Genomic Insights through Association Rule Mining for Gastroesophageal Reflux Disease and Diabetes Mellitus

Submitted by: Aqsa Fatima

Registration No: 59-FOC/MSBI/F22

Supervised by Dr. Mehrosh Khalid

Department of Bioinformatics,

Faculty of Computing and Information Technology,

International Islamic University Islamabad.

2025

A dissertation submitted to Department of Bioinformatics, International Islamic University, and Islamabad as partial fulfillment of the requirements for the award of the degree of MS in Bioinformatics

DEDICATION

My humble effort I dedicate to my lovely family and to my respectable supervisor.

Thanks to all of you!

Declaration

I hereby solemnly state that this thesis work “Genomic Insights through Association Rule Mining for Gastroesophageal Reflux Disease and Diabetes Mellitus”, nor overall neither as a section has been replicated from any source. I have done this research entirely based on my efforts and the proficient supervision of my research supervisor.

_____

Aqsa Fatima

59-FOC/MSBI/F22

Acknowledgement

As in Holy Quran, “The seven heavens and earth and all that is therein, glorify HIM and there is not a thing but glorifies HIS praise”.

I therefore, start my acknowledgement as a word of thanks to Almighty ALLAH, the Exalted, the Gracious and the Merciful, and to whom I owe my very existence. HE bestowed upon me the strength, wisdom and perseverance to complete my research work successfully. I am nothing without my ALLAH but can achieve everything with HIS assistance. I present my heartfelt gratitude from the bottom core of my heart to my Holy Prophet Hazrat Muhammad ﷺwhose life is an eternal source of guidance for me and for the entire world. Hisﷺ teachings always show me the right path in time of difficulty.

I express my sincere and respectful appreciation to my worthy supervisor Dr.Mehrosh Khalid for her scholarly guidance, mentorship and vast knowledge that helped me to embark upon this highly important work. She set her standards for her students and she not only encourages but also guides them to meet those standards. It was a great privilege and honor to work under her supervision. I am ever indebted and obliged to her.

Where would I be without my family? Words wane in expressing me veneration for my loving Ammi jee and Abu jee; I owe my heartiest gratitude for their assistance and never ending prayers for my success. Especially my father is the greatest blessing, which I am having in life. Abu jee is always there for me in any hurdle. He is the main reason behind my sparkling eyes and driving force for my success. Words can never explain my love and respect for my father. I highly commend the cooperative behavior of my siblings Dr. Sumaira Jabeen, M.Abdullah Ashraf and M. Abdul Rehman for their much needed support, patience understanding and encouragement in every possible way that I can say. Their endless love, priceless, perpetual, indispensable help, support and everything made all this possible.

I wish to express my special thanks to my friends whose company made me stay in university full of joys with everlasting memories.

PROJECT IN BRIEF

Project Title:	Genomic Insights through Association Rule Mining for Gastroesophageal Reflux Disease and Diabetes Mellitus.
Organization:	Department of Bioinformatics Faculty of computing and Information Technology International Islamic university H-10, Islamabad
Undertaken by:	Aqsa Fatima 59-FOC/MSB/F22
Supervised by:	Dr. Mehrosh Khalid
Start Date:	2023
Completion Date:	2025
Objective:	The aim of this study is to identify association between diabetes and gastroesophageal reflux disease through Bioinformatics analysis and Association rule mining.
Tools & Technologies:	RStudio, GEO2R, EnrichR, STRING, GENEMANIA.
Documentation Tool:	MSWord, MS PowerPoint
Operation System:	Windows 10
System Used:	Haier corei5

Table of Contents

List of Abbreviations

Acronym	Abbreviations	Acronym	Abbreviations
DM	Diabetes Mellitus	**T2DM**	Type 2 diabetes mellitus
GERD	Gastroesophageal reflux Disease	**LRYGB**	laparoscopic Roux-en-Y gastric bypass
DEGs	Differentially expressed genes	**PPI**	Protein-protein interactions
ARM	Association rule mining	**AXL**	AXL receptor tyrosine kinase
LES	Lower esophageal sphincter	**PTGER2**	Prostaglandin E receptor 2
TLESRs	transient lower esophageal sphincter relaxations	**GBP1**	Guanylate binding protein 1
T1DM	Type 1 diabetes mellitus	**LGALS3**	Lectin galactoside – binding soluble 3
GO	Gene ontology	**BE**	Barret's esophagus
KEGG	Kyoto Encyclopedia of Genes and Genomes	**GEO**	Gene expression omnibus
DisGeNet	Discovery of human Disease-GENE associations	**CHI3L1**	Chitinase-3-like protein 1

List of Figures

List of Tables

Abstract

Gastroesophageal reflux disease and diabetes mellitus are prevalent chronic conditions with significant global health implications. Studies suggest a strong association between the two, with DM patients exhibiting a 61% higher risk of GERD due to factors such as delayed gastric emptying, autonomic neuropathy, and poor glycemic control. This study aims to determine shared gene expression profiles and signaling mechanisms that regulate the relationship between GERD and DM. Using gene expression datasets from the Gene Expression Omnibus (GEO), we identified 47 DEGs in GERD (GSE9768) and 17 in DM (GSE161355) by applying a rigorous statistical threshold ($p < 0.002$). To identify gene co-expression patterns, we employed an unsupervised machine learning technique, the Apriori algorithm of association rule mining, combined with Jaccard similarity, with minimum support (0.2) and confidence (0.8). This approach revealed key genes including AXL, PTGER2, LGALS3, GBP1, and CHI3L1, which appeared in multiple rules, indicating their potential involvement in both diseases. Gene ontology and pathway enrichment analysis highlighted shared biological processes, particularly related to inflammation and prostaglandin signaling. Protein-protein interaction networks and gene co-expression analysis further supported the involvement of these genes in immune regulation and tissue remodeling. Gene-disease association analysis identified strong connections between the identified genes and both GERD and DM, offering potential therapeutic targets for further investigation. These findings suggest that common molecular mechanisms, particularly inflammatory pathways, underlie the coexistence of GERD and DM, and propose these genes as biomarkers for diagnosis and treatment strategies. Future studies should focus on experimental validation to confirm these results and explore potential therapeutic interventions.

Chapter 1

Introduction

1 Introduction

This section discussed the background and significance of the study, with the emphasis on the clinical and molecular aspect of gastroesophageal reflux disease and diabetes mellitus. It outlines the rationale for exploring their genetic linkage and establishes the foundation for applying computational approaches to identify shared disease mechanisms.

1.1Background of Study

Diabetes mellitus (DM) and gastro-esophageal reflux disease (GERD) are two chronic conditions that are common throughout the world and linked to serious morbidity and long-term consequences. Heartburn, regurgitation, chest pain, and in certain situations, consequences like Barrett's esophagus and esophageal cancer, are symptoms of GERD, a disorder marked by the retrograde migration of stomach contents into the esophagus. Chronic hyperglycemia brought on by insulin resistance or insufficiency is a hallmark of DM, a metabolic disease that can lead to retinopathy, nephropathy, cardiovascular disorders, and neuropathy.

Multiple clinical studies and systematic reviews have highlighted the increasing coexistence of GERD and DM in affected populations, particularly in type 2 diabetic individuals. For instance, studies report GERD symptoms in approximately 25–40% of diabetic patients, significantly higher than the general population. This association raises concern as GERD can negatively affect the quality of life and worsen metabolic control in diabetics. Moreover, evidence suggests that under diagnosis of GERD in diabetic patients due to overlapping autonomic neuropathy-related symptoms, often leading to silent esophageal injury.

1.2 Gastroesophageal Reflux Disease

Unusual reflux of stomach contents into the esophagus is a common gastrointestinal condition that can cause problematic symptoms and even problems. It effects millions of people around the world; estimates indicate that approximately 60 million Americans frequently suffer from GERD symptoms [1]. In addition to having a major impact on patients' quality of life, the illness is linked to high healthcare expenses, more than $75 billion is spent on it each year in the United States alone [2]. According to the Montreal Definition and Classification, GERD is a disorder that arises when reflux of stomach contents results in bothersome symptoms or complications. These

symptoms can include both common ones like heartburn and regurgitation as well as uncommon ones like laryngitis and persistent cough [1].

Lower esophageal sphincter (LES) dysfunction, elevated stomach acid production, and changes in esophageal motility are some of the intricately interacting variables that contribute to the pathophysiology of GERD. TLESRs, or transient lower esophageal sphincter relaxations, are thought to be the main physiological process causing acid regurgitation. Furthermore, dietary practices and lifestyle factors like obesity have been linked to the onset and aggravation of GERD symptoms. The genetic foundations of GERD have also been brought to light by recent research, which has used genome-wide association studies to identify links with psychological characteristics and obesity [3].

Given its significant prevalence and impact on health, effective management of GERD is critical. Treatment strategies typically include lifestyle modifications, pharmacotherapy with proton pump inhibitors (PPIs), and in some cases, surgical interventions. Despite the availability of various treatment options, many patients continue to experience refractory symptoms, necessitating ongoing research into the underlying mechanisms and innovative therapeutic approaches [4]. Overall, understanding GERD's multifaceted nature is essential for improving patient outcomes and minimizing its burden on healthcare systems.

1.2.1 Symptoms of GERD

A common gastrointestinal illness that is defined by the repeated regurgitation of foodstuff in the abdomen to the esophagus, which causes bothersome symptoms and possible complications. Heartburn is the most widespread symptom of GERD, which is usually referred to as a burning feeling in the chest and may spread to the neck and the throat. This feeling is also experienced after a meal and it might be aggravated by lying down or bending forward. In addition to heartburn, patients suffer regurgitation, which refers to the feeling of sour or bitter liquid regurgitation to the throat or mouth [5].

Additional signs of GERD are non-burning chest pain, dysphagia and Globus sensation. Patients with chronic cough, hoarseness, sore throat, and acid exposure on the teeth might also experience other unusual cases. The symptoms during nighttime may cause persistent cough or laryngitis and aggravate such conditions as asthma.

1.3 Diabetes Mellitus

A complicated metabolic condition with far-reaching ramifications for world health. It is primarily distinguished by chronic hyperglycemia caused by abnormalities in insulin secretion, insulin action, or both. The condition is divided into numerous categories, with Type 1 diabetes mellitus (T1DM) and Type 2 diabetes mellitus (T2DM) being the most common.

1.3.1 Type I Diabetes Mellitus (T1DM)

Type 1 diabetes is an autoimmune disease in which the immune system assaults and destroys insulin-producing beta cells in the pancreas. This leads in an absolute insulin shortage, which causes high blood glucose levels. T1DM often begins in infancy or adolescence, but it can develop at any age. This kind of diabetes develops due to a genetic predisposition as well as environmental influences [6].

1.3.2 Type II Diabetes Mellitus (T2DM)

Type 2 diabetes accounts for 90-95% of all diabetes cases and is characterized by insulin resistance. This syndrome frequently occurs in adulthood, although increased incidence are being detected in younger populations as obesity and sedentary lifestyles become more prevalent. T2DM develops gradually and can be influenced by genetics, lifestyle choices, and environmental variables. Lifestyle changes, oral medicines, and, on occasion, insulin therapy are used to manage the condition [6].

Diabetes has become a global epidemic. According to the World Health Organization, an estimated 422 million individuals had diabetes in 2014, and forecasts indicate that this figure could climb dramatically due to factors such as urbanization and ageing populations. The rise in obesity rates is especially worrying because it is closely linked to the development of T2DM. Diabetes pathogenesis is complicated, involving genetic predisposition and environmental causes.

