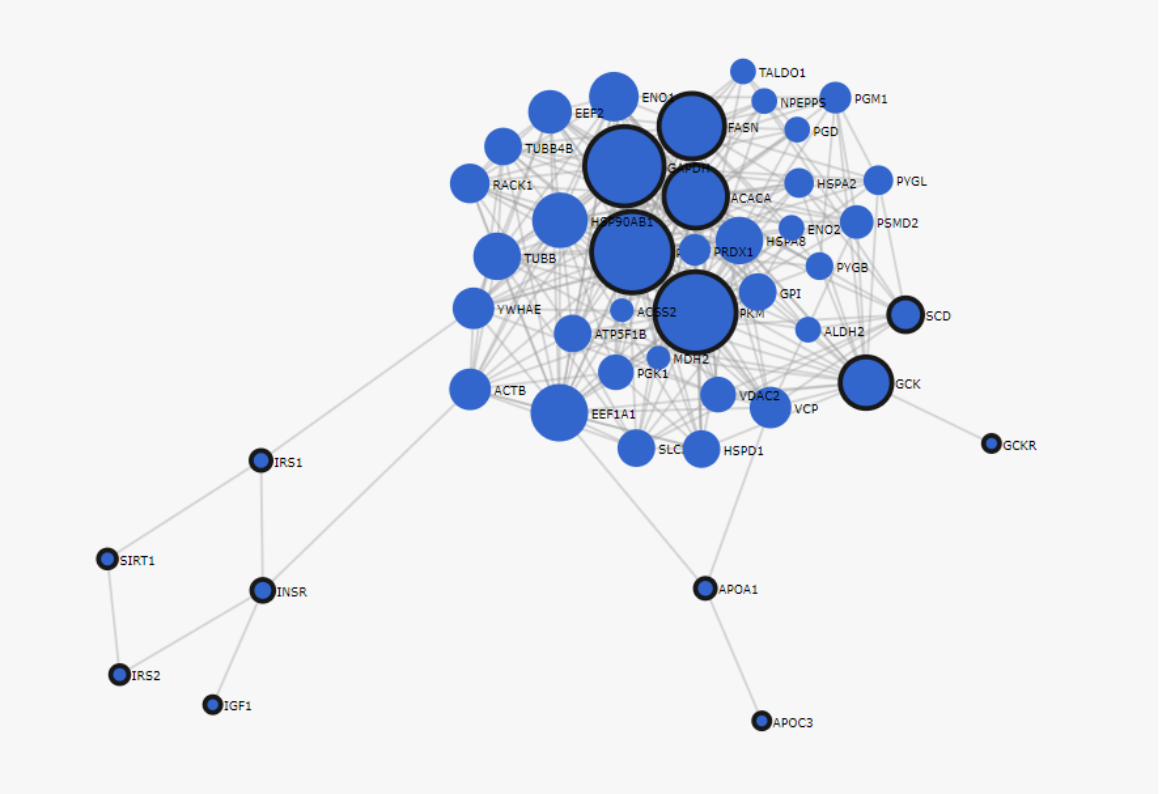


FIGURE 4.9: The interaction network of genes was analyzed using the FunCoup database (http://FunCoup.sbc.su.se). Bold and highlighted circles represent interactions with other genes, while the highlighted genes specifically belong to set number 1

In table 4.9, molecular annotations derived from enrichment analysis are presented, highlighting significant XD scores (in bold) and overlapped genes, with IRS1, IRS2, and INSR being the most frequently overlapping genes. Tissue specificity for all genes from Set 1 and Set 2 was also analyzed using the EnrichNet server, with XD scores ranging from 8.97 to 1.09 across different tissues. Additional results, including biological and cellular annotations, are provided in the supplementary data.

4.4.3 Retrieval of Protein Sequences for Set 1 and Set 2

The UniProt server was used to verify gene IDs and retrieve protein sequences. UniProt (Universal Protein Resource) offers a comprehensive, easily accessible resource for protein sequences and functional annotations [201]. Protein sequences for genes from Set 1 (selected manually) and Set 2 (interacting with Set 1 genes) were

downloaded from the UniProt Knowledgebase in FASTA (canonical) format. The gene IDs used were as follows: P35568, Q9Y4H2, P05019, Q14397, P35557, P14618, P49327, Q13085, P06213, P04406, P30613, P14618, P02647, Q16822, P02656, P09104, Q06830, P05091, P11216, P40926, P60709, Q7JJU6, P13639, P00558, P52209, P36871, P55072, P11142, P68104, P62258, P08238, P68371, P06733, P37837, Q13200, P40926, Q9NR19, P45880, P05141, P10809, Q92643. Protein sequences obtained from these gene IDs were compiled in a Word document, with the FASTA sequences provided in the supplementary data.

4.4.4 Gene Annotation

All genes obtained manually and through the interaction network were annotated using the Software Tool for Researching Annotations of Proteins (STRAP), provided by the European Bioinformatics Institute (Wellcome Trust Genome Campus, Hinxton, Cambridge, UK) [4]. Differential proteins were identified using UniProt IDs, which were then input into STRAP to generate protein annotation tables and diverse Gene Ontology (GO) charts for the proteomic data.

TABLE 4.10: Gene annotation through strap tool

Gene Name	**Uniprot Id**	**Name**	**Primary gene name**	**Length**	**Function**
L0HBA3	L0HBA3	Phospho transferase	GCK	51	Glucokinase (GCK) guides the release of insulin and the metabolism of the liver glucose.
P20823	P35568	Insulin receptor substrate 1	IRS1	1242	Can mediate insulin regulation of different cell processes.
Q9Y4H2	Q9Y4H2	Insulin receptor substrate 2	IRS2	1338	Regulation of different insulin cellular processes can be mediated.

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
P05019	P05019	Insulin-like growth factor I	IGF1		Insulin-like growth factors are structurally and functionally linked to insulin and are plasma isolated, but have dramatically increased activity to promote growth. Stimulates glucose delivery to osteoblastic bone-derived cells and operates at much lower levels than insulin.
Q14397	Q14397	Glucokinase regulatory protein	GCKR		Glucokinase (GCK) is regulated by forming an inactive enzyme complex.

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
	Q86Y46	Keratin, type II cytoskeletal 73	KRT73		Specific portion of intermediate keratin filaments in the hair follicle's internal root sheath (IRS).
P35557	P35557	Hexokinase-4	GCK		Phosphorylation to hexosis 6-phosphate catalyzed, including hexosis of D-glucose, D-fructose and D-mannose.
A7LFL1	A7LFL1	Phosphotransferase	GCK	465	Transfers sugars into the cell, including glucose, mannose and mannitol.

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
P14618	P14618	Pyruvate kinase PKM	PKM		Has a role in hair growth. Specific components of intermediary keratin filaments in the hair follicle's inner root sheath (IRS).
P49327	P49327	Fatty acid synthase	FASN		Fatty acid synthetase is a multifunctional enzyme that catalyze long-chain saturated fatty acid de novo biosynthesis in the presence of NADPH starting from the acetile-CoA and malonyl-CoA.

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
A0A0U1RQF0	A0A0U1RQF0	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	FASN	2509	The multifunctional protein contains 7 catalysts.
Q13085	Q13085	Acetyl-CoA carboxylase 1	ACACA		Cytosolic enzyme which catalyzes acetyl-CoA carboxylation into malonyl-CoA.
Q86SK9	Q86SK9	Stearoyl-CoA desaturase 5	SCD5		Stearyl-CoA desaturase that 4uses O(2) and reduced cytochrome b5 electrons to introduce saturated fatty acyl-CoA substrates with the first double connections.

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
O00767	000767	Acyl-CoA desaturase	SCD	359	The role of lipid biosynthesis plays a significant role. The role of genes involved in lipogenesis is significant in regulating the expression of mitochondrial oxidation of fatty acid.
P35558	P35558	Phosphoenolpyruvate carboxykinase, c ytosolic [GTP]	PCK1		Cytosolic phosphoenolpyruvate carboxykinase which catalyzes the reversible decarboxylation and oxaloacetate (OAA) phosphorylation and functions as the gluconeogensis-restrictive enzyme

Table 4.10 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
Q16822	Q16822	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	PCK2		A metabolism task that generates lactate glucose.
P06213	P06213	Insulin receptor	INSR		Tyrosine kinase receptor that mediates insulin pleiotropic activities.
P14672	P14672	Solute carrier family 2, facilitated glucose transporter member 4	SLC2A4		A glucose supplementary transporter regulated by insulin that plays a key role in blood circulation removal. Insulin response is 4 controlled by its intracellular role.

In table 4.10, gene annotation was performed using the STRAP tool, where gene IDs, primary names, primary lengths, and gene functions were identified. The STRAP tool revealed multiple functions for each gene. For example:

GCK plays a role in regulating insulin. **IRS1** is involved in mediating cellular processes related to insulin. **IGF1** (Insulin-like Growth Factor 1) functions in insulin-related processes and is derived from plasma. **GCKR** is associated with glucokinase regulation. **KRT73** is involved in hair root formation. **Hexokinase-4** plays a role in the formation of D-fructose and D-mannose. **Phosphotransferase** is involved in sugar transport within cells. **PKM** contributes to hair growth. **FASN** is crucial in the formation of long-chain fatty acids from acetyl-CoA. **3-Hydroxyacyl-dehydratase** acts as an acyl carrier and performs seven catalytic functions. **ACACA** is a cytosolic enzyme involved in the rate-limiting step of fatty acid synthesis. **SCD5** produces a mixture of unsaturated fatty acids. **PCK1** has different functions at low and high glucose levels, catalyzing cataplerotic reactions at low glucose levels and anaplerotic reactions at high glucose levels. **SLC2A4** facilitates the transport of glucose from the extracellular environment into cells.

4.4.5 Sequence Alignment to Identify Similar Proteins in PDB

Protein BLAST (BLASTP) was performed using the NCBI database to compare primary biological sequences, including proteins, amino acids, and nucleotides of DNA and RNA. BLAST (Basic Local Alignment Search Tool) is a program that compares primary biological sequence information to identify similarities with existing sequences in the database [204].

The BLAST database contains extensive protein and nucleotide sequences from various organisms. By comparing the subject sequence with sequences already present in the database, BLAST helps identify related sequences and provides insights into the subject sequence. Protein sequences collected from UniProt were analyzed using BLASTP to find similar three-dimensional structures in the Protein Data Bank (PDB).

The protein sequences, obtained from UniProt in FASTA format, were input into the query sequence box on the BLAST website (blast.ncbi.nlm.nih.gov) and subjected to BLAST analysis. This process identified similarities between the query protein sequences and existing protein sequences in the database across different species.

4.4.6 Sequence Similarity Check using BLAST

The sequence similarity of the query sequences with other genes was assessed using BLAST. The expected value (E-value) indicates the number of hits one can expect to see by chance when searching a database of a particular size.

A lower E-value signifies a closer match to the query sequence. In table 4.11, all query sequences had E-values of 0 or less, indicating that the query protein sequences were well-represented in the PDB, with identification rates ranging from 90% to 100%.

TABLE 4.11: Multiple sequence alignment of all the genes of set no 1 and 2

Gene name	**Description**	**Scientific name**	**Max score**	**Total score**	**Query cover**	**E value**	**Per ident**
GCK	Chain A, Glucokinase Isoform 3	Homo sapiens	102	102	100%	3e-28	98.04%
GCKR	Chain A, Glucokinase Regulatory Protein		1288	1288	100%	0.0	100.00%
IGF1	Chain A, Insulin-Like Growth Factor-I	Homo sapiens	149	149	35%	1e-47	

Table 4.11 continued from previous page

Gene name	Description	Scientific name	Max score	Total score	Query cover	E value	Per ident
IRS1	Chain A, Insulin Receptor Substrate 1	Homo sapiens	560	560	21%	0.0	100.00%
IRS2	Chain A, Insulin Receptor Substrate 1	Homo sapiens	320	320	20%	3e-100	60.44%
INSR	Chain A, Insulin receptor	Homo sapiens	2783	2783	98%	0.0	98.60%
PKM	Chain A, Pyruvate kinase PKM Isoform M2	Homo sapiens	1097	1097	100%	0.0	100.00%
PKLR	Chain A, Pyruvate kinase PKLR	Homo sapiens	1097	1097	94%	0.0	100.00%
ACACA	Chain A, Acetyl-CoA carboxylase 1	Homo sapiens	4888	4888	100%	0.0	99.96%

Table 4.11 continued from previous page

Gene name	Description	Scientific name	Max score	Total score	Query cover	E value	Per ident
SITR2	Sirt2 in complex with a 13-mer trifluoro acetylated Ran peptide	Homo sapiens	736	736	91%	0.0	100.00%

4.4.7 Domain Identification and Structure Prediction

Protein structure predictions were carried out using the Robetta server, which employs the de novo Rosetta method [206]. Robetta provides protein domain models both ab initio and via comparative methods. The energy profiles of the predicted protein structures were assessed using the ProSA server. ProSA is used to identify errors in theoretical and experimental protein structures and is known for its rapid results, even with large molecules [207]. This server calculates Z-scores and energy values for each protein structure, with lower Z-scores indicating more accurate structures. The protein sequences were entered into ProSA, and the following Z-scores were noted for the genes in Set 1 and Set 2 (table 4.12):

TABLE 4.12: Z-scores were noted for the genes in set 1 and set 2

S.No	Genes	Z- Scores
1	ACACA	-13.35
2	ACOD	-1.21
3	APOA	-5.07
4	APOC3	-0.37

5	FASN	-11.78
6	G3P	-9.9
7	GCK	-8.84
8	GCKR	-4.6
9	IGF1	-4.6
10	INSR	-9.83
11	IRS1	-5.18
12	IRS2	-5.38
13	KPYM	-10.89
14	PKLR	-10.89
15	SIRT1	-
16	ENO2	-5.34
17	PRDX1	-7.55
18	ALDH2	-9.85
19	PYGB	-11.52
20	MDH2	-9.45
21	ACTB	-10.09
22	TUBB	-8.78
23	EEF2	-12.41
24	PGK1	-11.52
25	RACK1	-3.3
26	PGD	-11.78
27	PGM1	-12.1
28	VCP	-12.96
29	HSPA8	-8.72
30	EEF1A1	-9.95
31	YWHAE	-6.8
32	HSP90AB1	-9.05
33	TUBB4B	-8.78
34	ENO1	-10.34

35	TALDO1	-8.8
36	MDH2	-9.45
37	ACSS2	-9.4
38	VDAC2	-4.74
39	SLC25A5	-4.9
40	HSPD1	-10.25

Given that all Z-scores were negative, the predicted protein structures are considered accurate.