In T1DM, autoimmune death of pancreatic beta cells causes a shortage of insulin production. Insulin resistance and, eventually, beta-cell failure, on the other hand, characterize T2DM. Chronic hyperglycemia can lead to significant consequences affecting multiple organ systems, including cardiovascular disease, neuropathy, nephropathy, and retinopathy [7]. Understanding diabetes mellitus is critical to creating effective preventative and management measures. With its

expanding prevalence worldwide, there is an urgent need for public health efforts that attempt to educate consumers about risk factors and promote healthier lifestyles.

1.3.3 Symptoms of Diabetes Mellitus

Diabetes mellitus causes a variety of symptoms that differ in degree and onset between individuals, mostly dependent on the type of diabetes. Common symptoms include excessive thirst (polydipsia), frequent urine (polyuria), severe hunger (polyphagia), and unexpected weight loss, all of which are caused by the body's inability to adequately use glucose for energy, resulting in high blood sugar levels. People may also feel weariness, hazy vision, and slow-healing wounds, as well as recurring infections and skin problems like itching or thrush. Type 1 diabetes symptoms frequently appear quickly and may involve additional indicators such as nausea and stomach pain caused by diabetic ketoacidosis, whereas Type 2 diabetes symptoms can be more gradual and subtle, often going unrecognized for years [8]. Recognizing these signs is critical for early detection and treatment to avoid major complications associated with uncontrolled diabetes.

1.4 History

The understanding of DM and GERD has evolved significantly over time, with early documentation of DM in ancient Egyptian texts and GERD being recognized in the context of digestive disorders since ancient times. Both diseases have since become the focus of extensive research due to their growing prevalence and clinical impact.

1.4.1 History of GERD

GERD has evolved significantly since its identification as a unique clinical entity in the 1970s. In 1976, Krejs et al. introduced the term "gastro-esophageal reflux disease" to the English language, thereby acknowledging GERD as a major health issue. GERD was initially studied primarily from a surgical standpoint, but it later received attention from gastroenterologists, resulting in a more complete understanding of its pathogenesis and clinical implications [9]. Over the decades, GERD has become one of the most common chronic gastrointestinal ailments, influencing about 60 million Americans and more than a billion people worldwide. The burden of GERD has grown significantly, with healthcare expenses exceeding $75 billion per year in the United States alone [10]. Treatment advances have included the introduction of PPIs and various surgical treatments, indicating an increasing acknowledgement of GERD's impact on quality of life and health. As

research advances, understanding the history of GERD is critical to improving management options and patient outcomes [11].

1.4.2 History of Diabetes Mellitus

Diabetes mellitus has been around for thousands of years, with the first recorded observations extending back to ancient civilizations. In the first century AD, the physician Aretaeus of Cappadocia coined the name "diabetes," which comes from the Greek word "syphon," to describe a disorder characterized by excessive urine. In 1674, Thomas Willis added "mellitus," which means "honey-sweet," after noticing the sweet taste of diabetic urine, which had been reported in Egyptian medical books as early as 1550 BC. Diabetes therapies have historically been rudimentary and sometimes ineffectual, ranging from starvation diets to opium use. In 1889, German researchers Joseph von Mering and Oskar Minkowski showed the link between the pancreas and diabetes by surgically removing it from dogs, resulting in diabetic symptoms. This finding cleared the path for Frederick Banting and Charles Best's isolation of insulin in the early 1920s, which revolutionized diabetic therapy. Since then, advances in glucose monitoring devices and insulin delivery systems have substantially improved diabetes treatment and results [12].

1.5 Prevalence of Gastroesophageal Reflux Disease and Diabetes Mellitus

GERD is a common digestive condition that has a substantial impact on quality of life and healthcare systems around the world, with prevalence varied according to area and demographic factors. A review of 30 studies found that GERD prevalence ranges from 18% to 28% in North America, 9% to 26% in Europe, 3% to 8% in East Asia, 9% to 33% in the Middle East, and 12% in Australia, with a significant increase since 1995. Recent research has found a prevalence as high as 44.8%, with regurgitation being the most prevalent symptom (4). Age is an important issue, as older adults had greater rates of GERD symptoms, with prevalence rising from 4.4% in people under 20 to more than 21 percent in those older than 80. Obesity, smoking, and dietary habits are all risk factors for GERD, with obesity increasing the incidence by 1.73 times. As public awareness of GERD rises, knowing its epidemiology becomes critical for successful care and preventative initiatives [13].

Diabetes mellitus has reached alarming levels globally; affecting roughly 537 million adults aged 20 to 79 years as of 2021, with forecasts indicating that this figure would climb to 643 million by

2030 and 783 million by 2045. In the United States, the prevalence of diagnosed diabetes among adults was reported at 11.3% in 2021, with significant variation among age categories; for example, it was 3.0% for those aged 18-44 and 24.4% for those aged 65 and more (14). Diabetes prevalence has continuously increased worldwide, particularly in low- and middle-income countries, where it accounts for around 80% of cases. This tendency is mostly due to lifestyle issues such as poor nutrition and physical inactivity, which are exacerbated by socioeconomic challenges [15].

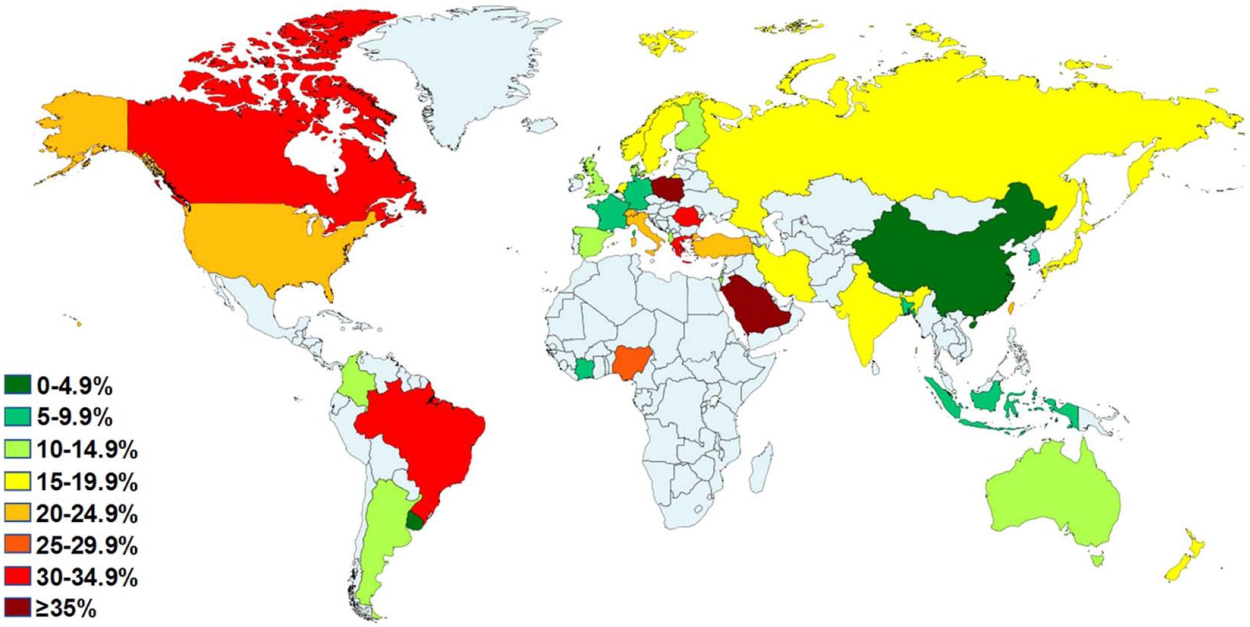


Figure 1.1: Global Prevalence of GERD by country [13].

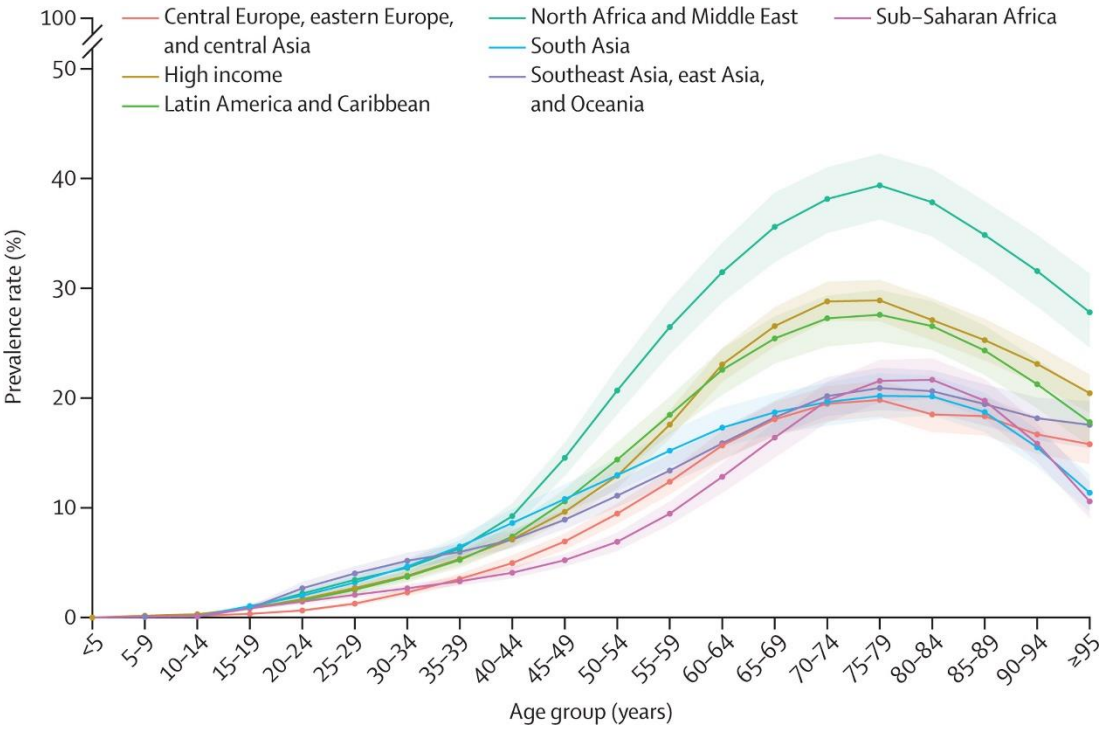


Figure 1.2: Age-Specific prevalence rate of Diabetes across global region, showing highest rate in older adults, particularly in Middle East and high-income countries [15].

1.6 Diagnosis of GERD and DM

A thorough review of clinical symptoms and medical history is usually required to diagnose gastro-esophageal reflux disease. Patients frequently appear with familiar symptoms such heartburn and regurgitation, that can lead to an empirical diagnosis and therapy without comprehensive testing. However, if symptoms are atypical, severe, or unresponsive to first treatment, more diagnostic testing may be necessary. Upper endoscopy is a common diagnostic procedure that allows for direct visualization of the esophagus and can detect issues such as esophagitis and Barrett's esophagus. Another crucial test is esophageal pH monitoring, which measures acid exposure in the esophagus over time and can be conducted with either a catheter or a wireless capsule. Additional tests, such as esophageal manometry, measure motility abnormalities and aid in the diagnosis of other illnesses that may mimic GERD. Barium esophagrams can also be used to assess structural problems; however, they are not typically indicated as a main diagnostic technique for GERD. Overall, combining clinical examination with focused diagnostic tests is critical for correctly diagnosing GERD and selecting the best therapeutic strategy [16].

Diabetes mellitus is diagnosed using several important tests that evaluate blood glucose levels and assess the body's ability to regulate glucose. The A1C, FPG, and OGTT are the three most widely utilized assays. The A1C test evaluates average blood glucose levels over the previous two to three months, and a result of 6.5% or higher indicates diabetes. The FPG test requires at least eight hours of fasting before measuring blood glucose; a result of 126 mg/dL or higher indicates diabetes. The OGTT entails fasting overnight, followed by the consumption of a glucose-rich beverage, with blood samples collected at regular intervals; a two-hour blood glucose level of 200 mg/dL or greater suggests diabetes. Furthermore, a random plasma glucose test can detect diabetes if the result is 200 mg/dL or greater, especially if it is accompanied by, traditional symptoms like increased thirst and frequent urination. Unless there are evident signs of hyperglycemia, any diagnosis should be confirmed with additional tests on a different day [17].