4.4.8 Additional Validation Using MetaMQAPII & PHYRE2 Servers

The MetaMQAPII and PHYRE2 servers were also used to validate the 3D structures of the proteins. Protein sequences in PDB format were entered along with the corresponding protein names. Results were sent to the provided email address, with processing times varying from a few minutes to several hours. These servers helped confirm the accuracy and reliability of the predicted protein structures.

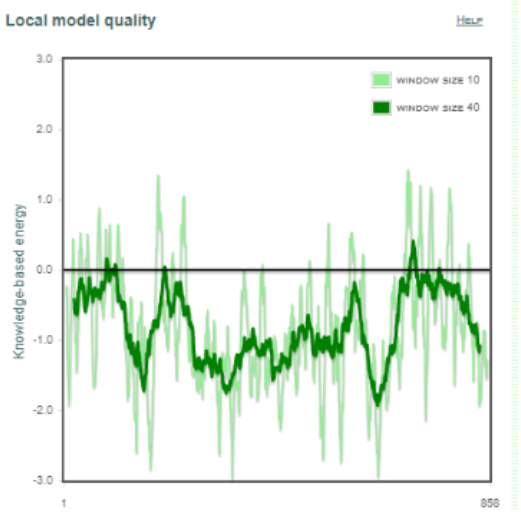


FIGURE 4.10: The protein structure validation of EEF2 were represent in this graph and Z score value is -12.41. Z-score is a numerical measurement that gives a value's relationship to the mean of a group of values.

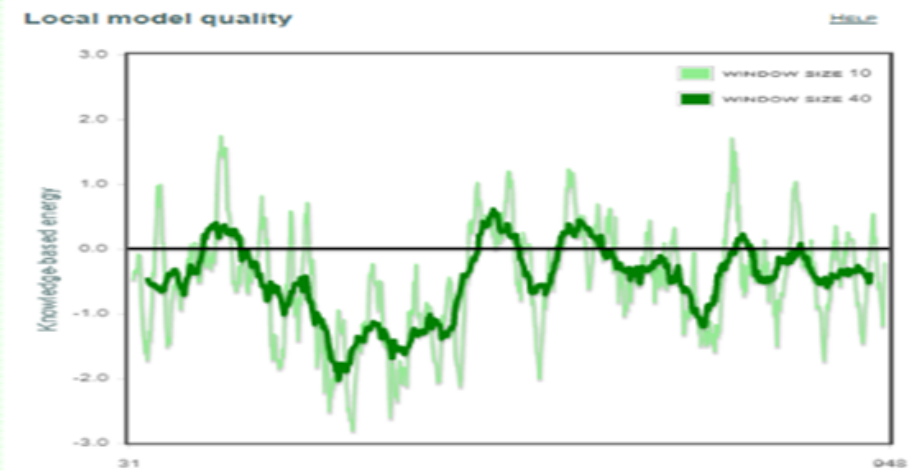


FIGURE 4.11: The protein structure validation of INSR were represent in this graph and Z score value is -9.83.

The Z-score, also known as the standard score, is a valuable statistic for two primary reasons: (a) it allows us to estimate the probability of a score occurring within a normal distribution, and (b) it enables the comparison of scores from different normal distributions. In this context, the Z-score was used to assess the predicted protein structures (figure 4.10, 4.11). A positive Z-score indicates that the protein structure's value is above the mean, while a negative Z-score suggests it is below the mean. This metric also helped determine the minimum threshold level for query proteins (figure 4.10, 4.11). Additional protein graphs are available in the supplementary data.

4.4.9 Protein-Protein Docking

In both computational and experimental biology, there is an increasing need for reliable computational methods to model and analyze protein interactions. The GRAMM-X web server [208] was utilized to investigate interactions between pairs of protein structures. This server facilitates ligand and receptor docking to analyze protein interactions. Genes from set 1 were used as receptors, while genes from set 2 served as ligands, enabling protein-protein docking. Results were collected via email. To cross-validate the docking results, ClusPro—a tool for automated docking and discrimination in predicting protein complexes—was also employed. Both GRAMM-X

and ClusPro yielded similar docking results. Predicting protein structures is a critical and challenging task in theoretical and computational biology.

Protein docking is essential for understanding cellular functions and organization. The ClusPro server, known for its ease of use in protein-protein docking, requires two files in PDB format: one for the receptor and one for the ligand. This method generates multiple docked protein structures and provides model scores for these structures [209][210].

4.4.10 Analysis of Docking

Docking analysis was conducted to examine the interactions between ligands and proteins. The docking results identified the amino acids involved in protein-protein interactions, highlighting amino acids as the functional units of proteins. This information is crucial for understanding how proteins interact at the molecular level.

Discovery Studio is a software suite designed for researchers to simulate small molecule and macromolecule systems. Developed and distributed by Dassault Systèmes BIOVIA, Discovery Studio offers a variety of applications, including simulation, ligand design, pharmacophore modeling, and structure-based design, which are essential tools in drug design and protein modeling research [211].

4.4.11 Identification of Common Amino Acids in Interacting Complexes to Confirm Set 2 Genes as Novel Biomarkers

The docking results were analyzed using Discovery Studio to identify the interacting amino acids in the docked complexes (figure 4.12, 4.13). Surface analysis was performed, and the interacting amino acids were noted. The molecular docking of set 1 biomarkers with set 2 novel biomarkers aimed to identify the most common interacting amino acid residues, as shown in table 4.13.

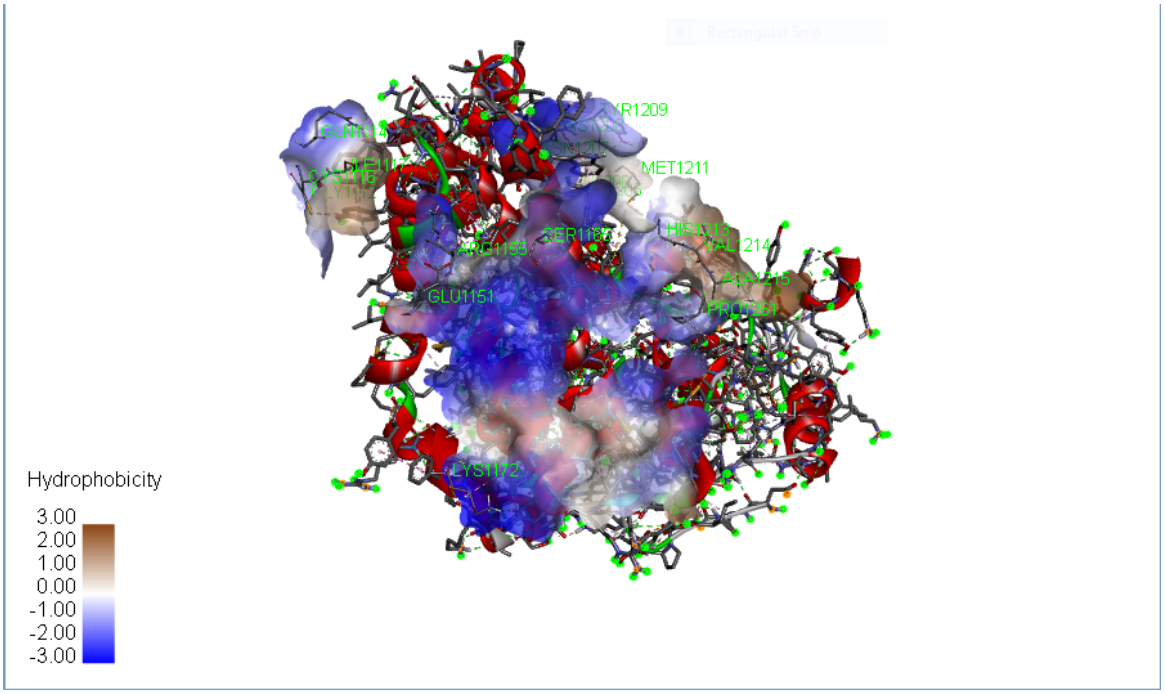


FIGURE 4.12: Docking result of ACACA and ENO1.The figure shows that there are many amino acids that help in interaction of two proteins. In the docking between ACACA-ENO1, ACACA is the receptor and ENO1 is ligand.

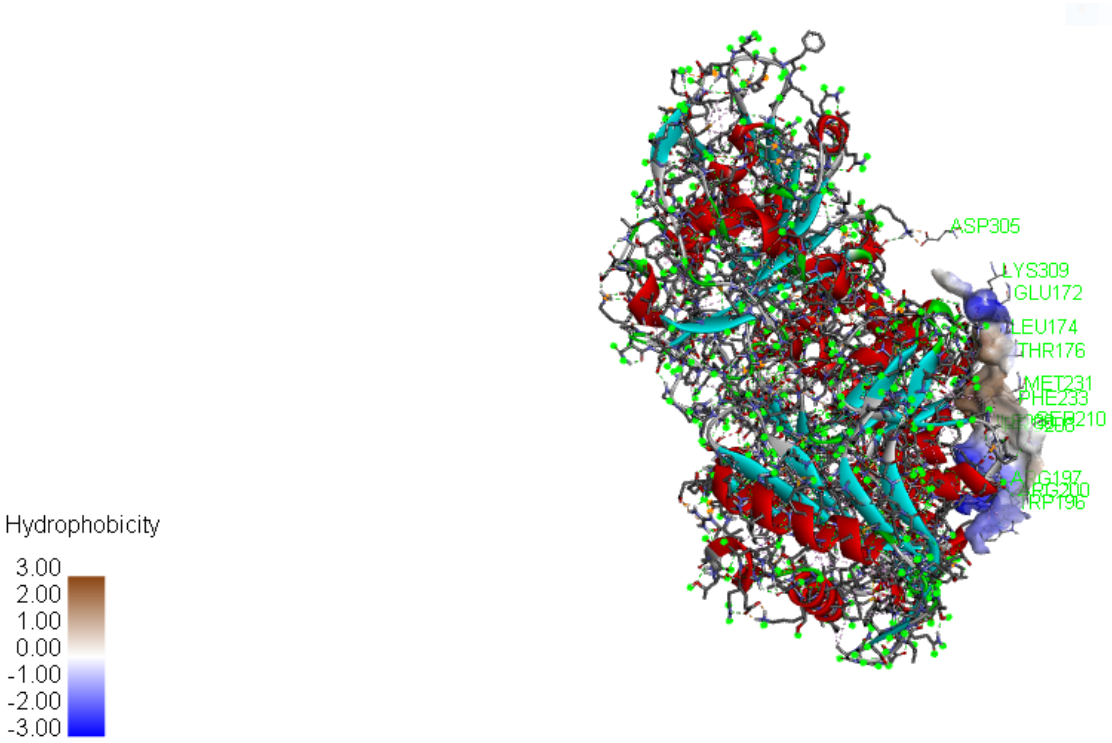


FIGURE 4.13: The docking result G3P AND PGM1The picture shows that there are many amino acids that help in interaction of two proteins. In the docking between G3P-PGM1, G3P is the receptor and PGM1 is ligand.

From table 4.13, it is evident that many of the set 2 novel biomarkers, which showed interactions with set 1 biomarkers in the gene interaction network, interact with the same amino acids of their corresponding receptors. Therefore, during the interaction of these two sets of biomarkers, the novel biomarkers can induce mutations in these amino acids, leading to the development of disease. Such mutations decrease the stability of the receptor genes, which may contribute to the pathological process.

TABLE 4.13: Most frequent amino acids as a result of docking

Sr.no	**Amino acids**	**Repetition**
1	TYR215	9
2	ARG400	10
3	ASP35	10
4	LEU10	10
5	LEU39	10
6	LYS194	10
7	MET231	10
8	THR187	10
9	ARG200	11
10	HIS218	11
11	LYS18	11
12	LYS186	11
131	PHE233	11
14	TYR45	11
15	ARG40	12
16	TRP15	12
17	ARG16	13
18	ARG21	13
19	TYR42	13
20	TRP196	14
21	ARG197	15
22	ARG13	23

Chapter 5

Discussion

A detailed meta-analysis has illuminated a broad spectrum of genes and proteins exhibiting dysregulation in both diabetes and prediabetes, underscoring the complexity of these metabolic disorders. This analysis has revealed that genes such as SGCZ, HPSE2, ADGRA1, GLB1L3, PCSK6, SIRT1, PPAR, PGC1 alpha, NRF1, TG gene, and HBA1c play significant roles. Each gene has distinct functions that contribute to the pathophysiology of diabetes and prediabetes.

For instance, SGCZ (Sarcoglycan Zeta) is involved in muscle integrity and function, and its dysregulation is linked to muscular dystrophies, potentially impacting insulin signaling pathways. HPSE2 (Heparanase 2) is associated with extracellular matrix remodeling, and alterations in this gene can affect cellular metabolism and glucose homeostasis. ADGRA1 (Adhesion G Protein-Coupled Receptor A1) plays a role in cell adhesion and signaling, influencing inflammatory responses in diabetes [239].

GLB1L3 (Galactosidase Beta 1 Like 3) is involved in carbohydrate metabolism, where its dysregulation can disrupt glucose breakdown. PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6) regulates the activation of precursor proteins, thereby influencing metabolic pathways. SIRT1 (Sirtuin 1) is a critical regulator of metabolic homeostasis, and its impaired function is closely linked to insulin resistance. PPAR (Peroxisome Proliferator-Activated Receptor) is central to lipid metabolism and glucose regulation, with its dysregulation leading to metabolic syndrome. PGC1 alpha

(Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha) is involved in mitochondrial biogenesis and oxidative metabolism. NRF1 (Nuclear Respiratory Factor 1) regulates mitochondrial function and energy metabolism. The TG gene (Thyroglobulin) is implicated in thyroid hormone synthesis, affecting metabolic rate. Lastly, HBA1c (Hemoglobin A1c) serves as a marker for long-term glucose levels, where elevated levels indicate poor glucose control in diabetes [239].