1.7 Association between Gastroesophageal Reflux Disease and Diabetes Mellitus

Recent studies have established a notable association between both diseases, particularly Type 2 diabetes. A study conducted in Shanghai, China, found that 16% of patients with Type 2 diabetes reported typical GERD indications, indicating a higher prevalence compared to the general

population [18]. This research highlights that while patients with diabetes are known to have increased rates of GERD, the underlying risk factors remain poorly understood. Another investigation revealed that 40% of asymptomatic diabetic patients exhibited signs of GERD upon further examination, emphasizing the need for proactive screening in this demographic. The mechanisms linking these conditions may involve obesity, which is prevalent among those with Type 2 DM, as well as autonomic neuropathy that can affect gastrointestinal motility and esophageal function. Furthermore, a meta-analysis indicated that people with diabetes are more susceptible to GERD due to shared pathophysiological factors such as insulin resistance and obesity. These findings underscore the importance of monitoring gastrointestinal symptoms in diabetic patients to improve their overall management and quality of life [19].

1.8 Problem Statement

Despite the frequent coexistence of gastroesophageal reflux disease and diabetes mellitus, their genetic and molecular links remain poorly understood. Most genomic studies have investigated these conditions in isolation, lacking the use of integrative computational methods to identify shared biological patterns. Without applying advanced techniques such as association rule mining, critical insight into common gene expression profiles, functional networks and potential therapeutic targets.

1.9 Research Questions

- What are the common candidate genes that are associated with both gastroesophageal reflux disease and diabetes mellitus?
- How the identified candidate genes can contribute to the pathogenesis and underlying mechanism of both diseases?
- What are the specific biological pathways shared between GERD and DM indicted by common candidate genes?

1.10 Aims and Objectives

The focus of our research is to investigate the common candidate genes of GERD and DM and asses their genetic association for preliminary identification and individual- based curative treatment for GERD with DM.

This research outlines the following main objectives:

- To identify the common candidate genes that exhibit associations with both gastroesophageal reflux disease and diabetes mellitus.
- To explore potential genetic links between gastroesophageal reflux disease and diabetes mellitus to uncover common candidate genes involve in the pathogenesis of both diseases.
- To provide valuable insight into shared genetic mechanism and potential therapeutic targets.

1.11 Proposed Solution

To explore the common genetic factors between gastroesophageal reflux disease and diabetes mellitus, multivariate covariance analysis, association rule mining and pathway enrichment analysis is proposed. This research strategy aims to understand the shared molecular mechanisms that contribute to the coexistence of GERD and DM and provides valuable understanding of complex relationships between these two diseases in a broader biological context.

1.12 Overview of Computational Approach

A thorough computational model was used to explore the common molecular pathways between GERD and DM. A framework combines statistical modeling, pattern discovery, and systems biology to derive meaningful knowledge in high-throughput gene expression data. It allows identifying differentially expressed genes, finding co-regulated gene sets, and interpreting functional relationships via enrichment and network analysis. The main aspects of this strategy are as follows:

1.12.1 Association Rule Mining (ARM)

ARM identifies co-occurring items in large datasets. In genomics, ARM reveals relationships between gene expression patterns and clinical variables. This method, often used in market basket analysis, has proven effective in identifying gene clusters that co-express under similar conditions [20].

1.12.2 Multivariate Covariance Analysis (MANCOVA)

MANCOVA evaluates group differences across multiple dependent variables simultaneously, allowing detection of subtle correlations between gene expression patterns across disease states. It adds robustness by accounting for variance and interdependencies between variables [21].

1.12.3 Pathway Enrichment Analysis

To contextualize gene-level findings, pathway enrichment tools such as EnrichR, STRING, and GeneMANIA are employed. These tools map identified genes to biological functions, interactions, and cellular processes, offering insight into shared disease mechanisms.

1.13 Scope & Limitations

The Scope of this research is to use computational methods to explore the genomic relationships between DM and GERD, with a particular emphasis on finding shared candidate genes that are important in the etiology of both diseases. The goal of the research is to identify common molecular processes and biological pathways that connect various disorders using multivariate covariance analysis, association rule mining, and pathway enrichment analysis. This will provide important information about possible therapeutic targets.

However, the study does, have certain limitations. Large and complicated datasets may be difficult for the computational tools such as the arules and gcrma packages to process effectively. Furthermore, the quality and accessibility of publically available datasets, such those from the GEO database, may limit the identification of all significant genes, so compromising the accuracy of the findings. Further experimental validation is necessary to define the biological importance of these genes in the context of GERD and DM, even though association rule mining can indicate genetic connections. Furthermore, by concentrating solely on genetic factors, the study may ignore the impact of lifestyle, environment, and other non-genetic factors that contribute to the co-occurrence of GERD and DM. These non-genetic factors may be crucial in the interaction and development of both diseases.

Chapter 02

Literature Review

2 Literature Review

Diabetes mellitus and Gastroesophageal Reflux Disease are two common chronic disorders that present a major health challenge to the world. Their comorbidity has gained more and more attention because of common risk factors, possible pathophysiological relationships, and the presence of emerging data on genetic interactions. A number of epidemiological, clinical and molecular studies have examined the relationship between GERD and DM, to determine whether diabetes causes an individual to develop GERD or there is a bilateral relationship between the two. This study reviews existing research on the association, focusing on prevalence, risk factors, underlying biological mechanism and genetic evidence**.**

2.1 Epidemiological Association between GERD and DM

Sun *et al.* [19] performed the wide scope meta-analysis to investigate the connection between DM and GERD. The study examined nine studies that included more than 90,000 participants and identified a statistically significant relationship between DM and the higher risk of GERD, with the overall odds ratio (OR) of 1.61. Subgroup analyses indicated that it was a stronger association in the group below 50 years of age and in the Asian population. The results indicate that diabetic complications like autonomic neuropathy, delayed gastric emptying and obesity are some of the possible causes in the pathophysiology of GERD among diabetic patients.

Natalini *et al* [22]. Further supported this relationship by identifying diabetes as predictor variable for GERD in African American populations. Similarly, Lee *et al*. found that diabetes patients, both with and without peripheral neuropathy had a higher prevalence of GERD than non-diabetic counterparts did. These findings highlight the necessity of proactive screening for GERD symptoms in diabetics.

A retrospective cross-sectional study conducted by Chang *et al*. [23] on more than 5,500 patients in Taiwan investigated the risk factors related to the development of diabetes among patients diagnosed with GERD. The researchers found out that 13.2 percent of patients with GERD also had a diagnosis of DM. It found older age, obesity, and insufficient physical activity along with family history to be critical predictors, indicating that the components of metabolic syndrome are crucial in the connection between them.

Lin *et al.* [24] conducted a cross-sectional study to assess GERD symptoms and esophageal

findings in diabetic patients on symptom scores and endoscopic analysis. The findings indicated a greater incidence of silent GERD in diabetics, which underscores the importance of endoscopic assessment of these patients.

2.2 Pathophysiological Mechanism Linking GERD and DM

2.2.1 Autonomic Neuropathy and Esophageal Dysfunction

Fujiwara *et al.* [25] carried out a clinical study that utilized trans-nasal endoscopy to assess GERD among patients with type 2 diabetes. They discovered that 42.1 percent of diabetic patients had GERD with majority of them having mild esophagitis. Younger and heavier diabetic patients had a higher prevalence of GERD and a large proportion of them had no symptoms indicating an esophageal dysfunction.

Lorentzen *et al.* [26] investigated the prevalence of GERD among morbidly obese patients with and without diabetes type 2. Though there was no significant difference in the prevalence of symptoms, more than half of diabetic individuals had acid reflux that was pathological. Most were asymptomatic possibly due to diabetic neuropathy and esophageal hyposensitivity.

2.2.2 Hormonal and Metabolic Dysregulation

Kumar *et al* [27]. Performed a systematic review to examine the relationship between GERD and DM in both directions and the potential common pathophysiological processes and approaches to co-management. The review revealed delayed gastric emptying (gastroparesis) and autonomic neuropathy as key contributors in the connection between the two conditions. These processes lead to poor esophageal motility and acid reflux, which are common in diabetic patients. Another point made in the review was that the clinical management of one condition usually influences the other-such as how some GERD medications can interfere with glycemic control and diabetic neuropathy can hide GERD symptoms and underdiagnoses. The seven studies have been reviewed, which include trials of prokinetic such as itopride and GLP-1 receptor agonists, and surgical procedures such as laparoscopic Roux-en-Y gastric bypass (LRYGB) that showed better weight loss and reflux control outcomes than other bariatric surgeries. The authors promote a multidisciplinary and individual approach to treatment, which implies pharmacological, surgical, and lifestyle interventions depending on the needs of diabetic patients with GERD. The review acknowledges the strengths but identifies limitations in sample diversity and long-term follow-up, which suggests

additional randomized trials and longitudinal studies to inform the best management practices and enhance patient outcomes

2.3 Clinical and Lifestyle Risk Factors

Dixon *et al* [28]. Have studied the association between diabetes mellitus and esophageal adenocarcinoma, particularly in the U.S. veterans. They discovered that diabetic patients experienced a far higher rate of developing esophageal cancer than non-diabetes patients, despite the obesity factor being kept in check. This posits a possible pathophysiological connection between oxidative stress associated with hyperglycemia and cancer of the esophagus lining. Although GERD is already known as a risk factor in developing esophageal cancer, the study further proposes that diabetes has the ability to independently increase this risk, through causing or enhancing esophageal inflammation or changes in the healing response. The study emphasizes the value of periodic endoscopic monitoring of diabetic patients with chronic reflux symptoms to detect pre-cancerous alterations at an early stage.

2.4 Genetic Insights and Causal Inference

In a complementary study by Shuai and Larsson [29], genetically predicted BMI (OR = 1.49) and T2DM (OR = 1.07) were both related to the higher risk of GERD. Causal risk factor was also determined as smoking initiation, but alcohol and coffee consumption were not significantly connected. In general, the literature has been consistent in showing that DM is a risk factor of GERD. Autonomic neuropathy, obesity, hormonal imbalance, and genetic factors play a role in this association. Additionally, diabetic patients have high subclinical GERD, and therefore, their gastrointestinal evaluation should be part of diabetes management. Table 2.1 summarize the most significant findings.

2.5 Research Gap

There is a limited understanding of genetic interactions associated with GERD and DM, and their genetic links not studied. There is need for integrated approaches like association rule mining approach and MANCOVA analysis, in order to identify the pathophysiology and candidate genes of GERD & DM. Comprehensive understanding of complexity of disease etiology is required that underlying the common genetic basis that provide insights into shared pathways and mechanisms of both conditions. Also need for the understanding of the genetic basis which underlying the co-

occurrence of GERD and DM and have significant clinical implications. Identification of common candidate genes may lead to the development of novel therapeutic targets or personalized treatment strategies that address the shared genetic vulnerabilities of both diseases.

2.6 Critical Analysis

Recent advances in Mendelian randomization have offered valuable causal insights into the relationship between GERD and diabetes mellitus. However, these approaches primarily focus on individual gene effects and fail to account for gene-gene interactions or co-expression networks, which are critical in understanding the complexity of multifactorial diseases. To address this limitation, the present study applies association rule mining, a novel data mining technique used to uncover the hidden genomic patterns and shared expression signatures between GERD and DM.