Proteins also play significant roles in these disorders. Branched-chain and aromatic amino acids such as leucine, valine, tyrosine, phenylalanine, and isoleucine are involved in protein synthesis and metabolic signaling. Fluorodeoxyglucose, a glucose analog used in PET scans, monitors glucose uptake in tissues. Nonesterified fatty acid reflects lipid metabolism and is associated with insulin resistance. The protein p66Shc is involved in oxidative stress responses and aging. Monocyte Chemoattractant Protein-1 (MCP-1) plays a role in inflammatory processes contributing to insulin resistance, while C-Reactive Proteins are inflammatory markers linked to cardiovascular risk in diabetes.

GROalpha (CXCL1) is a chemokine involved in inflammatory responses, and zeta-Sarcoglycan protein is essential for maintaining muscle integrity. Heparinase 2 regulates extracellular matrix dynamics and affects cell signaling, and ADGRA1 influences inflammatory responses. GLB1L3 participates in glycosylation processes, and PCSK6 activates other proteins through proteolytic cleavage [243].

Molecules identified in this analysis include heparan sulfate proteoglycans, key components of the extracellular matrix affecting cellular communication. Alkaline phosphatase (ALKP) is an enzyme linked to bone and liver health, and Interleukin-1 Receptor Antagonist (IL-1RA) modulates inflammatory responses. Vascular Cell Adhesion Molecule 1 (VCAM1) plays a role in leukocyte adhesion and inflammation, and pentosidine-glucuronide is an advanced glycation end product indicating oxidative stress. Glutamyl-lysine-sulfate is a marker for protein modification, and indoxyl sulfate is a uremic toxin associated with kidney dysfunction. Dihydroxyphenyl-gamma-valerolactone-glucuronide is a metabolite derived from dietary polyphenols, and fluoro-6-thia-heptadecanoic acid is an analog of fatty acids used in metabolic

studies. Kynurenine is a metabolite in the tryptophan degradation pathway, while lysoPC (lysophosphatidylcholine) is a lipid involved in cell membrane dynamics. Phosphatidylcholine acyl-alkyl and diacyl are major components of cell membranes influencing lipid signaling, and sphingomyelin is a type of sphingolipid involved in cell signaling and membrane structure. Carbohydrates such as glucose and fructose are primary energy sources whose dysregulation is central to diabetes. Lipids including phospholipids, sphingomyelins, and triglycerides are essential for energy storage and cellular structure, and amino acids like branched-chain, aromatic, glycine, and glutamine are building blocks of proteins and key metabolic intermediates [244–247].

The triglyceride-to-high-density lipoprotein cholesterol ratio is an important marker for assessing cardiovascular risk and metabolic health in diabetic conditions. This extensive analysis underscores the complex interplay between genetic, protein, and molecular factors in the pathophysiology of diabetes and prediabetes. Understanding these interactions is crucial for developing targeted therapies and managing these metabolic disorders effectively [244–247]..

In conclusion, the intricate network of genes, proteins, and molecules involved in diabetes and prediabetes highlights the need for a comprehensive approach to treatment and management. By elucidating the roles of these components, this meta-analysis provides valuable insights into the underlying mechanisms of these metabolic disorders, paving the way for more effective interventions and improved patient outcomes.

The results presented describe a systematic approach to identifying core genes, interaction networks, hub genes, functional annotations, and pathway cross-talk networks associated with insulin resistance and metabolic diseases.

Insulin resistance is a condition characterized by decreased responsiveness of cells to insulin, leading to impaired glucose uptake and metabolism [222], [223]. Several genes have been implicated in the pathogenesis of insulin resistance and metabolic diseases, including IRS1, IRS2, AKT2, FOXO1, TNF-α, DAG, and IKK-β [214], [216], [224].

IRS1 and IRS2 are two critical genes that play a role in insulin signaling. These genes encode proteins that are involved in the activation of downstream signaling pathways

that regulate glucose metabolism and lipid homeostasis [223]. Both IRS1 and IRS2 are expressed in multiple tissues, including adipose tissue, liver, and skeletal muscle. In insulin-resistant states, these genes are often dysregulated, leading to impaired insulin signaling and glucose uptake [225], [226] AKT2 is another important gene involved in insulin signaling, and it plays a crucial role in glucose metabolism and lipid homeostasis. AKT2 is primarily expressed in insulin-responsive tissues such as skeletal muscle, adipose tissue, and liver. Defects in AKT2 have been linked to insulin resistance and type 2 diabetes [223], [227].

FOXO1 is a transcription factor that regulates the expression of genes involved in glucose and lipid metabolism. FOXO1 is primarily expressed in liver and muscle tissue and is regulated by insulin signaling. In insulin-resistant states, FOXO1 activity is often dysregulated, leading to increased hepatic glucose production and impaired glucose uptake in muscle tissue [226], [228].

TNF-α is a pro-inflammatory cytokine that plays a key role in the development of insulin resistance and metabolic diseases [229]. TNF-α is primarily produced by adipose tissue and is involved in the regulation of adipocyte function and insulin signaling. Increased TNF-α levels have been linked to insulin resistance, and TNF-α inhibitors have shown promise in the treatment of metabolic diseases [225], [230].

DAG is a second messenger molecule that is involved in insulin signaling. DAG activates protein kinase C (PKC), which in turn inhibits insulin signaling and impairs glucose uptake. In insulin-resistant states, DAG levels are often elevated, leading to impaired insulin signaling and glucose metabolism [228], [231].

IKK-β is a kinase that is involved in the regulation of the NF-κB pathway, a key regulator of inflammation. IKK-β is primarily expressed in immune cells but is also present in other tissues, including adipose tissue and liver. In insulin-resistant states, IKK-β activity is often elevated, leading to increased inflammation and impaired insulin signaling [226], [230].

Through our systematic bioinformatic approach we identified 18 hub genes (UBB, UBA52, NRG1, AKT3, NFKBIL1, AKT1, RELA, MAPK1, MAPK14, PTK2B, GSK3B,

MAPK3, DOK1, SOS1, RAF1, SHC1, INSR, and PIK3R1) that play crucial roles in regulating the activity of other genes and pathways within the network. We were able to validate 8 genes SHC1, AKT1, PIK3R1, GSK3B, AKT3, RELA, MAPK1, and INSR of these 18 genes through published literature for playing a role in inflammation, either directly or indirectly that leads to pathogenesis of insulin resistance and metabolic diseases.

SHC1 is known to be involved in insulin signaling and has been shown to be a potential therapeutic target in insulin resistance [232]. AKT1, on the other hand, has been implicated in regulating glucose homeostasis, lipid metabolism, and inflammation [233]. PIK3R1 has also been found to play a role in insulin resistance by regulating glucose uptake and glycogen synthesis [232]. GSK3B is a key regulator of insulin signaling and glucose metabolism, and its dysregulation has been linked to the development of insulin resistance and metabolic diseases [234]. AKT3 has also been reported to play a role in insulin signaling and glucose metabolism [235]. RELA, a subunit of the transcription factor NF-κB, has been implicated in inflammation and insulin resistance [232]. MAPK1 has been shown to regulate glucose homeostasis and insulin sensitivity, while INSR has been identified as a key regulator of glucose uptake and insulin signaling [235].

Overall, these genes play crucial roles in regulating insulin signaling, glucose metabolism, and inflammation, and their dysregulation has been linked to the development of insulin resistance and metabolic diseases.

In our study, we focused on identifying biomarkers linked to prediabetic insulin resistance. Through extensive literature mining, we retrieved 24 distinct biomarkers associated with prediabetes and insulin resistance, which we designated as set no 1. This initial set of genes was manually curated based on their documented roles in prediabetes and insulin resistance.

To discover novel biomarkers related to prediabetic insulin resistance, we utilized a protein-protein interaction network generated through the FunCoup server. This network was constructed using the biomarkers from set no 1 as the starting point.

Within this interaction network, numerous genes were found to interact with the genes from set no 1, and these interacting genes were compiled into a new group, referred to as set no 2. These genes in set no 2 were then considered potential novel biomarkers for prediabetic insulin resistance. Next, we aimed to identify common interacting amino acid residues between the biomarkers from set no 1 and the novel biomarkers from set no 2. To achieve this, we performed protein-protein docking, where the genes from set no 1 served as receptors and the genes from set no 2 acted as ligands.

Through this docking analysis, we identified specific amino acids that frequently appeared in the interaction results. Notably, ARG13 emerged as the most recurrent amino acid. Our findings suggest that mutations in these frequently interacting amino acids, such as ARG13, could lead to decreased stability of the receptor genes. Such mutations in these biomarkers may contribute to the pathogenesis of prediabetes and insulin resistance. The significance of prediabetes and insulin resistance as precursors to type 2 diabetes has been well-documented by various researchers. Dr. Polonsky's work on MODY1 and MODY2 genes highlights their critical roles in diabetes. Mutations in these genes can lead to beta cell defects, resulting in excessive insulin production and subsequent insulin resistance [235, 236]. Our investigation into biomarkers for prediabetes and insulin resistance aligns with this understanding. Additionally, A.T. Hattersley's research on the GCK gene, which is involved in glucose phosphorylation, underscores its importance in hyperglycemia. Mutations in GCK can predispose individuals to prediabetes and diabetes [234]. We also identified GCK as a significant biomarker, emphasizing its role in maintaining glucose homeostasis [237].

Further supporting our findings, research conducted by Raquel Villegas and colleagues in Shanghai, China, examined genes such as IRS1, IRS2, INSR, and IRS3 in young individuals. Their study demonstrated that mutations in these genes disrupt glucose pathways, leading to insulin resistance, prediabetes, and eventually diabetes [238]. Our study also included these genes among others involved in glucose management, reinforcing their relevance. In summary, the primary aims and objectives of our study were to identify biomarkers linked to prediabetic insulin resistance and to discover

novel biomarkers through interaction network analysis. We then performed protein-protein docking to assess the impact of these biomarkers on the progression of insulin resistance and prediabetes. Our findings revealed that mutations in certain amino acids, particularly ARG13, play a crucial role in this process. By identifying these key biomarkers and understanding their interactions, we have contributed valuable insights into the mechanisms underlying prediabetic insulin resistance and its progression to type 2 diabetes.

To enhance the clinical relevance of the identified biomarkers, it is essential to establish their practical applications in early disease detection and patient management. The biomarkers characterized in this study hold significant potential for improving the early diagnosis of individuals at risk of developing type 2 diabetes. Integrating these molecular markers into routine clinical screening may facilitate the identification of high-risk individuals, allowing for timely interventions such as lifestyle modifications, targeted nutritional strategies, and pharmacological approaches to prevent disease progression. Furthermore, incorporating these findings into patient education programs can enhance awareness regarding metabolic health, promoting proactive measures to mitigate insulin resistance. By translating these insights into clinical practice, healthcare professionals can implement more precise risk assessment models, ultimately improving patient outcomes and reducing the burden of diabetes-related complications. Future studies should focus on validating these biomarkers in larger population cohorts and developing accessible diagnostic tools that can be seamlessly integrated into existing healthcare frameworks.

This study employed meta-analysis and in-silico approaches to identify biomarkers for prediabetic insulin resistance; however, certain limitations must be acknowledged. The reliability of the meta-analysis depended on the quality and consistency of included studies, with variations in study populations and methodologies potentially introducing bias. Computational analyses relied on existing databases, which may have contained incomplete or evolving datasets, impacting the accuracy of gene interaction networks and molecular docking results. To enhance result reliability, multiple validation steps were implemented. Homology assessments were conducted using BLASTP,

structural validation through ProSA, and docking analyses were cross-validated using GRAMM-X and ClusPro. Despite these measures, experimental validation remains necessary to confirm the clinical applicability of the identified biomarkers. Future studies should focus on functional validation and clinical trials to translate these findings into practical medical applications.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

A detailed meta-analysis has highlighted a broad spectrum of genes and proteins that exhibit dysregulation in both diabetes and prediabetes. This analysis revealed that the genes SGCZ, HPSE2, ADGRA1, GLB1L3, PCSK6, SIRT1, PPAR, PGC1 alpha, NRF1, TG gene, and HBA1c play significant roles in these metabolic disorders. SGCZ (Sarcoglycan Zeta) is involved in muscle integrity and function, with its dysregulation linked to muscular dystrophies and potential impacts on insulin signaling pathways [239]. HPSE2 (Heparanase 2) is associated with extracellular matrix remodeling, and its alteration can affect cellular metabolism and glucose homeostasis [239]. ADGRA1 (Adhesion G Protein-Coupled Receptor A1) plays a role in cell adhesion and signaling, impacting inflammatory responses in diabetes [239]. GLB1L3 (Galactosidase Beta 1 Like 3) is involved in carbohydrate metabolism, where its dysregulation can disrupt glucose breakdown [239]. PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6) regulates the activation of precursor proteins, influencing metabolic pathways. SIRT1 (Sirtuin 1) is a critical regulator of metabolic homeostasis, with impaired function closely linked to insulin resistance [240]. PPAR (Peroxisome Proliferator-Activated Receptor) is central to lipid metabolism and glucose regulation, and its dysregulation can lead to metabolic syndrome [240].

PGC1 Alpha (Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha) is involved in mitochondrial biogenesis and oxidative metabolism [241]. NRF1 (Nuclear Respiratory Factor 1) regulates mitochondrial function and energy metabolism [241]. The TG gene (Thyroglobulin) is implicated in thyroid hormone synthesis, affecting metabolic rate [241]. HBA1c (Hemoglobin A1c) serves as a marker for long-term glucose levels, with elevated levels indicating poor glucose control in diabetes [242].

The identified proteins also play significant roles in these disorders. Branched-chain and aromatic amino acids such as leucine, valine, tyrosine, phenylalanine, and isoleucine are involved in protein synthesis and metabolic signaling [243]. Fluorodeoxyglucose, a glucose analog used in PET scans, monitors glucose uptake in tissues. Nonesterified fatty acids reflect lipid metabolism and are associated with insulin resistance. The protein p66Shc is involved in oxidative stress responses and aging.

Monocyte Chemoattractant Protein-1 (MCP-1) plays a role in inflammatory processes contributing to insulin resistance, while C-Reactive Proteins are inflammatory markers linked to cardiovascular risk in diabetes [243]. GROalpha (CXCL1) is a chemokine involved in inflammatory responses, and zeta-Sarcoglycan protein is essential for maintaining muscle integrity. Heparinase 2 regulates extracellular matrix dynamics and affects cell signaling, while ADGRA1 influences inflammatory responses [244]. GLB1L3 participates in glycosylation processes, and PCSK6 activates other proteins through proteolytic cleavage [244].

Identified molecules include heparan sulfate proteoglycans, vital for extracellular communication, and ALKP, associated with bone and liver health, along with IL-1RA, an inflammation modulator [245]. VCAM1 promotes leukocyte adhesion, while pentosidine - glucuronide marks oxidative stress. Glutamyl-lysine-sulfate indicates protein modification; indoxyl sulfate relates to kidney dysfunction [245].

Metabolites such as dihydroxyphenyl-gamma-valerolactone-glucuronide and fluoro -6-thia- heptadecanoic acid support metabolic studies [246]. Kynurenine (tryptophan pathway) and lysoPC (membrane dynamics) were also noted [247]. Phosphatidylcholine variants and sphingomyelin impact lipid signaling [245]. Carbohydrates like

glucose and fructose are key diabetes-related energy sources [245]. Lipids (phospholipids, sphingomyelins, triglycerides) are essential for energy storage [245]. Amino acids—including branched-chain, aromatic, glycine, and glutamine—play metabolic roles [243]. The triglyceride-to-HDL cholesterol ratio is a critical cardiovascular and metabolic health indicator [245].

This extensive analysis underscores the complex interplay between genetic, protein, and molecular factors in the pathophysiology of diabetes and prediabetes. Understanding these interactions is crucial for developing targeted therapies and managing these metabolic disorders effectively.

The study aimed to investigate the role of insulin resistance in the onset of metabolic syndrome. To achieve this, we identified key pathways and core genes involved in insulin resistance pathways and retrieved gene interaction data. They then identified and validated hub genes, which are genes that play a central role in the network of interactions between genes. We identified 18 hub genes that were involved in insulin resistance pathways. These genes were UBB, UBA52, NRG1, AKT3, NFKBIL1, AKT1, RELA, MAPK1, MAPK14, PTK2B, GSK3B, MAPK3, DOK1, SOS1, RAF1, SHC1, INSR and PIK3R1 [239].

To validate these hub genes, we looked for evidence in the literature that supported their involvement in insulin resistance or related pathways. we found that eight of the hub genes (SHC1, AKT1, PIK3R1, GSK3B, AKT3, RELA, MAPK1, and INSR) had been previously identified as playing a role in inflammation, either directly or indirectly [243]. These genes collectively participate in complex networks that influence inflammation and metabolic processes, with their dysregulation contributing to the development and progression of metabolic diseases.

Inflammation appears to be a key contributor to the onset of insulin resistance, which is central to the development of metabolic syndrome. The hub genes identified in this study may offer promising targets for therapeutic strategies aimed at preventing or managing this condition. Understanding the molecular framework of insulin resistance is critical in unraveling the mechanisms behind metabolic syndrome. While this study

highlights several key genes and pathways, further research is essential to validate these findings and explore their relevance across diverse populations.

An in-silico strategy was applied to pinpoint biomarkers linked to prediabetic insulin resistance. Initially, a literature-based gene set (set 1) was compiled. A gene interaction network was then generated, from which additional interacting genes (set 2) were identified. Some genes from set 1 were absent in the network, yielding 40 distinct genes in total.

To determine the biological relevance of these genes, enrichment analysis was carried out using various functional annotations. UniProt was employed to retrieve protein sequences and IDs. Structural verification was performed through BLASTP against the Protein Data Bank (PDB), and protein models were evaluated for stability using ProSA, with all models showing acceptable negative Z-scores.

Protein docking simulations, conducted using GRAMM-X and ClusPro and analyzed via Discovery Studio, identified ARG13 as the most frequently involved amino acid in binding interactions. The docking results indicated that several set 2 biomarkers consistently interacted with the same residues of set 1 biomarkers. These findings suggest that common residues, particularly ARG13, may play a crucial role in receptor-ligand interactions between the two biomarker sets. This supports their potential utility as early diagnostic or therapeutic biomarkers in insulin resistance linked to prediabetes.

6.2 Future Work and Recommendation

One possible future direction for this project is to conduct a genetic analysis of the identified hub genes in different populations and individuals. This can be done using PCR (polymerase chain reaction) to amplify and detect specific DNA sequences associated with the hub genes. By analyzing the genetic variation in these genes, researchers can determine which alleles are present and how they contribute to insulin resistance and metabolic syndrome. This type of genetic analysis can provide valuable information for developing personalized interventions and treatments for metabolic syndrome.

For example, individuals with certain alleles of the identified hub genes may benefit more from specific dietary or exercise interventions or may require different medication regimens. Additionally, understanding the genetic basis of insulin resistance can help identify new drug targets and improve the development of novel therapies for metabolic syndrome.

Identified biomarkers associated with prediabetes could further be confirmed or validated in our population using different wet lab technologies. The identified biomarker was in different form it may be a protein, gene, micro-RNA and any metabolite; so, there were different methods in wet lab to validate all type of biomarkers. However, if the biomarker was a gene then sequence alignment and polymorphism PCR done; if the biomarker was a metabolite then electrophoresis and spectrophotometry was done, if the biomarker was a protein then ELISA (enzyme linked immunosorbent assay) were done and if the biomarkers was a micro-RNA then PCR (polymerase chain reaction) was done. This will further lead to identification of real biomarkers in our population that could be used for designing diagnosis techniques specific for pre-diabetic insulin resistance. Preventing prediabetic insulin resistance will inhibit the progression of deadly diseases like diabetes, cardiovascular diseases and complex syndromes such as metabolic syndrome.

Bibliography

[1] U. Smith and B. B. Kahn. Adipose tissue regulates insulin sensitivity: role of adipogenesis, de novo lipogenesis and novel lipids. *Journal of Internal Medicine*, 280(5):465–475, 2016. doi: 10.1111/joim.12540.

[2] A. R. Saltiel and C. R. Kahn. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414(6865):799–806, December 2001. doi: 10.1038/414799a.

[3] Lili Ding, Zhong Wang, Xiaoli Zhong, Yingxiang Wu, Meihua Zhu, Tao Liu, Wei Shen, Jingmin Zhang, Meian He, Cuihong Zheng, Jianping Weng, Hongxi Yang, and Weiping Jia. Serum lipoprotein (a) concentrations are inversely associated with t2d, prediabetes, and insulin resistance in a middle-aged and elderly chinese population. *Journal of Lipid Research*, 56(4):920–926, 2015. doi: 10.1194/jlr.P049015.

[4] M. Buysschaert and M. Bergman. Definition of prediabetes. *Medical Clinics of North America*, 95(2):289–297, 2011. doi: 10.1016/j.mcna.2010.11.002.

[5] A. R. Saltiel. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *World Journal of Diabetes*, 104:517–529, 2001.

[6] M. R. Rooney, M. Fang, K. Ogurtsova, B. Ozkan, J. B. Echouffo-Tcheugui, E. J. Boyko, D. J. Magliano, and E. Selvin. Global prevalence of prediabetes. *Diabetes Care*, 46(7):1388–1394, July 2023. doi: 10.2337/dc22-2376.

[7] Centers for Disease Control and Prevention. National diabetes statistics report: Estimates of diabetes and its burden in the united states. 2024. URL https://

www.cdc.gov/diabetes/php/data-research/index.html. Accessed: March 18, 2025.

[8] Xia Wu, Guo hui Li, Lan Liu, Hui fang Li, Yi Zhao, Allison Rabkin Golden, and Le Cai. Changes in pre-diabetes and diabetes prevalence and diabetes self-management behaviors across socioeconomic spectra in rural southwest china: 2013–2022. *BMC Public Health*, 25(1):594, 2025. doi: 10.1186/s12889-025-21763-y. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-21763-y.

[9] A. Basit, A. Fawwad, H. Qureshi, A. S. Shera, and N. Members. Prevalence of diabetes, pre-diabetes and associated risk factors: second national diabetes survey of pakistan (ndsp), 2016-2017. *British Medical Journal*, 8:20961, 2018. doi: 10.1136/bmjopen-2017-020961.

[10] U. Hostalek. Global epidemiology of prediabetes - present and future perspectives. *Clinical Diabetes and Endocrinology*, 5(1), 2019. doi: 10.1186/s40842-019-0080-0.

[11] A. G. Tabák, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimäki. Prediabetes: A high-risk state for diabetes development. *Lancet*, 379(9833):2279–2290, 2012. doi: 10.1016/S0140-6736(12)60283-9.

[12] M. A. Abdul-Ghani and R. A. Defronzo. Pathophysiology of prediabetes. *Current Diabetes Reports*, 9(3):193–199, 2009. doi: 10.1007/s11892-009-0032-7.

[13] A. Pérez-García, M. Torrecilla-Parra, M. Fernández-De Frutos, Y. Martín-Martín, V. Pardo-Marqués, and C. M. Ramírez. Posttranscriptional regulation of insulin resistance implications for metabolic diseases. *Biomolecules*, 12(2):208, January 2022. doi: 10.3390/biom12020208.

[14] Lei Tian, Ruixiang Zhao, Xinyi Xu, Zhiwei Zhou, Xiaofang Xu, Dongmei Luo, and Zhiqiang Zhou. Modulatory effects of lactiplantibacillus plantarum on chronic metabolic diseases. *Food Science and Human Wellness*, 12(4):959–974, 2023. doi: 10.1016/j.fshw.2022.10.018.

[15] American Diabetes Association. Prediabetes — ada. 2020. URL https://diabetes.org/diabetes/prediabetes.

[16] Centers for Disease. Control, "prediabetes: Your chance to prevent type 2 diabetes — cdc. 2018. URL https://www.cdc.gov/diabetes/basics/prediabetes.html. Diabetes.

[17] National Institute of Health. Insulin resistance and prediabetes niddk. 2020. URL https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance. 'National Institutes of Health.

[18] Hao Sun, R. Anum, and S. Morghan. Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice*, 183:109119, 2022. doi: 10.1016/j.diabres.2021.109119.

[19] World. Obesity federation. 2021. URL https://www.worldobesity.org/about/about-obesity. About Obesity.

[20] C. Lobstein. World obesity federation projections. 2021. URL https://www.worldobesity.org/about/about-obesity. About Obesity.

[21] Katherine T. Mills, Tanika N. Kelly, Joshua E. D. Reed, P. Muntner, Paul K. Whelton, and Jiang He. Global disparities of hypertension prevalence and control. *Circulation*, 134(6):441–450, August 2016. doi: 10.1161/CIRCULATIONAHA.115.018912.

[22] "Cardiovascular. Cardiovascular diseases types, causes & symptoms, " 2022. URL https://my.clevelandclinic.org/health/diseases/21493-cardiovascular-disease.

[23] Gregory A. Roth, Catherine Johnson, Adam O. Adebayo, Foad Abd-Allah, Abayomi O. Abebe, Cristiana Abbafati, Mohsen Naghavi, and Christopher J. L. Murray. Global, regional, and national age and sex specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis

for the global burden of disease study 2017. *The Lancet*, 392(10159):1736–1788, November 2018. doi: 10.1016/S0140-6736(18)32203-7.

[24] S. A. Lear and D. Gasevic. Ethnicity and metabolic syndrome: Implications for assessment, management and prevention. *Nutrients*, 12(1), January 2020. doi: 10.3390/nu12010015.

[25] M. G. Saklayen. The global epidemic of the metabolic syndrome. *Current Hypertension Reports*, 20(2):1–8, February 2018. doi: 10.1007/s11906-018-0812-z.

[26] American. Heart association. 2020. URL https://www.heart.org/en/health-topics/metabolic-syndrome. Why Metabolic Syndrome Matters.

[27] A. R. Gosmanov, E. O. Gosmanova, and E. Dillard-Cannon. Management of adult diabetic ketoacidosis. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 7:255–264, June 2014. doi: 10.2147/DMSO.S50516.

[28] Alisa T. Kraja, Sharon L. R. Kardia, Charles F. Sing, Xia Li, Hemant K. Tiwari, Donna K. Arnett, Myriam Fornage, and Myron D. Gross. A bivariate genome-wide approach to metabolic syndrome: Stampeed consortium. *Diabetes*, 60(4): 1329–1339, April 2011. doi: 10.2337/db10-1011.

[29] B. V. K. S. Lakkakula, M. Thangavelu, and U. R. Godla. Genetic variants associated with insulin signaling and glucose homeostasis in the pathogenesis of insulin resistance in polycystic ovary syndrome: A systematic review. *Journal of Assisted Reproduction and Genetics*, 30(7):883–895, July 2013. doi: 10.1007/s10815-013-0030-1.

[30] Guangyu Yu, Yuxi Chen, Yimeng Wang, Xiaoping Wu, Xuying Zhang, and Qi Yang. Lessons and applications of omics research in diabetes epidemiology. *Current Diabetes Reports*, 24(3):27–44, 2024. doi: 10.1007/s11892-024-01533-7.

[31] Ali Tayanloo Beik, Fariba Mahmoudi, Niloofar Soltani, Elham Asadollahi, Mahdi S. Sadeghi, and Ramin Shamsaei. Diabetes and heart failure: Multi omics approaches. *Frontiers in Physiology*, 12:705424, August 2021. doi: 10.3389/fphys.2021.705424.

[32] T. J. Wang. Metabolite profiles and the risk of developing diabetes. *Nature Medicine*, 17(4):448–453, March 2011. doi: 10.1038/nm.2307.

[33] E. Ahlqvist, T. S. Ahluwalia, and L. Groop. Genetics of type 2 diabetes. *Clinical Chemistry*, 57(2):241–254, February 2011. doi: 10.1373/clinchem.2010.157016.

[34] W. T. Cefalu. Insulin resistance: Cellular and clinical concepts. pages 13–26, 2001. doi: 10.1177/153537020122600103.

[35] J. H. Koeslag, P. T. Saunders, and E. Terblanche. A reappraisal of the blood glucose homeostat which comprehensively explains the type 2 diabetes mellitus-syndrome x complex. *Journal of Physiology*, 549(2):333–346, 2003. doi: 10.1113/jphysiol.2002.037895.