Creighton and Hanash [30] used the Apriori algorithm on yeast expression data and were able to mine association rules that indicated co-expression between genes in certain conditions. They showed that randomized datasets did not give any meaningful rules, which confirmed the statistical power of ARM. Anandhavalli *et al* [21] noted that ARM has the capacity to detect condition-specific interactions of genes in large data. The algorithm identifies novel associations in genomics by treating gene expressions as item sets and uncovers hidden association inaccessible to traditional statistical approaches. The review determines that ARM has potential to identify candidate biomarkers and regulatory modules in genomic studies.

ARM identifies frequent gene expression rules and conditional dependencies, enabling the detection of co-regulated genes that conventional methods, such as differential expression analysis or GWAS, often overlook. This method supports a more comprehensive understanding of disease comorbidity at the molecular level. This study integrates association rule mining (ARM) with tools like GeneMANIA and STRING, and applies algorithms such as Apriori to highlight novel regulatory interactions and potential biomarkers. This approach defines previously unrecognized molecular relationships and functional networks, offering new directions for biomarker discovery and personalized medicine.

Table 2.1: The summary of the Literature Review

S.No	Authors	Study Type	Key Findings
1.	Sun *et al*	Meta- analysis (n>90,000)	DM was related to 1.61x increased risk of GERD, with the greatest effect being in <50 and Asian populations
2.	Natalini *et al*	Population study(African Americans)	DM is an independent predictor of GERD
3.	Chang *et al.*	Cross-sectional study (n=5,500 +)	13.2% GERD patients had DM: older age, obesity, inactivity & family history linked.
4.	Lin *et al.*	Cross sectional Study and Endoscopy study	Higher incidence of silent GERD in diabetic patients.
5.	Fujiwara *et al.*	Clinical endoscopy study	42.1 % diabetic patients have GERD; many show asymptomatic symptoms.
6.	Lorentzen *et al.*	Cross sectional study (morbid obesity)	It finds that pathological acid reflux is common in diabetic patients, mostly asymptomatic.
7.	Kumar *et al.*	Systematic review	Shared mechanism include gastroparesis and neuropathy; it recommend multidisciplinary care
8.	Dixon *et al.*	Veteran cohort study	DM independently increased the risk of esophageal cancer in GERD patients.
9.	Larsson *et al.*	Mendelian randomization study	Genetically predicted BMI and T2DM increase the GERD risk whereas smoking is a causal factor of GERD.
10.	Anandhavalli *et a.l*	Review of ARM in genomics	ARM detects hidden gene interactions that are valuable for biomarker discovery

Chapter 03

Materials &Methods

3.1 Study Design

This study utilized computational bioinformatics approach to analyze gene expression data from GERD and DM using datasets from GEO database. After preprocessing and identified DEGs, association rule mining was applied to detect shared gene patterns. Functional analysis included GO and pathway enrichment using EnrichR, and gene-disease association were explored using CTD, PubMed and DisGeNet. PPI and co-expression networks were visualized using STRING and GeneMANIA.

3.2 Data Collection

Gene Expression datasets for Gastroesophageal Reflux Disease and Diabetes Mellitus were obtain from Gene Expression omnibus database. The dataset GSE9768 contains 11 GERD patient ad 2 healthy controls and GSE161355 dataset include 18 DM patients and 15 non-DM patients, both generated using the GPL570 platform [HG-U133_Plus_2] Affymetrix Human Genome U133 plus 2.0 Array. Table 3.1 describe the summary of datasets.

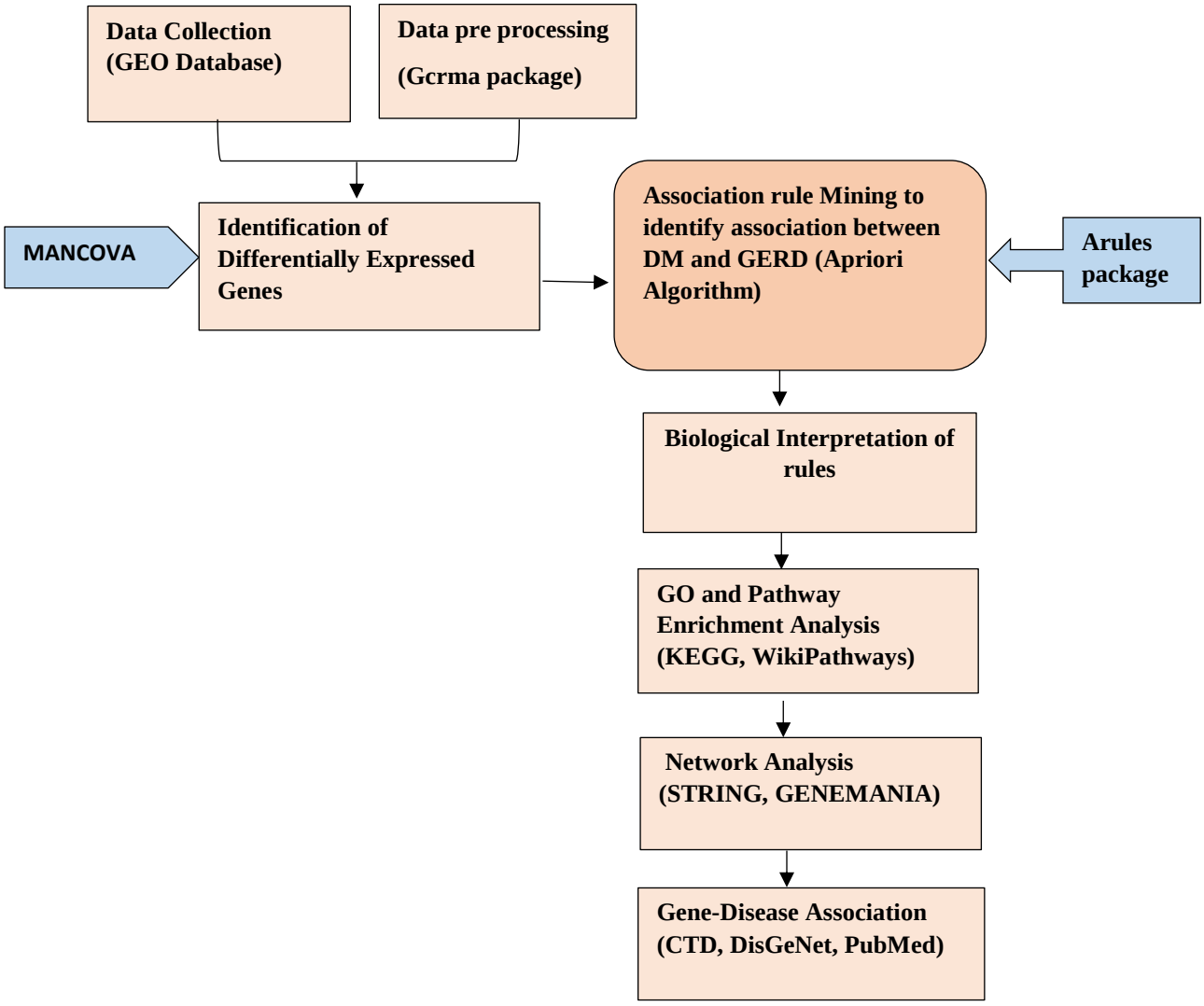


Figure 3.1 Methodological approach for genomic insight through ARM for GERD and DM

Table 3.1: Summary of Datasets

Dataset	Disease	Samples	Array	Platform	Probe ID	File	Summary
GSE161355	Diabetes Mellitus	33 Samples (15 Control Samples &18 DM Samples)	[HG-U133A_2] Affymetrix Human Genome Array	GPL570	54675 IDs	.CEL	The study design of the GSE161355 dataset involves profiling gene expression in cortical neurons, astrocytes, and endothelial cells from the neurovascular unit of aging brains. It includes 33 samples, with 18 from patients diagnosed with Type 2 diabetes mellitus and 15 from non-diabetic controls. This design enables the investigation of transcriptomic differences between diabetic and normal samples to understand T2D-related molecular changes in neurovascular cells.
GSE9768	Gastroesophageal Reflux Disease	13 Samples (2 Control Samples &11 GERD Samples)	[HG-U133A_2] Affymetrix Human Genome Array	GPL570	54675 IDs	.CEL	The study design of the GSE9768 dataset involves analyzing gene expression in a BE cell line (CP-A hTERT) exposed to acid and bile, and comparing RNA expression to controls using Affymetrix arrays. Additionally, Real-Time PCR was used to validate expression changes in 12 genes across biopsies from 110 patients with varying stages of reflux-related diseases.

3.3 Identification of Differentially Expressed Genes

DEGs are key indicators of altered transcriptional activity between healthy and diseased states. Identifying DEGs enables the characterization of disease specific molecular pathways and potential biomarkers. This process forms the foundation for understanding disease mechanisms at the gene expression level.

3.3.1 Data Visualization

We retrieved raw datasets for GERD (GSE9768) and DM (GSE161355) from the GEO repository and aligned them with their corresponding conditions. Probe intensity values were reviewed to ensure the correct mapping of sample to their respective conditions. Boxplot and histogram were generated to inspect expression distribution, detect outliers and check for batch effects. These plots guided subsequent normalization and analysis steps.

3.3.2 Data Preprocessing

Data preprocessing were conducted in R (v4.3.2) using Bioconductor package including affy, gcrma, genefilter and affyPLM. Background correction methods (MAS5, RMA, and GCRMA) and normalization techniques (MAS5, quantile) were applied, followed by log2 transformation. Genes with low expression or limited variability were filtered out to retain biologically meaningful features [32].

3.3.3 DEG Analysis

Differential expression was determined using multivariate analysis of covariance (MANCOVA), incorporating a model matrix to distinguish between groups. Genes meeting the statistical threshold ($p<0.002$) were classified as significant. Expression profiles were visualized through heatmaps to illustrate group specific patterns [33].

3.3.4 Annotations and Co-expression Analysis

Significant genes were annotated using the hgu133plus2 platforms to retrieve genes symbols and biological identifiers. Correlation analysis was applied to assess co-expression among DEGs. Covariance, dissimilarity, and interaction matrices were constructed to investigate functional gene-gene associations.

3.4 ARM for Identifying Shared Genes Expression Patterns between DM and GERD

Association Rule Mining (ARM) is a data mining technique used to uncover hidden relationship between variables in large datasets .In bioinformatics, ARM can be applied to gene expression data to detect co-occurrence patterns between genes under different biological conditions. It is especially effective in revealing complex gene interactions that may not be captured by traditional gene expression analysis [34]. In this study, ARM was employed to identify gene expression patterns shared between GERD and DM, based on differentially expressed genes from both datasets.

3.4.1 Data Import and Preprocessing

Gene expression matrices of GERD and DM were imported and formatted to ensure numerical consistency .columns with near-constant or irrelevant values were removed to improve data quality.to prepare the data for rule mining ,continuous expression values were discretized intro three ranked categories : Low-Medium, Medium and High.

3.4.2 Rule Generation Using the Apriori Algorithm

The Apriori Algorithm, implanted through arules package in R, was used to generate association rules separately for each datasets. A minimum support of 0.2 and minimum confidence of 0.8 were applied to extract robust and statistically significant rules. These rules were derived from the discretized expression data, capturing gene co-expression patterns within each condition.

3.4.3 Rules comparison and Similarity Assessment

To identify shared gene expression patterns between GERD and DM, rules from both datasets were compared using Jaccard similarity coefficient. Rules with similarity score below 0.2 were retained as biologically relevant overlaps. For each rule, metrics including support, confidence ,lift and count were calculated and stored in CSV format for downstream analysis.

3.4.4 Rules Filtering and Integration

To enhance the reliability of shared rules, only those with confidence above and equal to 0.8 in both datasets were selected. From each matching pair, the rule with higher confidence were retained to prioritize the robust association. This filtering step helped eliminate weak and redundant patterns.