[36] U. Smith. Impaired diabetic insulin signaling and action occur in fat cells long before glucose intolerance. is insulin resistance initiated in the adipose tissue? *International Journal of Obesity*, 26(7):897–904, 2002. doi: 10.1038/sj.ijo.0802028.

[37] S. M. Grundy. What is the contribution of obesity to the metabolic syndrome? *Endocrinology and Metabolism Clinics of North America.*, 33(2):267–282, 2004. doi: 10.1016/j.ecl.2004.03.001.

[38] R. M. Krauss and P. W. Siri. Metabolic abnormalities: Triglyceride and low density lipoprotein. *Endocrinology and Metabolism Clinics of North America.*, 33(2):405–415, 2004. doi: 10.1016/j.ecl.2004.03.016.

[39] H. Ginsberg, J. M. Olefsky, and G. M. Reaven. Further evidence that insulin resistance exists in patients with chemical diabetes. *Diabetes*, 23(8):674–678, 1974. doi: 10.2337/diab.23.8.674.

[40] R. A. DeFronzo and E. Ferrannini. Insulin resistance: A multifaceted syndrome responsible for niddm, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care*, 14(3):173–194, 1991. doi: 10.2337/diacare.14.3.173.

[41] M. Krssak, K. Falk Petersen, A. Dresner, L. DiPietro, S. M. Vogel, D. L. Rothman, M. Roden, and G. I. Shulman. Intramyocellular lipid concentrations are correlated with insulin sensitivity in humans: a 1h nmr spectroscopy study. *Diabetologia*, 42(1):113–116, January 1999. doi: 10.1007/s001250051123.

[42] G. Perseghin, P. Scifo, F. De Cobelli, E. Pagliato, A. Battezzati, C. Arcelloni, A. Vanzulli, G. Testolin, G. Pozza, A. Del Maschio, and L. Luzi. Intramyocellular triglyceride content is a determinant of in vivo insulin resistance in humans: a 1h-13c nuclear magnetic resonance spectroscopy assessment in offspring of type 2 diabetic parents. *Diabetes*, 48(8):1600–1606, August 1999. doi: 10.2337/diabetes.48.8.1600.

[43] Frank A. Bovey, Peter A. Mirau, and H. S. Gutowsky. Nuclear magnetic resonance spectroscopy in lean. *Nucleic Acids Research*.

[44] V. T. Samuel and G. I. Shulman. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *Nucleic Acids Research*. doi: 10.1172/JCI77812.thesis.

[45] N. Tiffin, E. Adie, F. Turner, H. G. Brunner, M. A. van Driel, M. Oti, N. Lopez-Bigas, C. Ouzounis, C. Perez-Iratxeta, M. A. Andrade-Navarro, and A. Adeyemo. Computational disease gene identification: A concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Research*, 34(10):3067–3081, 2006. doi: 10.1093/nar/gkl381.

[46] S. Ellard, C. Bellanne-Chantelot, A. T. Hattersley, and European Molecular Genetics Quality Network (EMQN) MODY group. Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. *Diabetologia*, 51(4):546–553, 2008. doi: 10.1007/s00125-008-0942-y.

[47] S. Cauchi, Y. El Achhab, H. Choquet, C. Dina, F. Krempler, R. Weitgasser, C. Nejjari, W. Patsch, M. Chikri, D. Meyre, and P. Froguel. Post genome wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS One*, 3(5):e2031, May 2008. doi: 10.1371/journal.pone.0002031.

[48] Y. Tabara, H. Osawa, R. Kawamoto, H. Onuma, I. Shimizu, T. Miki, and K. Kohara. Replication study of candidate genes associated with type 2 diabetes based on genome wide screening. *Diabetes*, 58(2):493–498, February 2009. doi: 10.2337/db07-1785.

[49] A. Ighbariya and R. Weiss. Insulin resistance, prediabetes, metabolic syndrome: What should every pediatrician know? *Journal of Clinical Research in Pediatric Endocrinology*, 9(2):49–57, 2017. doi: 10.4274/JCRPE.2017.S005.

[50] Taiyi Kuo, Allison McQueen, Tzu-Chieh Chen, and Jen-Chywan Wang. Regulation of glucose homeostasis by glucocorticoids. pages 99–126, 2015.

[51] Michael P. Czech. Insulin action and resistance in obesity and type 2 diabetes. *Nature Medicine*, 23(7):804–814, 2017. doi: 10.1038/nm.4350.

[52] S. Bibi. In silico screening of novel drug candidates for diabetes mellitus. 1456505, 2017.

[53] R. K. Nicholas Dias and Yung Peng. The vasculature in prediabetes. *Physiological Behaviour*, 176(3):139–148, 2017. doi: 10.1016/j.physbeh.2017.03.040.

[54] S. Rajak, S. Banerjee, A. Saxena, N. Sinha, M. Maji, P. Mukhopadhyay, A. Mukherjee, R. S. Dey, S. K. Bhattacharyya, and A. Manna. Cellular fibronectin containing extra domain a causes insulin resistance via toll-like receptor 4. *Science Reports*, 10(1):1–11, 2020. doi: 10.1038/s41598-020-65970-6.

[55] Z. Punthakee, R. Goldenberg, and P. Katz. Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome. *Cancer Journal of Diabetes*, 42:S10–S15, 2018. doi: 10.1016/j.jcjd.2017.10.003.

[56] Z. Xia D. Liang Y. Tang J. Wei X. Wang L. Kong, P. Zhu and T. Xia. Significance of serum microrna s in prediabetes and newly diagnosed type 2 diabetes: A clinical study. *Acta Diabetologica*, 48(1):61–69, 2011. doi: 10.1007/s00592-010-0226-0.

[57] I. van Rooij M. Heeringa Z. Chen, V. van der A and O. H. Franco. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes, and

type 2 diabetes: The rotterdam study. *Clinical Nutrition*, 39(1):242–249, 2020. doi: 10.1016/j.clnu.2019.01.021.

[58] C. M. Edwards and K. Cusi. Prediabetes: A worldwide epidemic. *Endocrinology and Metabolism Clinics of North America*, 45(4):751–764, 2016. doi: 10.1016/j.ecl.2016.06.007.

[59] A. G. Tabak, C. Herder, and M. Kivimaki. Prediabetes: A high-risk state for developing diabetes progression from prediabetes to diabetes reversion to normoglycaemia risk prediction. *Lancet*, 379(9833):2279–2290, 2017. doi: 10.1016/S0140-6736(12)60283-9.Prediabetes.

[60] G. Vila, J. O. L. Jorgensen, A. Luger, and G. K. Stalla. Insulin resistance in patients with acromegaly. *Frontiers in Endocrinology*, 10, July 2019. doi: 10.3389/fendo.2019.00509.

[61] R. M. M. Khan, Z. J. Y. Chua, J. C. Tan, Y. Yang, Z. Liao, and Y. Zhao. From pre-diabetes to diabetes: Diagnosis, treatments and translational research. *Medicine*, 55(9):1–30, 2019. doi: 10.3390/medicina55090546.

[62] K. Audouze, S. Brunak, and P. Grandjean. A computational approach to chemical etiologies of diabetes. *Science Reports*, 3(3):1–6, 2013. doi: 10.1038/srep02712.

[63] S. Schinner, W. Scherbaum, S. Bornstein, and A. Barthel. Molecular mechanisms of insulin resistance in glucagon-producing. *Diabet. Med.*, 22:674–682, 2005.

[64] R. Weiss, N. Santoro, C. Giannini, A. Galderisi, G. R. Umano, and S. Caprio. Prediabetes in youths: mechanisms and biomarkers. *The Lancet Child and Adolescent Health*, 1(3):240–248, 2017. doi: 10.1016/S2352-4642(17)30044-5.

[65] C. Gar, M. Rottenkolber, C. Prehn, J. Adamski, J. Seissler, and A. Lechner. Serum and plasma amino acids as markers of prediabetes, insulin resistance, and incident diabetes. *Critical Reviews in Clinical Laboratory Sciences*, 55(1):21–32, 2018. doi: 10.1080/10408363.2017.1414143.

[66] M. Larissa Aviles-Santa, Cynthia M. Perez, Neil Schneiderman, Peter J. Savage, Robert C. Kaplan, Yanping Teng, and Erick L. Suárez. Detecting prediabetes among hispanics/latinos from diverse heritage groups: does the test matter? findings from the hispanic community health study/study of latinos. *Preventive Medicine*, 95:110–118, 2017. doi: 10.1016/j.ypmed.2016.12.009.

[67] L. D. Mastrandrea, K. A. Bree, M. H. Greenawald, and W. K. Bickel. Role of delay discounting in predicting change in hba1c for individuals with prediabetes. 42(5):851–859, 2020. doi: 10.1007/s10865-019-00026-3.Role.

[68] A. G. Tabak, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimaki. Diabetes 1 prediabetes: a high-risk state for diabetes development. *Lancet*, 379 (9833):2279–2290, 2012. doi: 10.1016/S0140-6736(12)60283-9.

[69] H. Yki-Jarvinen. Insulin resistance in type 2 diabetes. pages 174–190, 2010. doi: 10.1002/9781444324808.ch11.

[70] M. Bergman. Pathophysiology of prediabetes and treatment implications for the prevention of type 2 diabetes mellitus. *Endocrine*, 43(3):504–513, 2013. doi: 10.1007/s12020-012-9830-9.

[71] Diabetes. URL https://www.who.int/news-room/fact-sheets/detail/diabetes.

[72] S. Dagogo-Jack, C. Edeoga, E. Nyenwe, E. Chapp-Jumbo, and J. Wan. Pathobiology of prediabetes in a biracial cohort (pop-abc): Design and methods. *Ethnicity and Disease*, 21(1):33–39, 2011.

[73] E. Ferrannini, A. Gastaldelli, and P. Iozzo. Pathophysiology of prediabetes. *Medical Clinics of North America*, 95(2):327–339, 2011. doi: 10.1016/j.mcna.2010.11.005.

[74] B. Martino, L. De Luca, G. Vitale, M. Borrelli, A. Rizzo, A. Di Palma, M. Romano, A. Paldino, F. Riccardi, and L. Iuliano. Genetic risk of prediabetes and diabetes development in chronic myeloid leukemia patients treated with nilotinib. *Experimental Hematology*, 55:71–75, 2017. doi: 10.1016/j.exphem.2017.07.007.

[75] K. G. M. M. Alberti and P. Z. Zimmet. Definition, diagnosis and classification of diabetes mellitus and its complications part 1: Diagnosis and classification of diabetes mellitus provisional report of a who consultation. *Diabetes*, pages 539–553, 1998.

[76] C. Of, D. Mellitus, O. C. Of, G. Regulation, and D. O. F. Diabetes. *Diabetes Mellitus and Other Catagories of Description of Diabetes*, volume 28. 2005.

[77] C. Yang and D. Tang. Patient-specific carotid plaque progression simulation. *Computational Modeling in Engineering and Sciences*, 1(2):119–131, 2000. doi: 10.1016/j.biotechadv.2011.08.021.Secreted.

[78] S. N. Seclen, M. E. Rosas, A. J. Arias, E. Huayta, and C. A. Medina. Prevalence of diabetes and impaired fasting glucose in peru: report from perudiab, a national urban population-based longitudinal study. *BMJ Open Diabetes Ressearch and Care*, 3(1):e000110, 2015. doi: 10.1136/bmjdrc-2015-000110.

[79] M. Stumvoll, B. J. Goldstein, and T. W. Van Haeften. Type 2 diabetes: principles of pathogenesis and therapy. 365, 2010.

[80] B. Paulweber, P. Valensi, P. Poredos, A. Ceriello, O. Dapunt, A. Drev, J. Faber, C. J. Greaves, S. M. Haffner, H. Hiemstra, S. E. Kahn, R. Kahn, R. Ma, C. Nolan, L. O. Ohlson, C. Pan, E. Porrini, S. Putz, J. Repke, R. Rivas, M. Shabana, W. Sutherland, and M. van Loon. A european evidence-based guideline for the prevention of type 2 diabetes. *Diabetologia*, 42:3–36, 2010.

[81] M. Mota, D. Lixandru, M. Oprea, D. Păcurar, C. Deaconu, S. Bălănescu, C. Luchian, D. Dumitrescu, R. Mateescu, C. Voicu, A. Dima, E. Enache, C. Toma, A. Nechita, M. Laza, E. Boia, C. Cătană, C. Stanciu, A. Păduraru, and C. Bădescu. Prevalence of diabetes mellitus and prediabetes in the adult romanian population: Predatorr study. *Diabetes and Metabolism*, pages 1–9, 2015. doi: 10.1111/1753-0407.12297.

[82] B. Sv, C. Ng, S. Aung, N. Mahadi, J. Lee, A. Alavi, and A. Ibrahim. The association between diabetes mellitus and depression: Pathophysiological mechanisms. *Journal of Diabetes and Metabolic Disorders*, 9(2):120–125, 2016.

[83] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw. Idf diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Research Clinical Practice*, 94(3):311–321, 2011. doi: 10.1016/j.diabres.2011.10.029.

[84] L. M. Fa. Outcomes of diabetes care: a population-based study. *Diabetes Endocrinology*, 15(4), 2003.

[85] T. Ninomiya, V. Perkovic, D. de Zeeuw, W. Smith, K. Dwyer, and et al. Albuminuria and kidney function independently predict cardiovascular and renal outcomes in diabetes. *Journal of the American Society of Nephrology*, pages 1813–1821, 2009. doi: 10.1681/ASN.2008121270.

[86] C. D. Moulton, J. C. Pickup, and K. Ismail. Depression and diabetes 2: The link between depression and diabetes: the search for. *LANCET Diabetes Endocrinology*, 3(6):461–471, 2015. doi: 10.1016/S2213-8587(15)00134-5.

[87] P. Pathway. Pathogenesis of diabetic nephropathy: Ryuichi kikkawa and masakazu haneda. 1997.

[88] S. M. Gold, I. Dziobek, V. Sweat, and A. Tirsi. Hippocampal damage and memory impairments as possible early brain complications of type 2 diabetes. pages 711–719, 2007. doi: 10.1007/s00125-007-0602-7.

[89] R. Kikkawa. Chronic complications in diabetes mellitus. *Diabetes Endocrinology*, 2000:1996–1998, 2017.

[90] R. Donnelly, A. M. Emslie-smith, I. D. Gardner, and A. D. Morris. *Vascular complications of diabetes: Microvascular complications Macrovascular complications*, volume 320. 2000.

[91] F. Facts and O. N. Diabetes. Centers for disease control and prevention. national diabetes fact sheet. *Diabetes Endocrinology*, 07470:973–696, 2011.

[92] American Diabetes Association. *Definition and Description of Diabetes and Other Categories of Glucose Regulation*, volume 33. 2010. doi: 10.2337/dc10-S062.

[93] A. D. A. Statements. American association of clinical endocrinologists and american diabetes control improve clinical outcomes for inpatients with. *Diabetes Endocrinology*, 32(6), 2009. doi: 10.2337/dc09-9029.

[94] A. Nuzulia. Type 2 diabetes. *Angewandte Chemie International Edition*, 6(11): 951–952, 1967.

[95] A. Qaseem, T. J. Wilt, D. Kansagara, and C. Horwitch. Hemoglobin a1c targets for glycemic control with pharmacologic therapy for nonpregnant adults with type 2 diabetes mellitus: A guidance statement update from the american college of physicians. *Science*, 2018, March 2017. doi: 10.7326/M17-0939.

[96] K. M. V. Narayan and M. B. Weber. Clinical risk factors, dna variants, and the development of type 2 diabetes. *New England Journal of Medicine*, 360(13): 1360–1361, 2009. doi: 10.1056/NEJMc082624.

[97] P. R. Cavanagh. Therapeutic footwear for people with diabetes. *Science*, 20: 51–55, December 2003. doi: 10.1002/dmrr.435.

[98] R. E. Chance and B. H. Frank. Research, development, production, and safety of biosynthetic human insulin. *Diabetes Care*, 16(3):133–142, 1993. doi: 10.2337/diacare.16.3.133.

[99] I. Classification. Standards of medical care in diabetes 2014. *Science*, 37:14–80, October 2013. doi: 10.2337/dc14-S014.

[100] Metabolism Society for Endocrinology and Diabetes of Southern Africa. Abstracts of papers: The following are abstracts of papers presented at the 10th meeting of the society for endocrinology, metabolism and diabetes of southern africa held at the athe. In *Metabolism and Diabetes of Southern Africa*, pages 25–29, September 1970.

[101] I. S. Johnson. Human insulin from recombinant dna technology. *Science*, 219 (4585):632–637, 1983.

[102] V. B. Khyade. Herbals and their compounds targeting pancreatic beta cells for the treatment of diabetes. *International Journal of Scientific Study*, 6(3):1–44, 2018.

[103] P. Sonksen, J. Sonkserr, C. London, S. T. Hospital, and L. Sei. Insulin: Understanding its action in health and disease. *British Journal of Anaesthesia*, 85(1): 69–79, 2000. doi: 10.1093/bja/85.1.69.

[104] K. Docherty and D. F. Steiner. Post-translational proteolysis in polypeptide hormone biosynthesis. *Annual Review of Physiology*, 44(4):625–638, 1982. doi: 10.1146/annurev.ph.44.030182.003205.

[105] M. Shiao, B. Liao, M. Long, and H. Yu. Adaptive evolution of the insulin two-gene system in mouse. *Nature Reviews Drug Discovery*, 2008. doi: 10.1534/genetics.108.087023.

[106] A. Wollmer, B. Rannefeld, B. R. Johansen, K. R. Hejnaes, P. Balschmidt, and F. B. Hansen. Phenol-promoted structural transformation of insulin in solution. *Nature Reviews Drug Discovery*, 368:903–911, August 1987.

[107] L. Langkjar. Insulin structure and stability. *Nature Reviews Drug Discovery*, 1922, 1988.

[108] U. Derewenda, Z. Derewenda, G. G. Dodson, R. E. Hubbard, and F. Korber. *Molecular structure of insulin: The insulin monomer and its assembly*, volume 45. 1989.

[109] P. De Meyts. Insulin and its receptor: Structure, function and evolution. *Nature Reviews Drug Discovery*, 7:1351–1362, 2004. doi: 10.1002/bies.20151.

[110] P. De Meyts and J. Whittaker. Structural biology of insulin and igf1 receptors: Implications for drug design. *Nature Reviews Drug Discovery*, 1(10):769–783, 2002. doi: 10.1038/nrd917.

[111] C. M. Yip and M. D. Ward. Atomic force microscopy of insulin single crystals: Direct visualization of molecules and crystal growth. *Biophysical Journal*, 71 (2):1071–1078, 1996. doi: 10.1016/S0006-3495(96)79307-4.

[112] P. F. Egea, R. M. Stroud, and P. Walter. Targeting proteins to membranes: Structure of the signal recognition particle. *Diabetes*, pages 213–220, 2005. doi: 10.1016/j.sbi.2005.03.007.

[113] Z. Vajo, J. Fawcett, and W. C. Duckworth. Recombinant dna technology in the treatment of diabetes: Insulin analogs. *Diabetes*, 22(5):706–717, 2001.

[114] A. P. K. Abdul, B. Ahmad, P. R. D. I. R. I., and O. A. B. Ghani. Pharmacology of insulin. *Diabetes Management*, 44:10–14, 2003.

[115] W. C. Duckworth, R. G. Bennett, and F. G. Hamel. Insulin degradation: Progress and potential. *Diabetes Management*, 2014, November 1998. doi: 10.1210/edrv.19.5.0349.

[116] S. Najjar. Insulin action: Molecular basis of diabetes. In *Encyclopedia of Life Sciences*, pages 1–10. 2003. doi: 10.1038/npg.els.0001402.

[117] Sarah K. Hamza N. Marium R., Aisha T. and Faizan A. From pre-diabetes to diabetes: Diagnosis. *Diabetes Management*, pages 1–30, 2023.

[118] T. Pros and D. D. With. The pros and cons of diagnosing diabetes with a1c. *Diabetes Care*, 34, 2011. doi: 10.2337/dc11-s216.

[119] A. Nuzulia. Standards of medical care in diabetes e. *Angew. Chemie Int. Ed.*, 6(11):951–952, 1967.

[120] D. Care and S. S. Suppl. Classification and diagnosis of diabetes: Standards of medical care in diabetes 2018. *Diabetes Care*, 41(Suppl 1):S13–S27, January 2018. doi: 10.2337/dc18-S002.

[121] S. E. Kahn. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia*, 46:3–19, 2003. doi: 10.1007/s00125-002-1009-0.

[122] C. M. Edwards. Prediabetes: A worldwide epidemic. *Diabetes*, 2016. doi: 10.1016/j.ecl.2016.06.007.

[123] N. Mangalakumar. Scholarship at uwindsor: An adaptive clustering algorithm for gene expression time-series data analysis. *Diabetes*, 2018.

[124] J. C. C. Immunol, C. M. Mcgeough, and A. J. Bjourson. Clinical and cellular immunology diagnostic, prognostic and theranostic genetic biomarkers for rheumatoid arthritis. *Diabetes*, 2012. doi: 10.4172/2155-9899.S6-002.

[125] J. R. Prensner, M. A. Rubin, J. T. Wei, and A. M. Chinnaiyan. *Beyond PSA: The Next Generation of Prostate Cancer Biomarkers*, volume 3. 2012. doi: 10.1126/scitranslmed.3003180.

[126] D. Kaufman. Journal of clinical and cellular immunology. *Diabetes*, 2015. doi: 10.4172/2155-9899.1000117.

[127] Egidio Ferrannini. Insulin resistance, insulin response, and obesity as indicators of metabolic risk. *Journal of Clinical Endocrinology and Metabolism*, 92(8): 2885–2892, 2007. doi: 10.1210/jc.2007-0334.

[128] G. Reaven. The metabolic syndrome or the insulin resistance syndrome? different names, different concepts, and different goals. 2004. doi: 10.1016/j.ecl.2004.03.002.

[129] C. E. Pearson. Repeat associated non-atg translation initiation: One dna, two transcripts, seven reading frames, potentially nine toxic entities! *Diabetes*, 7(3): 1–5, 2011. doi: 10.1371/journal.pgen.1002018.

[130] Flockhart D.A. Kesteven P., McLeod H.L. A randomized trial of genotype-guided dosing of warfarin. *New England Journal of Medicine*, 2013. doi: 10.1056/NEJMoa1311386.

[131] J. K. Aronson and R. E. Ferner. Biomarkers — a general review. *Diabetes*, pages 1–17, 2017. doi: 10.1002/cpph.19.

[132] B. Dorcely, K. Katz, R. Jagannathan, S. S. Chiang, I. J. Goldberg, and M. Bergman. Novel biomarkers for prediabetes, diabetes, and associated complications. *Diabetes*, pages 345–361, 2017.

[133] O. Articles. Use of hba1c for diagnoses of diabetes and prediabetes: Comparison with diagnoses based on fasting and 2-hr glucose values and effects of gender, race, and age. 12(5):258–268, 2014. doi: 10.1089/met.2013.0128.

[134] R. M. Bookchin and P. M. Gallop. Diagnoses of diabetes and prediabetes. *Diabetes*, 32(1):86–93, 1968.

[135] R. T. Ribeiro, M. P. Macedo, and J. F. Raposo. Hba1c, fructosamine, and glycated albumin in the detection of dysglycaemic conditions. *Diabetes*, pages 14–19, 2016.

[136] E. Danese, M. Montagnana, A. Nouvenne, and G. Lippi. Advantages and pitfalls of fructosamine and glycated albumin in the diagnosis and treatment of diabetes. 2015. doi: 10.1177/1932296814567227.

[137] C. Clark, M. C. Riddell, T. Ransom, G. M. Barlow, J. Rarick, K. Broughton, and S. Morrow. Screening for type 2 diabetes and prediabetes in obese youth: Evaluating alternate markers fructosamine, and glycated albumin. *Pediatric Diabetes*, 16:1–6, 2015. doi: 10.1111/pedi.12258.

[138] N. Furusyo, T. Koga, M. Ai, and S. Otokozawa. Utility of glycated albumin for the diagnosis of diabetes mellitus in a japanese population study: Results from the kyushu and okinawa population study (kops). *Diabetes*, pages 3028–3036, 2011. doi: 10.1007/s00125-011-2310-6.

[139] A. D. Pradhan, J. E. Manson, J. E. Buring, and P. M. Ridker. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *Diabetes*, 286(3):327–334, 2001.

[140] P. Zavodszky and P. Gal. Multiple roles of complement masp-1 at the interface of innate immune response and coagulation. *Molecular Immunology*, pages 1–10, 2014. doi: 10.1016/j.molimm.2014.05.013.

[141] T. Suzuki, M. Nakano, Y. Takeuchi, M. Matsumoto, S. Yamamoto, T. Matsuda, T. Maekawa, F. Tashiro, T. Shibasaki, Y. Yamada, and T. Kawai. Interleukin-6

enhances glucose-stimulated insulin secretion from pancreatic beta-cells: Potential involvement of the plc-ip3 dependent pathway. *Diabetes*, 60:537–547, February 2011. doi: 10.2337/db10-0796.

[142] J. Kim, R. A. Bachmann, and J. Chen. Interleukin-6 and insulin resistance. *Diabetes*, 80(08):613–633. doi: 10.1016/S0083-6729(08)00621-3.

[143] B. Vozarova, C. Weyer, R. S. Lindsay, R. E. Pratley, C. Bogardus, and P. A. Tataranni. High white blood cell count is associated with a worsening of insulin sensitivity and predicts the development of type 2 diabetes. *Diabetes*, 51:3038–3043, October 2002. doi: 10.2337/diabetes.51.10.3038.

[144] C. Mertoglu and M. Gunay. Neutrophil-lymphocyte ratio and platelet-lymphocyte ratio as useful predictive markers of prediabetes and diabetes mellitus. *Diabetes Metabolism Syndrome: Clinical Research and Reviews*, 2016. doi: 10.1016/j.dsx.2016.12.021.

[145] M. Verdoia, A. Schaffer, L. Barbieri, G. Aimaretti, P. Marino, and F. Sinigaglia. Impact of diabetes on neutrophil-to-lymphocyte ratio and its relationship to coronary artery disease. *Diabetes Metabolisum*, pages 1–8, 2015. doi: 10.1016/j.diabet.2015.01.001.

[146] Volker Grossmann, Petra Achenbach, S. Dorr, B. Wiegand, J. Wolf, H. Zoller, W. Z., and T. H. Profile of the immune and inflammatory response in individuals with prediabetes and type 2 diabetes. *Diabetes Care*, 64(5):1552–1560, May 2015. doi: 10.2337/dc14-3008.

[147] A. Festa, R. D. Agostino, R. P. Tracy, and S. M. Haffner. *Plasminogen Activator Inhibitor 1 Predict the The Insulin Resistance Atherosclerosis Study*, volume 51. no. APRIL, 2002.

[148] Angela Festa, Ralph D'Agostino, Russell Tracy, Steven Haffner, Kevin Mather, J. S. C., and M. H. States of glucose tolerance. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 19(3):562–568, 1999. doi: 10.1161/01.ATV.19.3.562.

[149] Barbara Thorand, Hubert Kolb, Jens Baumert, Wolfgang Koenig, Lloyd Chambless, Christa Meisinger, Thomas Illig, Stephan Martin, and Christian Herder. Elevated levels of interleukin-18 predict the development of type 2 diabetes: results from the monica/kora augsburg study, 1984–2002. *Diabetes*, 54(10): 2932–2938, October 2005. doi: 10.2337/diabetes.54.10.2932.

[150] M. Y. Donath. Inflammation as a sensor of metabolic stress in obesity and type 2 diabetes. *Diabetes Care*, 152:4005–4006, November 2011. doi: 10.1210/en.2011-1691.

[151] Adam G. Tabak, Maren Carstensen, Daniel R. Witte, Eric J. Brunner, Martin J. Shipley, Markus Jokela, Michael Roden, Mika Kivimäki, and Christian Herder. Adiponectin trajectories before type 2 diabetes diagnosis: Whitehall ii study. *Diabetes Care*, 35(12):2540–2547, 2012. doi: 10.2337/dc11-2263.