3.4.5 Visualization of Common Gene Pattern

The top 20 most similar rules were visualized using bar plot to illustrate overlapping gene expression pattern across GERD and DM based on lift values. This visualization facilitated the interpretation of gene connectivity, highlighting potential co-regulated or functionally related gene across the two diseases.

3.5 Biological Interpretation of Common Rule

Biological interpretation of Common rule involved Gene disease association, ontology and pathway analysis and Network analysis. As these rules contain genes that will be common in both diseases based on support and confidence parameters of ARM.so there biological, interpretation is required to understand their shared molecular function, there association with disease and biological process in which these genes were involved, cellular components and their Networks that explain the connection of these genes with other genes.

3.5.1 Gene Disease Association

The identification of Gene-disease associations is fundamental in understanding the genetic etiology of complex disorders and their phenotypic outcomes. Such associations also underline the connection between particular genetic factors, including single nucleotide polymorphisms (SNPs), and disease phenotypes, which can form the basis of the strategic development of therapeutics and the maturation of precision medicine [35].

3.5.1.1 Importance of Gene - Disease Association

The gene-disease relationships are important in understanding disease pathology and possible biomarkers and therapeutic targets. These associations can help in developing personalized medicine where medicine is adjusted according to the genetic profile of the individual since they help in understanding how genetic differences promote the development of diseases and their progression. Such accuracy allows the improvement of interventions and risk prediction approaches in clinical practice.

3.5.1.2 Comparative Toxicogenomics Databases (CTD)

CTD [https://ctdbase.org/] is freely available database that curates data on the connections between genes, chemicals, phenotypes and diseases. CTD combines data on environmental exposure with

the data on molecular pathways to study the relationship between the environmental factors and the gene activities and predisposition towards diseases [36]. It offers:

- Manually determined gene disease and chemical disease connections
- Pathway and functional data integration to interpret environmental effects
- An exposome module on chemical phenotype and early exposure biomarkers

These qualities aid in the formulation of hypothesis on environmentally triggered illnesses and contributes towards early detection and prevention.

3.5.1.3 Disease Gene Network (DisGeNet)

A comprehensive information platform, DisGeNet [http://www.disgenet.org/] incorporates gene-disease and variant-disease connections from a variety of sources, such as curated databases, GWAS catalogues, and scholarly publications. The most recent version consists of:

- More than 24,000 illnesses and characteristics
- Details about 170,000 genomic variations and about 17,000 genes
- A modernized user interface featuring data prioritization tools and APIs

Through the unification of data across normal and abnormal phenotypes, DisGeNet facilitates investigation across a broad range of human disorders and allows for thorough analysis of disease mechanisms [37].

3.5.1.4 Integration into the Current Study

The gene-disease correlations of differentially expressed genes found in the GERD and DM datasets were examined in this study using both CTD and DisGeNet. These databases were utilized for:

- Connect DEGs to established disease correlations
- Examine how genetic and environmental factors affect disease processes.
- For GERD and DM, reinforce the biological interpretation of common gene expression patterns.

The findings of this study were more translationally relevant due to the combined application of CTD and DisGeNet, which facilitates a more thorough investigation of gene–environment interactions.

3.5.2 Gene Ontology and Pathway Enrichment analysis

To identify the shared function and pathways between diabetes and GERD, biological process, cellular component, and molecular functions and enrichment analysis (WikiPathways, Reactome, BioCarta, and (KEGG)) were carried out using the EnrichR online tool. The P value less than 0.05 was greatly enhanced.

3.5.2.1. Gene Ontology (GO)

GO analysis was also conducted through EnrichR to explore overrepresented biological functions, cellular components, and molecular processes associated with the set of differentially expressed genes (DEGs). Using GO's structured vocabulary, EnrichR provides insights into specific biological roles and localizations of genes within cells, further contextualizing their potential involvement in the condition studied. EnrichR compares the gene list with the GO terms in each category (BP, CC, and MF) and performs statistical tests, such as Fisher's exact test, to assess the importance of each GO term's enrichment. Multiple testing corrections, like the Benjamin-Hochberg adjustment, ensure that the identified GO terms are statistically reliable and not due to random chance. This analysis adds depth to functional interpretations, highlighting pathways and cellular functions where DEGs may play significant roles. By integrating GO terms with EnrichR's platform, this approach provides a comprehensive, function-based view of gene activity, aiding in understanding molecular mechanisms and revealing potential areas for further research.

3.5.2.2 Pathway Enrichment Analysis

KEGG, Reactome, BioCarta, WikiPathways Pathway enrichment analysis was performed using EnrichR, a widely used bioinformatics tool that facilitates the identification of overrepresented biological pathways among sets of genes or proteins. EnrichR integrates gene lists with multiple pathway databases, including KEGG, Reactome, WikiPathways, and BioCarta, to determine if certain pathways are statistically enriched within the gene set. After differential expression analysis identified genes of interest, these genes were input into EnrichR to assess their involvement in known pathways, allowing for a broader understanding of the biological processes potentially relevant to the study's condition. EnrichR applies statistical testing, such as Fisher's exact test, combined with adjustments for multiple comparisons (e.g., Bonferroni or Benjamin-Hochberg corrections), to ensure the robustness of enrichment results. This analysis enables the identification of biologically significant pathways, providing insight into coordinated gene

functions and revealing potential molecular mechanisms or therapeutic targets associated with the phenotype under investigation. EnrichR's streamlined access to comprehensive pathway databases makes it a valuable tool for pathway-centric interpretations in genomics and biomedical research [50].

3.5.3. Protein –Protein Interaction and Functional Network

Protein–Protein Interaction Networks (PPI) are essential for understanding the functional relationship between proteins within biological system. These networks provide insight into how protein work together to carry out cellular process and how disruptions in these interactions may contribute to disease mechanism. In a PPI network, proteins are represented as nodes, and interactions are represented as edges. Genes corresponding to highly connected nodes (hubs) are often considered biologically significant, as they may serve as a key regulators or points of convergence in disease pathways.

3.5.3.1. STRING Database for PPI Analysis

The protein–protein interactions of the shared genes between GERD and DM were examined through STRING database. To create a thorough interaction complex, STRING combines data from publicly accessible databases, computational prediction techniques, and experiments. Only interactions with strong biological importance were included after filtering using confidence values, which ranged from 4.00 (low confidence) to 9.00 (high confidence) [51].

3.5.3.2. GeneMANIA for Co-expression Network Construction

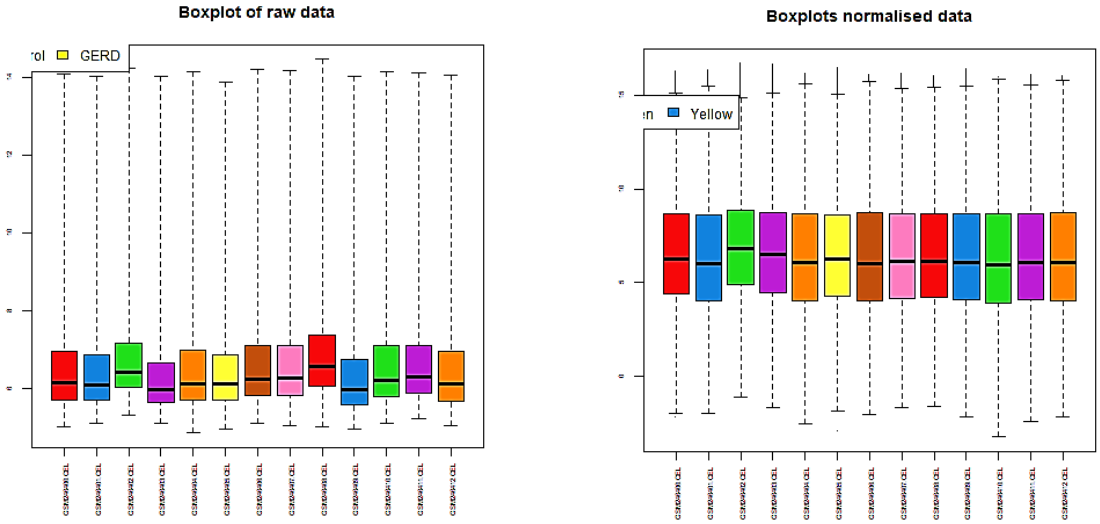
A co-expression network of the shared genes was constructed using GeneMANIA. This tool incorporates a variety of data kinds, such as shared pathways, physical interactions, genetic interactions, gene co-expression, and co-localization. By finding related genes and displaying the kind and degree of connections between them, GeneMANIA improves functional prediction and provides important information on biological connectivity and gene function [52].

Chapter 04

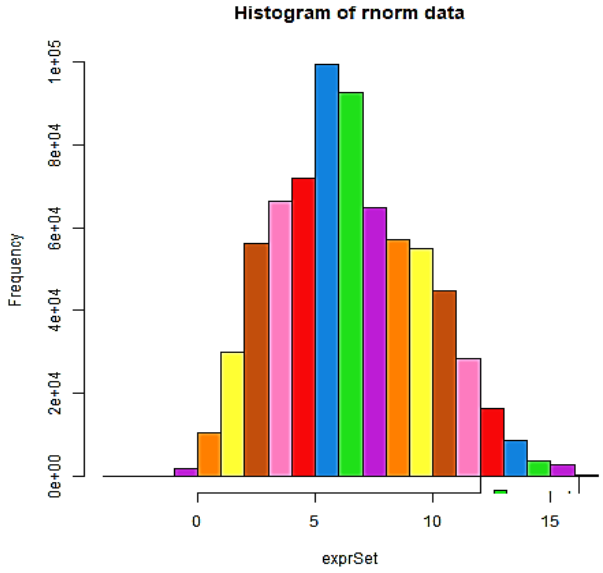
Results

4.1 Identification of Differentially Expressed Genes (DEGs)

Int this study we identified 47 DEGs from GSE9768 dataset of GERD and 17 DEGs were identified in the GSE161355 datasets of DM stringent p-value threshold of <0.002(Figure 4.3 &4.4). Boxplots and histogram of the raw and normalized expression data for both datasets were presented in Figure 4.1 and 4.2, respectively. Whereas, Table 4.1 and 4.2 provide the list of significant DEGs from both datasets.



(a) (b)



(c)

Figure 4.1 (a) Boxplot of raw data (b) Each boxplot shows the spread of normalized gene expression across samples, with a consistent range indicating effective normalization and minimized technical biases." (c) Histogram of normalized data for GERD datasets.