[152] Norbert Stefan, Qi Sun, Andreas Fritsche, Jürgen Machann, Fritz Schick, Felicia Gerst, and Charlotte Jeppesen. Impact of the adipokine adiponectin and the hepatokine fetuin-a on the development of type 2 diabetes: prospective cohort- and cross-sectional phenotyping studies. *PLOS ONE*, 9(3):e92238, 2014. doi: 10.1371/journal.pone.0092238.

[153] Debanjan Pal, Juan Carlos S. G. Carrillo, Rajesh S. Ghosh, Min Zhang, Xue Zhang, Anuja S. Bhattacharya, Steven D. S. Schneider, Graciela H. H. Minervini, Georgiana L. R. Gonzalez, Jack L. S. R. Sweet, William A. G. D. T. O. Campbell, K. Suvannakitti, Laura M. Miller, Aidan L. C. M. Chueh, Andrew J. E. D. R. P. Zimmet, Eric D. A. W. M. F. W. H. Hartley, and Michael K. H. B. G. Chen. Fetuin-a acts as an endogenous ligand of tlr4 to promote lipid-induced insulin resistance. *Nature Medicine*, 18(8):1279–1285, March 2010. doi: 10.1038/nm.2851.

[154] S. D. Aas. The formation of 2-hydroxybutyric acid in experimental animals. *Journal of Biological Chemistry*, 58, 1944.

[155] Joshua Cobb, Manju Gopalan, Mahmoud Moustafa, Surya Ghosh, Vijaya Reddy, Debra Wexler, Shrinidhi Adhikari, and Jennifer Hays. Journal of diabetes science and technology. *Journal of Diabetes Science and Technology*, 7 (1):1–5, 2013. doi: 10.1177/193229681300700112.

[156] T. J. Cunningham, L. Yao, and A. Lucena. Product inhibition of secreted phospholipase a2 may explain lysophosphatidylcholines ' unexpected therapeutic properties. 10:1–10, 2008. doi: 10.1186/1476-9255-5-17.

[157] A. Dotevall, S. Johansson, and A. Rosengren. Increased levels of triglycerides, bmi, and blood pressure and low physical activity increase the risk of diabetes in swedish women: A prospective 18-year follow-up of the beda study. *Diabetes Care*, 27(3):615–622, 2004.

[158] T. D. Filippatos, E. C. Rizos, V. Tsimihodimos, I. F. Gazi, A. D. Tselepis, and M. S. Elisaf. Small high-density lipoprotein (hdl) subclasses are increased with decreased activity of hdl-associated phospholipase a2 in subjects with prediabetes. *Lipids*, 48(6):547–555, 2013. doi: 10.1007/s11745-013-3787-1.

[159] M. Guasch-ferr and M. A. Mart. Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. 39:833–846, May 2016. doi: 10.2337/dc15-2251.

[160] G. Balducci, Stefano, Sacchetti, Massimo, Haxhi, Jonida, Orlando, Giorgio, D'Errico, Valeria, Fallucca, Sara, Menini, Stefano, and Pugliese. Physical exercise as therapy for type ii diabetes. *Diabetes Metabolism Research and Reviews*, 32(30):13–23, 2014. doi: 10.1002/dmrr.

[161] Jing Huang, Yi Zhang, Zhaohui Tong, Qing Wang, Xue Li, Xu Wang, Zhong Zhang, Yanhui Zhang, Jun Li, Shuang Zhao, Dongdong Liao, Yi Wei, and Yi Lin. Iron overload and diabetes risk: A shift from glucose. *Diabetes*, 60:1175–1180, 2011. doi: 10.2337/db10-0593.

[162] Christine von Toerne, Cornelia Huth, Tonia de Las Heras Gala, Florian Kronenberg, Christian Herder, Wolfgang Koenig, and Christa Meisinger. Masp1, thbs1,

gpld1 and apoa-iv are novel biomarkers associated with prediabetes: the kora f4 study. *Diabetologia*, 59:1882–1892, 2016. doi: 10.1007/s00125-016-4024-2.

[163] X. Zhang, C. Zhang, L. Chen, X. Han, and L. Ji. Human serum acylcarnitine profiles in different glucose tolerance states. *Diabetes Research and Clinical Practice*, 103(11):1–7, 2014. doi: 10.1016/j.diabres.2014.04.013.

[164] P. Kovacs, M. Stumvoll, G. M. Fiedler, M. Mai, A. To, and A. B. Leichtle. Serum levels of acylcarnitines are altered in prediabetic conditions. *PLOS ONE*, 8(12), 2013. doi: 10.1371/journal.pone.0082459.

[165] S. H. Adams, C. Ketterer, R. G., K. A., T. E., K. E., D. L., and J. D. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic african-american women. *The Journal of Nutrition*, 139(6):1073–1081, 2009. doi: 10.3945/jn.108.103754.1073.

[166] S. A. Summers. Ceramides in insulin resistance and lipotoxicity. *Progress in Lipid Research*, 45:42–72, 2006. doi: 10.1016/j.plipres.2005.11.002.

[167] M. Parrizas, A. Rodriguez, A. B. Ropero, A. Cebrian, E. de Lemos, C. Gonzalez, A. Davila, R. M. Luque, and J. M. Fernandez-Real. Circulating mir-192 and mir-193b are markers of intervention. *The Journal of Clinical Endocrinology & Metabolism*, 100(2):407–415, 2015. doi: 10.1210/jc.2014-2574.

[168] L. Kong, J. Zhu, W. Han, X. Jiang, M. Xu, and Y. Zhao. Significance of serum microrna in pre-diabetes and newly diagnosed type 2 diabetes: A clinical study. *Diabetologia*, 54(1):61–69, 2011. doi: 10.1007/s00592-010-0226-0.

[169] L. Sun, B. Jiang, W. Li, and J. Zou. Microrna-15a positively regulates insulin synthesis by inhibiting uncoupling protein-2 expression. *Diabetes Research and Clinical Practice*, 91(1):94–100, 2011. doi: 10.1016/j.diabres.2010.11.006.

[170] M. A. Alfaqih, F. Al-Mughales, O. Al-Shboul, M. Al Qudah, Y. S. Khader, and M. Al-Jarrah. Association of adiponectin and rs1501299 of the adipoq gene with prediabetes in jordan. *Biomolecules*, 8(4):1–9, 2018. doi: 10.3390/biom8040117.

[171] C. Giannini, M. Manco, M. Marigliano, G. Montalto, G. Pugliese, M. Mottadelli, M. Patera, L. Ippolito, N. Santoro, and S. Caprio. Co-occurrence of risk alleles in or near genes modulating insulin secretion predisposes obese youth to prediabetes. *Diabetes Care*, 37(2):475–482, 2014. doi: 10.2337/dc13-1458.

[172] H. U. Haring. Novel phenotypes of prediabetes? *Diabetologia*, 59(9):1806–1818, 2016. doi: 10.1007/s00125-016-4015-3.

[173] D. Dutta, A. Bandyopadhyay, S. Dutta, S. Roy, A. Ghosh, S. Basak, S. Banerjee, A. Das, S. Mukherjee, and P. Mukhopadhyay. Tumor necrosis factor alpha -238g/a (rs 361525) gene polymorphism predicts progression to type-2 diabetes in an eastern indian population with prediabetes. *Diabetes Research and Clinical Practice*, 99(3):e37–e41, 2013. doi: 10.1016/j.diabres.2012.12.007.

[174] G. Hou, Y. Jin, M. Liu, C. Wang, and G. Song. Ucp2–866g/a polymorphism is associated with prediabetes and type 2 diabetes. *Archives of Medical Research*, 51(6):556–563, 2020. doi: 10.1016/j.arcmed.2020.06.001.

[175] Haffner S.M. Larkin M.E. Sumner A.E., Buse J.B. A1c combined with glycated albumin improves detection of prediabetes in africans: The africans in america study. *Diabetes Care*, 39:271–277, February 2016. doi: 10.2337/dc15-1699.

[176] Brenda C. T. Kieboom, Symen Ligthart, Abbas Dehghan, Steef Kurstjens, Jeroen H. F. de Baaij, Oscar H. Franco, Albert Hofman, Robert Zietse, Bruno H. Stricker, and Ewout J. Hoorn. Serum magnesium and the risk of prediabetes: A population-based cohort study. *Diabetologia*, 60(5):843–853, 2017. doi: 10.1007/s00125-017-4224-4.

[177] J. W. Choi, S. Moon, E. J. Jang, C. H. Lee, and J. S. Park. Association of prediabetes-associated single nucleotide polymorphisms with microalbuminuria. *PLoS One*, 12(2):1–13, 2017. doi: 10.1371/journal.pone.0171367.

[178] C. L. H. Goo Jun, David Aguilar, Charles Evans, and Charles F. Burant. Metabolomic profiles associated with subtypes of prediabetes among mexican americans in starr county, texas, usa. *Physiological Behaviour*, 176(1):100–106, 2016. doi: 10.1007/s00125-019-05031-4.Metabolomic.

[179] A. Jaiswal, R. Tabassum, A. Podder, S. Ghosh, N. Tandon, and D. Bharadwaj. Elevated level of c-reactive protein is associated with risk of prediabetes in indians. *Atherosclerosis*, 222(2):495–501, 2012. doi: 10.1016/j.atherosclerosis.2012.02.034.

[180] A. Bhatti. The mantle of advanced glycation end products in micro-and macrovascular complications of type 2 diabetes mellitus. *Current Research in Diabetes and Obesity Journal*, 2(5):10–14, 2017. doi: 10.19080/crdoj.2017.02.555596.

[181] K. Luc, A. Schramm-Luc, T. J. Guzik, and T. P. Mikolajczyk. Oxidative stress and inflammatory markers in prediabetes and diabetes. *Journal of Physiology and Pharmacology*, 70(6):809–824, 2019. doi: 10.26402/jpp.2019.6.01.

[182] T. S. Assmann, M. Recamonde-Mendoza, B. M. de Souza, A. C. Bauer, and D. Crispim. Micrornas and diabetic kidney disease: Systematic review and bioinformatic analysis. In *Molecular and Cellular Endocrinology, vol. 477. Ireland Ltd*, pages 90–102. Elsevier, December 2018. doi: 10.1016/j.mce.2018.06.005.

[183] M. H. Medema, R. Van Raaphorst, E. Takano, and R. Breitling. Computational tools for the synthetic design of biochemical pathways. *Nature Reviews Microbiology*, 10(3):191–202, 2012. doi: 10.1038/nrmicro2717.

[184] Ziva Rejc, Lidija Magdevska, Tilen Trselic, Timotej Osolin, Jakob Mraz, Eva Pavliha, Nikolaj Zimic, and Matjaz S. Klancnik. Computational modelling of genome-scale metabolic networks and its application to cho cell cultures. *Computers in Biology and Medicine*, 88:150–160, 2017. doi: 10.1016/j.compbiomed.2017.07.005.

[185] K. M. Sas, A. Karnovsky, G. Michailidis, and S. Pennathur. Metabolomics and diabetes: Analytical and computational approaches. *Diabetes*, 64(3):718–732, 2015. doi: 10.2337/db14-0509.

[186] Guohua Wang, Xiong Wang, Min Zhang, Jing Wang, Ping Ma, Yan Wang, and Yongjie Li. Strategy for early identification of prediabetes in lean populations:

New insight from a prospective chinese twin cohort of children and young adults. *Diabetes Research and Clinical Practice*, 146:101–110, 2018. doi: 10.1016/j.diabres.2018.10.003.

[187] M. Franz. Genemania update 2018. *Nucleic Acids Research*, 46(W1):W60–W64, July 2018. doi: 10.1093/NAR/GKY311.

[188] B. Mathieu. Gephi - the open graph viz platform. 2008. URL https://gephi.org/. Accessed: Feb. 24.

[189] P. Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, October 2007. doi: 10.1016/j.socnet.2007.04.002.

[190] E. Estrada and N. Hatano. Communicability in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 77(3):036111, March 2008. doi: 10.1103/PhysRevE.77.036111.

[191] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1):361–362, March 2009. doi: 10.1609/icwsm.v3i1.13937.

[192] ToppGene. Suite. 2023. URL https://toppgene.cchmc.org/help/publications.jsp. p. 2170870, 2014. Accessed: Feb. 24.

[193] V. Kaimal, E. E. Bardes, S. C. Tabar, A. G. Jegga, and B. J. Aronow. Toppcluster: A multiple gene list feature analyzer for comparative enrichment clustering and networkbased dissection of biological systems. *Nucleic Acids Research*, 38, May 2010. doi: 10.1093/NAR/GKQ418.

[194] T. Schmitt, C. Ogris, and E. L. L. Sonnhammer. Funcoup 3.0: Database of genome-wide functional coupling networks. *Nucleic Acids Research*, 42(D1): 380–388, 2014. doi: 10.1093/nar/gkt984.

[195] E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, and A. Valencia. Enrichnet: Network-based gene set enrichment analysis. *Bioinformatics*, 28(18):451–457, 2012. doi: 10.1093/bioinformatics/bts389.

[196] The UniProt Consortium. The universal protein resource (uniprot). *Nucleic Acids Research*, 36:190–195, November 2007. doi: 10.1093/nar/gkmmp598.

[197] The UniProt Consortium. The universal protein resource (uniprot) in 2010. *Nucleic Acids Research*, 38:142–148, October 2009. doi: 10.1093/nar/gkp846.

[198] V. N. Bhatia, D. H. Perlman, C. E. Costello, and M. E. McComb. Software tool for researching annotations of proteins: Open-source protein annotation software with data visualization. *Analytical Chemistry*, 81(23):9819–9823, 2009. doi: 10.1021/ac901335x.

[199] J. Ye, S. Mcginnis, and T. L. Madden. Blast: improvements for better sequence analysis. 34:6–9, 2006. doi: 10.1093/nar/gkl164.

[200] D. E. Kim, D. Chivian, and D. Baker. Protein structure prediction and analysis using the robetta server. *Nucleic Acids Research*, 32:526–531, 2004. doi: 10.1093/nar/gkh468.

[201] M. Wiederstein and M. J. Sippl. Prosa-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35:407–410, 2007. doi: 10.1093/nar/gkm290.

[202] A. Tovchigrechko and I. A. Vakser. Gramm-x public web server for protein-protein docking. *Nucleic Acids Research*, 34:310–314, 2006. doi: 10.1093/nar/gkl206.

[203] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho. Cluspro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20(1):45–50, 2004. doi: 10.1093/bioinformatics/btg371.

[204] D. Kozakov, D. R. Hall, Y. Wu, S. E. Mottarella, L. Luo, H. A. Carlson, C. A. Rohl, A. Kolinski, R. V. Pappu, and D. W. R. R. The cluspro web server for protein-protein docking. *Nature Protocols*, 12(2):255–278, 2017. doi: 10.1038/nprot.2016.169.

[205] V. K. Vyas, R. D. Ukawala, M. Ghate, and C. Chintha. Homology modeling a fast tool for drug discovery: current perspectives. *Indian Journal of Pharmaceutical Sciences*, 74(1):1, 2012.

[206] "Advancing. Heart, lung, blood, and sleep research & innovation — nhlbi, nih. 2023. URL https://www.nhlbi.nih.gov/. Heart, Lung Blood Inst., Accessed: Mar. 14.

[207] Veronique L. Roger, Alan S. Go, Donald M. Lloyd-Jones, Robert J. Adams, Jessie D. Berry, Timothy M. Brown, and et al. Heart disease and stroke statistics-2011 update: A report from the american heart association. *Circulation*, 123(4):18–209, 2011. doi: 10.1161/CIR.0b013e3182009701.

[208] Benjamin F. Voight, Michael Culpan, Andrew R. Dyer, J. L. Follis, Paul W. Franks, Kyle J. Gaulton, Soumya Ghosh, and et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*, 42(7):579–589, June 2010. doi: 10.1038/ng.609.

[209] Johan Rung, Stephane Cauchi, Anders Albrechtsen, Lishuang Shen, Ghislain Rocheleau, Christine Cavalcanti-Proenca, and Francois Bacot. Genetic variant near irs1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nature Genetics*, 41(10):1110–1115, September 2009. doi: 10.1038/ng.443.

[210] J. W. Knowles, S. S. Baker, C. E. Barlow, A. R. Dyer, C. Gu, D. M. Harlan, X. Ruan, J. Hsieh, A. K. Sampson, C. Li, S. E. Kahn, N. Beck, R. M. Watanabe, J. C. Florez, S. Ghosh, G. Michailidis, J. O. Hill, T. Mott, M. I. McCarthy, N. D. Palmer, and D. J. Hunter. Identification and validation of n-acetyltransferase 2 as an insulin sensitivity gene. *Journal of Clinical Investigation*, 125(4):1739–1751, April 2015. doi: 10.1172/JCI74692.

[211] Guojun Chen, Hitesh K. Tiwari, Juying Hu, Themistocles L. Assimes, A. R. Dyer, R. M. Watanabe, I. Baric, Feng Lu, K. L. Edwards, C. J. Rodriguez, A. R. Dyer, K. Salter, D. R. Jacobs, T. H. Mosley, S. E. Kahn, H. H. Goring, H. S. Ahn, H. Zhang, M. Watanabe, Q. Wong, D. O'Connor, P. W. Franks, N. D. Palmer, E. G. Burchard, D. L. Koller, L. Wang, A. B. Goldfine, N. Risch, M. I.

McCarthy, and N. D. Palmer. Genome-wide association study identifies novel loci associated with fasting insulin and insulin resistance in african americans. *Human Molecular Genetics*, 21(20):4530–4536, October 2012. doi: 10.1093/hmg/dds282.

[212] Leif C. Groop, Roberto C. Bonadonna, Stefano Del Prato, Per Hjemdahl, John Hinkley, David R. Hodge, C. R. Kahn, Jerrold M. Olefsky, Richard E. Pratley, G. Toffolo, and M. Vranic. Glucose and free fatty acid metabolism in non-insulin-dependent diabetes mellitus. evidence for multiple sites of insulin resistance. *Journal of Clinical Investigation*, 84(1):205–213, July 1989. doi: 10.1172/JCI114142.

[213] Elisabetta Bugianesi, Alan J. McCullough, and Salvatore Fargion. Insulin resistance in non-diabetic patients with non-alcoholic fatty liver disease: Sites and mechanisms. *Diabetologia*, 48(4):634–642, April 2005. doi: 10.1007/s00125-005-1682-x.

[214] Maria J. Redondo and Noel G. Morgan. Heterogeneity and endotypes in type 1 diabetes mellitus. *Nature Reviews Endocrinology*, 19(9):542–554, 2023.

[215] Zari Dastani, Lise Laurin, Andrew R. E. Sinclair, Anna Glazer, Mark O. Goodarzi, Loïc Yengo, Robin G. Walters, and Nicholas J. Timpson. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genetics*, 8(3):e1002607, March 2012. doi: 10.1371/journal.pgen.1002607.

[216] A. T. Hattersley, S. Bevan, C. Lynas, C. R. C., H. E. M., and W. S. K. Linkage of type 2 diabetes to the glucokinase gene. *Lancet*, 339(8805):1307–1310, 1992. doi: 10.1016/0140-6736(92)91958-B.

[217] B. I. Kahn and E. H. Morgan. Hospital of the programmed failure of the insulin resistance. *New England Journal of Medicine*, 334(12):777–783, 1996.

[218] P. R. Langlais, L. J. Mandarino, and W. T. Garvey. Normal physiology of insulin action: Mechanisms of insulin signal transduction. *Endocrine Reviews*, 36(1):45–65, 2015.

[219] R. Villegas, X. Wang, Y. Chen, L. Zhang, Y. Liu, Z. Li, J. Li, J. Lin, Y. He, and Y. Wang. Genetic variation and insulin resistance in middle-aged chinese men. *Annals of Human Genetics*, 79(5):357–365, 2015. doi: 10.1111/ahg.12124.

[220] M. Benito. Tissue specificity on insulin action and resistance: Past to recent mechanisms. *Acta Physiologica*, 201(3):297–312, 2011. doi: 10.1111/j.1748-1716.2010.02201.x.

[221] K. D. Pruitt, T. Tatusova, and D. R. Maglott. Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. 33:501–504, 2005. doi: 10.1093/nar/gki025.

[222] G. S. Hotamisligil. Inflammation and metabolic disorders. *Nature*, 444(7121):860–867, December 2006. doi: 10.1038/nature05485.

[223] J. Hirosumi, F. W. Goetz, K. Kato, H. C., H. S., D. H., and K. Y. A central role for jnk in obesity and insulin resistance. *Nature*, 420(6913):333–336, November 2002. doi: 10.1038/nature01137.

[224] G. Boden. Obesity and free fatty acids. *Endocrinology and Metabolism Clinics of North America*, 37(3):635–646, September 2008. doi: 10.1016/J.ECL.2008.06.007.

[225] D. Cai, M. Yuan, D. Frantz, S. Pavlides, E. J. Keliher, A. Kade, J. Lee, W. R. E., and G. S. Hotamisligil. Local and systemic insulin resistance resulting from hepatic activation of ikk-beta and nf-kappab. *Nature Medicine*, 11(2):183–190, January 2005. doi: 10.1038/nm1166.

[226] A. R. Saltiel and C. R. Kahn. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414(6865):799–806, 2001. doi: 10.1038/414799a.

[227] C. Martinez Calejman, S. Trefely, S. W. Entwisle, A. Luciano, S. M. Jung, W. Hsiao, and D. A. Guertin. mtorc2-akt signaling to atp-citrate lyase drives brown adipogenesis and de novo lipogenesis. *Nature Communications*, 11(1):575, January 2020. doi: 10.1038/s41467-020-14430-w.

[228] S. Patel, B. W. Doble, K. MacAulay, E. M. Sinclair, D. J. Drucker, and J. R. Woodgett. Tissue-specific role of glycogen synthase kinase 3beta in glucose homeostasis and insulin action. *Molecular and Cellular Biology*, 28(20):6314–6328, October 2008. doi: 10.1128/mcb.00763-08.

[229] S. A. Muhammad, W. Raza, T. Nguyen, B. Bai, X. Wu, and J. Chen. Cellular signaling pathways in insulin resistance-systems biology analyses of microarray dataset reveals new drug target gene signatures of type 2 diabetes mellitus. *Frontiers in Physiology*, 8:13, January 2017. doi: 10.3389/fphys.2017.00013.

[230] Bilun Ke, Zhiyun Zhao, Xin Ye, Zhanguo Gao, Vincent Manganiello, Bin Wu, and Jianping Ye. Inactivation of nf-κb p65 (rela) in liver improves insulin sensitivity and inhibits camp/pka pathway. *Diabetes*, 64(10):3355–3362, October 2015. doi: 10.2337/db15-0242.

[231] Martine Vaxillaire, Valerie Boccio, Anne Philippi, Corinne Vigouroux, Joe Terwilliger, Philippe Passa, Jacques S. Beckmann, Gilberto Velho, G. Mark Lathrop, and Philippe Froguel. A gene for maturity onset diabetes of the young (mody) maps to chromosome 12q. *Nature Genetics*, 9(4):418–423, 1995. doi: 10.1038/ng0495-418.

[232] M. A. Magnuson. Glucokinase gene structure: Functional implications of molecular genetic studies. *Diabetes*, 39(5):523–527, 1990. doi: 10.2337/diab.39.5.523.

[233] R. E. Toribio. Insulin resistance. *Nature Reviews Endocrinology*, 7(2):90–103, 2011.

[234] Valeriya Lyssenko, Anna Jonsson, Peter Almgren, Nicoló Pulizzi, Bo Isomaa, Tiinamaija Tuomi, Göran Berglund, David Altshuler, Peter Nilsson, and Leif Groop. Clinical risk factors, dna variants, and the development of type 2 diabetes. *New England Journal of Medicine*, 359(21):2220–2232, 2008. doi: 10.1056/NEJMoa0801822.

[235] S. M. Haffner. Insulin resistance, inflammation, and the prediabetic state. *American Journal of Cardiology*, 92(4):18–26, 2003. doi: 10.1016/S0002-9149(03)00612-X.

[236] G. I. Shulman. Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. *New England Journal of Medicine*, 371(12):1131–1141, September 2014. doi: 10.1056/nejmra1011035.

[237] P. E. Scherer. Adipose tissue: From lipid storage compartment to endocrine organ. *Diabetes*, 55(6):1537–1545, June 2006. doi: 10.2337/db06-0263.

[238] M. P. Czech. Insulin action and resistance in obesity and type 2 diabetes. *Nature Medicine*, 23(7):804–814, July 2017. doi: 10.1038/nm.4350.

[239] T. Liu, H. Li, Y. P. Conley, B. A. Primack, J. Wang, W. Lo, and C. Li. A genome-wide association study of prediabetes status change. *Frontiers in Endocrinology*, 13:881633, 2022. doi: 10.3389/fendo.2022.881633. URL https://doi.org/10.3389/fendo.2022.881633.

[240] S. Tulipani, M. Palau-Rodriguez, A. Miñarro Alonso, F. Cardona, A. Marco-Ramell, B. Zonja, M. Lopez de Alda, A. Muñoz-Garach, A. Sanchez-Pla, F. J. Tinahones, and C. Andres-Lacueva. Biomarkers of morbid obesity and prediabetes by metabolomic profiling of human discordant phenotypes. *Clinica Chimica Acta*, 463:53–61, 2016. doi: 10.1016/j.cca.2016.10.005. URL https://doi.org/10.1016/j.cca.2016.10.005.

[241] W. Zhou, M. R. Sailani, K. Contrepois, Y. Zhou, S. Ahadi, S. R. Leopold, M. J. Zhang, V. Rao, M. Avina, T. Mishra, J. Johnson, S. Chen, A. A. Metwally, T. D. Tran, H. Nguyen, X. Zhou, B. Albright, B. Hong, L. Petersen, and M. Snyder. Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature*, 569 (7758):663–671, 2019. doi: 10.1038/s41586-019-1236-x. URL https://doi.org/10.1038/s41586-019-1236-x.

[242] S. Dagogo-Jack, A. A. Brewer, I. Owei, L. French, N. Umekwe, R. Rosenthal, and J. Wan. Pathobiology and reversibility of prediabetes in a biracial cohort (prop-abc) study: Design of lifestyle intervention. *BMJ Open Diabetes Research & Care*, 8(1):e000899, 2020. doi: 10.1136/bmjdrc-2019-000899. URL https://drc.bmj.com/content/8/1/e000899.

[243] H. F. Jelinek, C. Helf, and K. Khalaf. Human shc-transforming protein 1 and its isoforms p66shc: A novel marker for prediabetes. *Journal of Diabetes Investigation*, 12(10):1881–1889, 2021. doi: 10.1111/jdi.13551. URL https://pubmed.ncbi.nlm.nih.gov/33759377/.

[244] Z. Li, Y. Zhang, M. Hoene, L. Fritsche, S. Zheng, A. Birkenfeld, A. Fritsche, A. Peter, X. Liu, X. Zhao, L. Zhou, P. Luo, C. Weigert, X. Lin, G. Xu, and R. Lehmann. Diagnostic performance of sex-specific modified metabolite patterns in urine for screening of prediabetes. *Frontiers in Endocrinology*, 13: 935016, 2022. doi: 10.3389/fendo.2022.935016. URL https://pubmed.ncbi.nlm.nih.gov/35909528/.

[245] M. Guasch-Ferré, A. Hruby, E. Toledo, C. B. Clish, M. A. Martínez-González, J. Salas-Salvadó, and F. B. Hu. Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. *Diabetes Care*, 39(5):833–846, 2016. doi: 10.2337/dc15-2251. URL https://doi.org/10.2337/dc15-2251.

[246] H. Lee, T. Park, J. Kim, J. H. Yun, H. Yu, Y. Kim, and B. Kim. Identification of metabolic markers predictive of prediabetes in a korean population. *Scientific Reports*, 10(1):1–9, 2020. doi: 10.1038/s41598-020-78961-4. URL https://doi.org/10.1038/s41598-020-78961-4.

[247] Y. Sun, Z. Wang, Z. Huang, H. Hu, and Y. Han. The association between the triglyceride-to-high-density lipoprotein cholesterol ratio and the risk of progression to diabetes from prediabetes: A 5-year cohort study in chinese adults. *Frontiers in Endocrinology*, 13:947157, 2022. doi: 10.3389/fendo.2022.947157. URL https://pubmed.ncbi.nlm.nih.gov/35923622/.