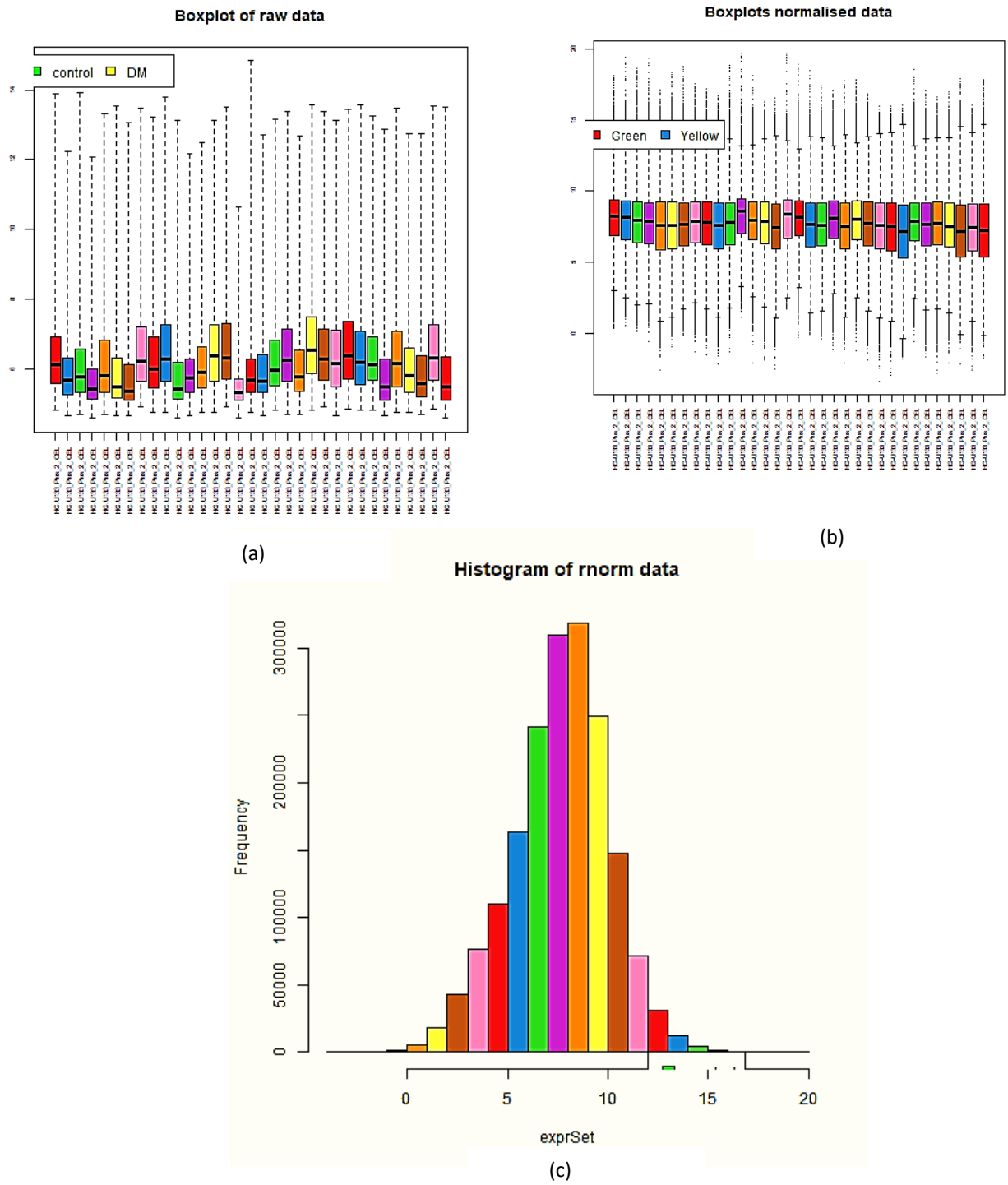


Figure 4.2 (a) Boxplot of raw data (b) Each boxplot shows the spread of normalized gene expression across samples, with a consistent range indicating effective normalization and minimized technical biases." (c) Histogram of normalized data for DM datasets.

Table 4.1 47 Differentially Expressed Genes of GSE9768 dataset

S.NO	ID	Symbol	Name	Ensembl
1	1552424_at	KLHL10	kelch like family member 10	ENSG00000161594
2	1553147_at	RANBP3L	RAN binding protein 3 like	ENSG00000164188
3	1554916_a_at	JRK	Jrk helix-turn-helix protein	ENSG00000234616
4	1557734_s_at	LOC100130548	uncharacterized LOC100130548	ENSG00000235138
5	1559021_at	C3orf52	chromosome 3 open reading frame 52	ENSG00000114529
6	1560794_at	TMEM252-DT	TMEM252 divergent transcript	ENSG00000226337
7	1562689_at	GCSIR	GPR55 cis regulatory suppressor of immune response RNA	ENSG00000232520
8	1563014_at	RPS15	ribosomal protein S15	ENSG00000115268
9	1563532_at	HMCN2	hemicentin 2	ENSG00000148357
10	1567622_at	ABLIM2	actin binding LIM protein family member 2	ENSG00000163995
11	203299_s_at	AP1S2	adaptor related protein complex 1 subunit sigma 2	ENSG00000182287
12	205244_s_at	SLC13A3	solute carrier family 13 member 3	ENSG00000158296
13	205365_at	HOXB6	homeobox B6	ENSG00000108511
14	206125_s_at	KLK8	kallikrein related peptidase 8	ENSG00000129455
15	206242_at	TM4SF5	transmembrane 4 L six family member 5	ENSG00000142484
16	206402_s_at	NPFF	neuropeptide FF-amide peptide precursor	ENSG00000139574
17	207177_at	PTGFR	prostaglandin F receptor	ENSG00000122420
18	208177_at	SLC34A1	solute carrier family 34 member 1	ENSG00000131183
19	209843_s_at	SOX10	SRY-box transcription factor 10	ENSG00000100146

S.NO	ID	Symbol	Name	Ensembl
20	210096_at	CYP4B1	cytochrome P450 family 4 subfamily B member 1	ENSG00000142973
21	211166_at	FAM153A	family with sequence similarity 153 member A	ENSG00000170074
22	211807_x_at	PCDHGB5	protocadherin gamma subfamily B, 5	ENSG00000276547
23	213674_x_at	IGHD	immunoglobulin heavy constant delta	ENSG00000211898
24	213692_s_at	VDR	vitamin D receptor	ENSG00000111424
25	214003_x_at	RPS20	ribosomal protein S20	ENSG00000008988
26	215130_s_at	IQCK	IQ motif containing K	ENSG00000174628
27	219761_at	CLEC1A	C-type lectin domain family 1 member A	ENSG00000150048
28	225323_at	CC2D1B	coiled-coil and C2 domain containing 1B	ENSG00000154222
29	226724_s_at	PSPC1	paraspeckle component 1	ENSG00000121390
30	229081_at	SLC25A13	solute carrier family 25 member 13	ENSG00000004864
31	229123_at	ZNF224	zinc finger protein 224	ENSG00000267680
32	229222_at	ACSS3	acyl-CoA synthetase short chain family member 3	ENSG00000111058
33	229273_at	SALL1	spalt like transcription factor 1	ENSG00000103449
34	229391_s_at	CALHM6	calcium homeostasis modulator family member 6	ENSG00000188820
35	230611_at	SYPL2	synaptophysin like 2	ENSG00000143028
36	231439_at	LRATD1	LRAT domain containing 1	ENSG00000162981
37	231653_at	ITPRID1	ITPR interacting domain containing 1	ENSG00000180347
38	234919_s_at	SNTG1	syntrophin gamma 1	ENSG00000147481

S.NO	ID	Symbol	Name	Ensembl
39	236672_at	ZNF681	zinc finger protein 681	ENSG00000196172
40	236885_at	MEX3A	mex-3 RNA binding family member A	ENSG00000254726
41	238113_at	SMIM2-AS1	SMIM2 antisense RNA 1	ENSG00000227258
42	238518_x_at	GLYCTK	glycerate kinase	ENSG00000168237
43	239183_at	ANGPTL1	angiopoietin like 1	ENSG00000116194
44	240869_at	KCNB1	potassium voltage-gated channel subfamily B member 1	ENSG00000158445
45	241552_at	AA06	uncharacterized LOC100506677	ENSG00000265544
46	241755_at	UQCRC2	ubiquinol-cytochrome c reductase core protein 2	ENSG00000140740
47	242253_at	PPP5D1P	PPP5 tetratricopeptide repeat domain containing 1, pseudogene	ENSG00000291145

Table 4.2 17 Differentially Expressed Genes of GSE161355 dataset

S.NO	ID	Symbol	Name	Ensembl
1	1552605_s_at	LINC00308	long intergenic non-protein coding RNA 308	c("ENSG00000184856", "ENSG00000262510")
2	1568784_at	LOC102725116	uncharacterized LOC102725116	
3	1569149_at	PDLIM7	PDZ and LIM domain 7	ENSG00000196923
4	202269_x_at	GBP1	guanylate binding protein 1	ENSG00000117228
5	202685_s_at	AXL	AXL receptor tyrosine kinase	ENSG00000167601
6	206631_at	PTGER2	prostaglandin E receptor 2	ENSG00000125384
7	206826_at	PMP2	peripheral myelin protein 2	ENSG00000147588
8	207720_at	LORICRIN	loricrin cornified envelope precursor protein	ENSG00000203782
9	208825_x_at	RPL23A	ribosomal protein L23a	ENSG00000198242
10	208949_s_at	LGALS3	galectin 3	ENSG00000131981
11	209395_at	CHI3L1	chitinase 3 like 1	ENSG00000133048
12	210091_s_at	DTNA	dystrobrevin alpha	ENSG00000134769
13	213526_s_at	LIN37	lin-37 DREAM MuvB core complex component	ENSG00000267796
14	213560_at	GADD45B	growth arrest and DNA damage inducible beta	ENSG00000099860
15	218149_s_at	ZNF395	zinc finger protein 395	ENSG00000186918
16	219052_at	HPS6	HPS6 biogenesis of lysosomal organelles complex 2 subunit 3	ENSG00000166189
17	224954_at	SHMT1	serine hydroxymethyltransferase 1	c("ENSG00000176974", "ENSG00000284320")

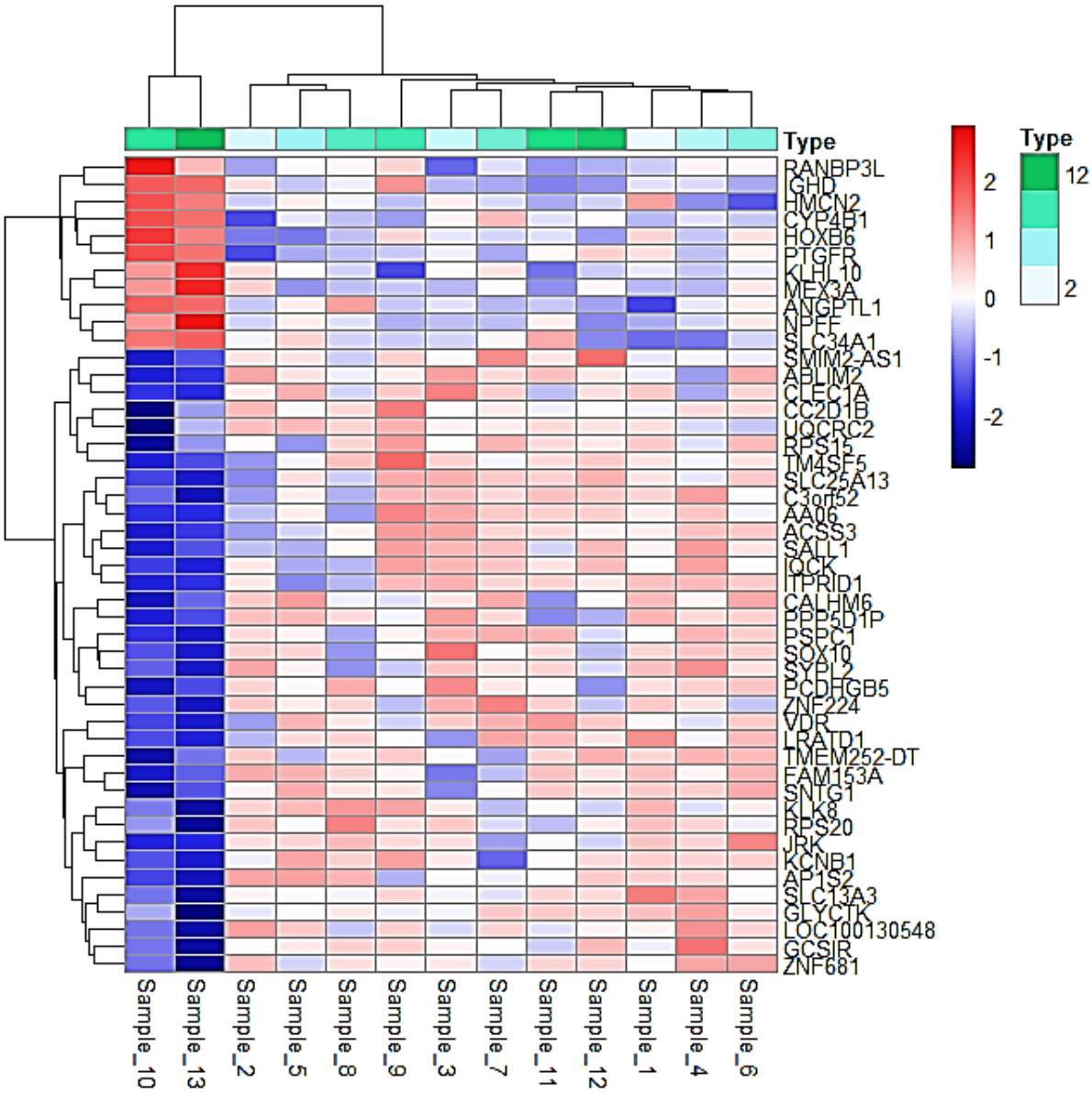


Figure 4.3: DEGs for GERD patients were selected using a stringent P-value cutoff of less than 0.002. Heat map visualizations illustrate gene expression levels, with blue representing lower expression and red indicating higher expression across samples.

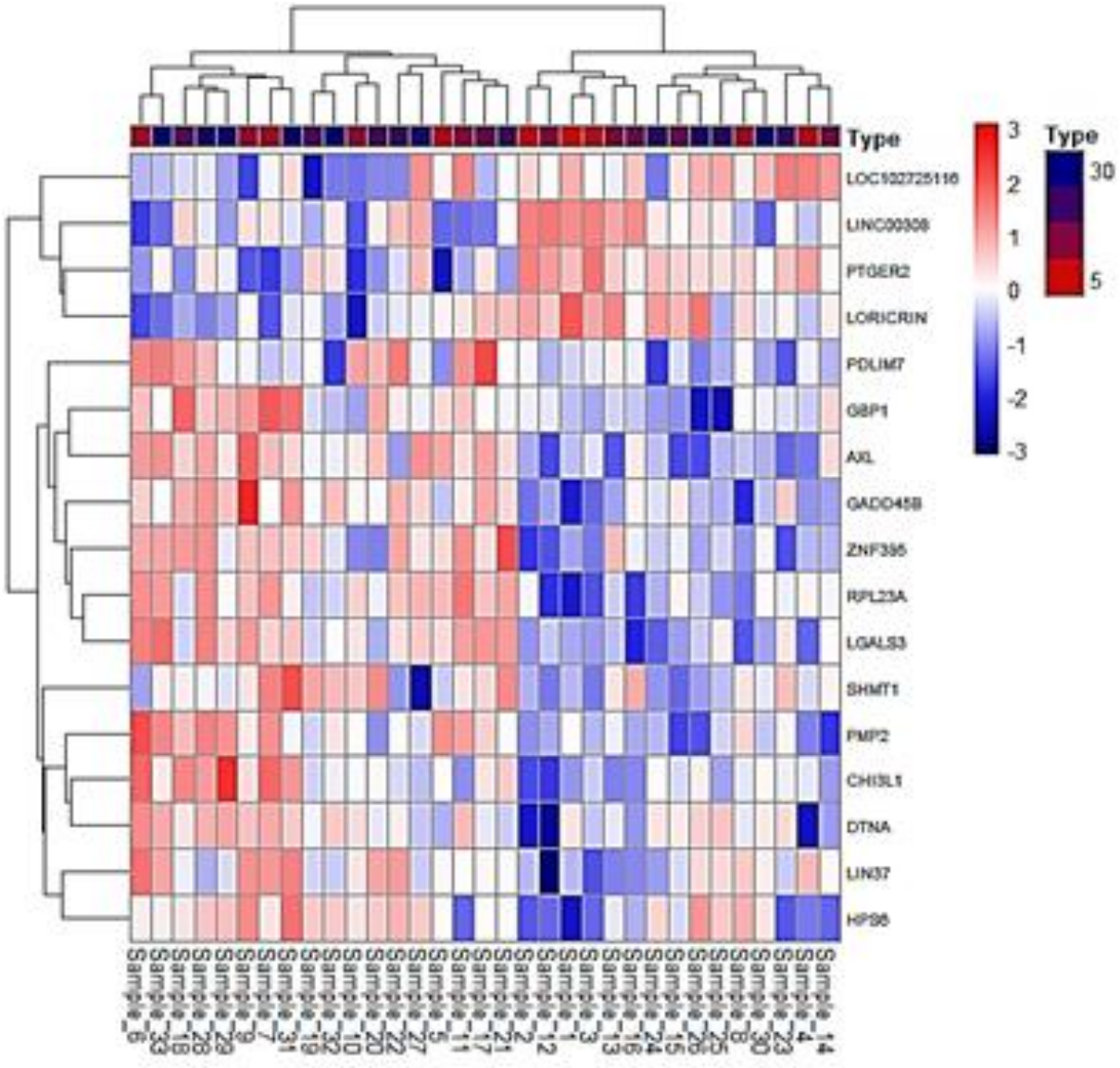


Figure 4.4: DEGs for DM patients were identified using a threshold P-value less than 0.002.Heat map visualizations illustrate gene expression levels, with blue representing lower expression and red indicating higher expression across samples.

4.2 Multivariate Covariance Analysis of Diabetes and GERD DEGs

The multivariate covariance analysis of diabetes-associated DEGs focusses on the top ten differentially expressed genes(shown in Table 4.3) chosen from a list of 17 relevant genes. A symmetric covariance matrix was used to investigate the interrelationships between these genes. The diagonal members in the matrix have values of 1.00, indicating the variance between each gene and itself. Off-diagonal elements display covariance values, which describe the extent to which gene pairs' expression levels change concurrently. Several gene pairings showed significant positive covariance, including AXL and LGALS3 (0.63), RPL23A and LGALS3 (0.77), AXL and GBP1 (0.61), and PMP2 and LGALS3 (0.65), indicating a high degree of coordinated gene expression.

In contrast, negative covariance, such as those between RPL23A and LINC00308 (-0.60) and PTGER2 and GBP1 (-0.56), indicate inverse expression relationships. For example, the positive covariance between AXL and LGALS3 indicates that both genes may have comparable expression trends, whereas the negative covariance between PTGER2 and GBP1 shows the opposite pattern of regulation. These findings contribute to a better understanding of the complex regulatory behavior of diabetes-related genes, as well as additional investigation into their biological functions and connections.

The interrelationships between differentially expressed genes (DEGs) related to GERD are explored, with a total of 47 being identified as substantial. To be clear and illustrative, the top 10 DEGs are shown in this section(Table 4.4) through a symmetric multivariate covariance matrix. The diagonal entry of the matrix contains all 1.00, which is the self-variance of all the genes and the off-diagonal elements are the covariance between pairs of genes. Good positive correlations like KLHL10 and TMEM252-DT (0.71), RANBP3L and TMEM252-DT (0.77), JRK and LOC100130548 (0.70), LOC100130548 and ABLIM2 (0.86), and GCSIR and RPS15 (0.76) indicate potential co-regulation or biological commonality. Conversely, notable negative covariance such as KLHL10 and C3orf52 (-0.84), KLHL10 and ABLIM2 (-0.76), and TMEM252-DT and HMCN2 (-0.71) indicate inverse expression patterns. As an example, the positive covariance between the genes LOC100130548 and ABLIM2 is strongly positive, which suggests that these genes are coordinated in their expression patterns and the negative covariance between

the genes KLHL10 and C3orf52 is high and means that these genes are antagonistically regulated. The results provide significant insight into the intricate molecular interactions of the GERD.

Table 4.3 Covariance Matrix of Significant Differential Expressed Genes of GERD

	KLHL10	RANBP3L	JRK	LOC100130548	C3orf52	TMEM252-DT	GCSIR	RPS15	HMCN2	ABLIM2
KLHL10	1	0.434194	-0.69983	-0.65993	-0.84428	0.706809	-0.69554	-0.61157	-0.70253	-0.760457857
RANBP3L	0.434194	1	-0.55692	-0.43337	-0.55117	0.772632	-0.55064	-0.79998	-0.70108	-0.403932607
JRK	-0.69983	-0.55692	1	0.69891	0.591186	-0.50178	0.660032	0.778237	0.79661	0.728044548
LOC100130548	-0.65993	-0.43337	0.69891	1	0.723468	-0.50693	0.700427	0.579589	0.627894	0.859636628
C3orf52	-0.84428	-0.55117	0.591186	0.723468	1	-0.76479	0.750678	0.576364	0.612298	0.785634092
TMEM252-DT	0.706809	0.772632	-0.50178	-0.50693	-0.76479	1	-0.65463	-0.68955	-0.70639	-0.607072402
GCSIR	-0.69554	-0.55064	0.660032	0.700427	0.750678	-0.65463	1	0.760232	0.505099	0.682574138
RPS15	-0.61157	-0.79998	0.778237	0.579589	0.576364	-0.68955	0.760232	1	0.6401	0.454855165
HMCN2	-0.70253	-0.70108	0.79661	0.627894	0.612298	-0.70639	0.505099	0.6401	1	0.690904614
ABLIM2	-0.76046	-0.40393	0.728045	0.859637	0.785634	-0.60707	0.682574	0.454855	0.690905	1

Table 4.4 Covariance Matrix of Significant Differentially Expressed Genes of Diabetes Mellitus

	LINC00308	LOC102725116	PDLIM7	GBP1	AXL	PTGER2	PMP2	LORICRIN	RPL23A	LGALS3
LINC00308	1	0.156128	-0.32457	-0.1885	-0.36942	0.491407	-0.49508	0.506048	-0.60118	-0.4656
LOC102725116	0.156127838	1	-0.26922	-0.21591	-0.30711	0.29017	-0.26425	0.280883	-0.11392	-0.1844
PDLIM7	-0.324566193	-0.26922	1	0.349592	0.428472	-0.26037	0.406748	-0.32311	0.43922	0.435183
GBP1	-0.188502395	-0.21591	0.349592	1	0.606136	-0.56288	0.515168	-0.38658	0.45885	0.387928
AXL	-0.369422399	-0.30711	0.428472	0.606136	1	-0.58885	0.577701	-0.50356	0.565216	0.62695
PTGER2	0.491406729	0.29017	-0.26037	-0.56288	-0.58885	1	-0.47676	0.518916	-0.56199	-0.47434
PMP2	-0.495078892	-0.26425	0.406748	0.515168	0.577701	-0.47676	1	-0.50557	0.509454	0.652259
LORICRIN	0.506048201	0.280883	-0.32311	-0.38658	-0.50356	0.518916	-0.50557	1	-0.50623	-0.44187
RPL23A	-0.601182288	-0.11392	0.43922	0.45885	0.565216	-0.56199	0.509454	-0.50623	1	0.771032
LGALS3	-0.465601951	-0.1844	0.435183	0.387928	0.62695	-0.47434	0.652259	-0.44187	0.771032	1

4.3 Association Rule Mining on Differentially Expressed Genes

In this study, Association Rule Mining was used to explore associations between two diseases, GERD and DM, through their Differentially Expressed Genes (DEGs). To achieve this, the Apriori algorithm was applied alongside Jaccard similarity to identify gene expression patterns with similar profiles. This approach was used to generate frequent item sets with a minimum support of 0.2 and a confidence threshold of 0.8. Once the frequent item sets were identified, association rules were generated based on these patterns. 1282 rules were initially generated. These rules were then filtered by applying a minimum confidence threshold of greater than or equal to 0.8, ensuring that only the most reliable associations were retained. After filtering, 88 high-confidence rules were selected. To remove redundancy, duplicate rules were eliminated, leaving a final set of 15 unique rules. Table 4.5 shows that dataset was divided into three sets of association rules based on the number of genes involved: One-Gene Rules, Two-Gene Rules, and Three-Gene Rules. Each set of rules was assessed using key metrics such as support, confidence, and lift, which indicate the frequency, reliability, and significance of the associations between genes.

4.3.1 Rules interpretation

The One-Gene Rules identify associations where a single gene in the antecedent is linked to a combination of genes in the consequent. A notable finding was the strong association between CHI3L1=High and both AXL=High and LGALS3=High, which had a support of 0.29, confidence of 1, and lift of 4.5. This suggests that whenever CHI3L1 is highly expressed, AXL and LGALS3 are always co-expressed, and this association occurs 4.5 times more frequently than would be expected by random chance. Similarly, LGALS3=High was strongly associated with CHI3L1=High and DTNA=High (lift = 5.4), indicating that LGALS3 plays a central role in regulating CHI3L1 and DTNA expression in this dataset.

The Two-Gene Rules involve associations between two genes in the antecedent and other genes in the consequent. These rules reveal potential interactions between gene pairs that influence gene expression. For example, the rule {AXL=High, UQCRC2=Low Medium} => {GBP1=High, PTGER2=Low Medium} (support = 0.2, confidence = 1, lift = 4.3) suggests that the co-expression of AXL and UQCRC2 leads to the always observed co-expression of GBP1 and PTGER2 at specific expression levels. This association occurs 4.3 times more often than expected by random

Chance, highlighting a potentially important regulatory interaction between these genes. Additionally, AXL=High, MEX3A=High was associated with GBP1=High, PTGER2=Low Medium (support = 0.25 confidence = 1, lift = 4.3), suggesting that these gene combinations could jointly influence the expression of GBP1 and PTGER2. The strong confidence values for all rules in this set further support the reliability of these associations.

The Three-Gene Rules represent complex relationships, involving interactions between three genes in the antecedent and a set of genes in the consequent. One notable rule is {GBP1=High, AXL=High, IGHD=High} => {PTGER2=Low Medium}, with a support of 0.29, confidence of 1, and lift of 4.15. This rule indicates that the combination of GBP1, AXL, and IGHD expression levels leads to the consistent co-expression of PTGER2 at a specific expression level. The lift value of 4.15 suggests that this association is 4.15 times more likely to occur than by random chance. Another significant rule involves {GBP1=High, AXL=High, NPFF=High} => {PTGER2=Low Medium} (support = 0.21, confidence = 1, lift = 4.14). This reinforces the idea that GBP1, AXL, and NPFF are strong regulators of PTGER2 expression, with the relationship occurring with a high degree of confidence and significantly more often than expected by chance.

The rule {GCSIR=High, RPL23A=High, CHI3L1=High} => {LGALS3=High} (support = 0.21, confidence = 1, lift = 5.49) highlights another important relationship, where the combination of GCSIR, RPL23A, and CHI3L1 expression consistently leads to LGALS3 expression, again showing a high lift and suggesting a significant biological interaction.

4.3.2 Key Gene Associations

Across all rule types, AXL and GBP1 emerged as central genes, frequently appearing in both the antecedents and consequents of the association rules. This suggests that both AXL and GBP1 play critical roles in the regulation of other genes, such as PTGER2, LGALS3, and CHI3L1. Their consistent presence in strong association rules suggests potential functional interactions or pathways involving these genes. Moreover, the high confidence values across the rules indicate that these associations are robust and reliable in the context of the dataset.

4.3.3 Biological Interpretation and Further Analysis

These rules have undergone biological interpretation, where we explored each gene involved in the associations. All the genes that appeared in the rules were further examined using pathway analysis, protein-protein interaction, and disease gene association studies. By investigating these frequent genes in various biological contexts, we aimed to identify potential pathways, interactions, and mechanisms relevant to the diseases studied.

.

Table 4.5 Selected Association Rules mined from DEGS of GERD and DM

Rules	support	confidence	lift
{CHI3L1=High}=>{AXL=High,LGALS3=High}}	0.29516	1	4.5248
{LGALS3=High}=>{CHI3L1=High,DTNA=High}	0.28103	1	5.497854
{DTNA=High}=>{GBP1=High,LGALS3=High}	0.288837	1	3.846154
{RPL23A=High,CHI3L1=High} => {LGALS3=High}	0.238642	1	5.497854
{AXL=High}} => {GBP1=High,PTGER2=Low-Medium}	0.229508	1	4.357143
{AXL=High,UQCRC2=Low-Medium} => {GBP1=High,PTGER2=Low-Medium}	0.201483	1	4.357143
{AXL=High,MEX3A=High} => {GBP1=High,PTGER2=Low-Medium}	0.257611	1	4.357143
{HOXB6=High,UQCRC2=Low-Medium} => {ANGPTL1=High}	0.230769	1	3.702312
{GBP1=High, AXL=High, IGHD=High}=>{PTGER2=Low-Medium}	0.292896	1	4.145631
{GBP1=High, AXL=High, NPFF=High}=>{PTGER2=Low-Medium}	0.214832	1	4.145631
{AXL=High,PTGER2=Low-Medium,ZNF224=Low-Medium}=> {{GBP1=High}}	0.249024	1	3.702312
{GBP1=High, AXL=High,ITPRID1=Low-Medium} => {PTGER2=Low-Medium}	0.249024	1	4.145631
{GCSIR=High,RPL23A=High,CHI3L1=High} => {LGALS3=High}	0.217096	1	5.497854
{GBP1=High, AXL=High,PTGFR=High} => {PTGER2=Low-Medium}	0.219516	1	4.145631
{GBP1=High, AXL=High,RANBP3L=High} => {PTGER2=Low-Medium}	0.232709	1	4.145631
{ABLIM2=Low-Medium,GBP1=High, AXL=High}=> {PTGER2=Low-Medium}	0.292896	1	4.145631
{ANGPTL1=High,GBP1=High, AXL=High} => {PTGER2=Low-Medium}	0.28103	1	4.145631

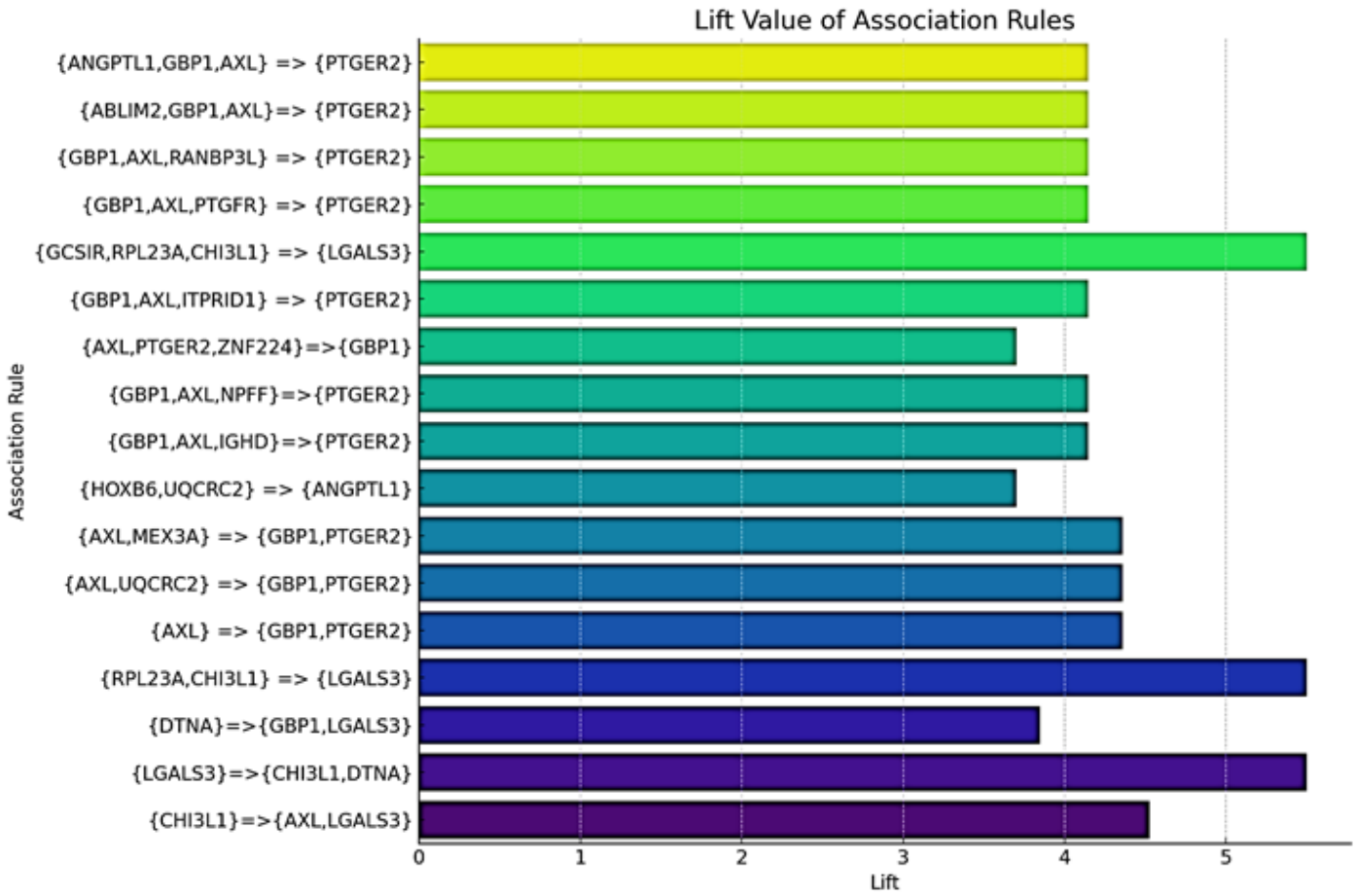


Figure 4.5 Bar graph showing lift values of key gene association rules linked to GERD and DM. Higher lift indicates stronger non-random associations among genes like LGALS3, CHI3L1 and AXL

4.4 Functional Annotation and Pathway Enrichment Analysis

In this study, EnrichR was used to perform pathway and gene ontology enrichment analyses to determine biological pathways and functional categories that are significantly associated with the common genes between GERD and DM. Each of the 20 genes found in this study was analyzed using EnrichR. Pathway enrichment with KEGG, Reactome and WikiPathways (Figure 4.7) identified the important molecular pathways that could be related to both conditions, especially focusing on the inflammatory and metabolic signaling cascades. Remarkably, amongst the key-shared pathways discovered was the prostaglandin-mediated inflammatory cascade, namely, prostaglandin E2 (PGE2) and the enzyme that catalyzes its production, cyclooxygenase-2 (COX-2). Literature supports the evidence that acid-mediated epithelial damage contributes to the upregulation of COX-2 expression and the production of PGE2 in GERD, leading to inflammation in the mucosa and esophagus dysfunction. Hyperglycemia in diabetes mellitus also enhances COX-2 activity, and this provokes prostaglandin-driven alterations in gastric motility, such as dysrhythmia and slowed gastric emptying, characteristic of diabetic gastroparesis. These effects have been partially mitigated by the use of COX inhibitors, including indomethacin, which highlights the importance of this pathway in both diseases. These results strengthen the molecular relationship between GERD and DM via a common COX-2/prostaglandin signaling pathway.

GO enrichment analysis supported this link by highlighting terms such as cellular response to prostaglandin stimulus (GO:0071379), prostaglandin receptor activity (GO:0004955), and cytoskeletal structures (GO:0005856), indicating prostaglandin's involvement in inflammation, signalling, and cellular dynamics. Figure 4.6 depicts these significant GO keywords, which demonstrate the functional convergence of GERD and DM at both the molecular and cellular levels.

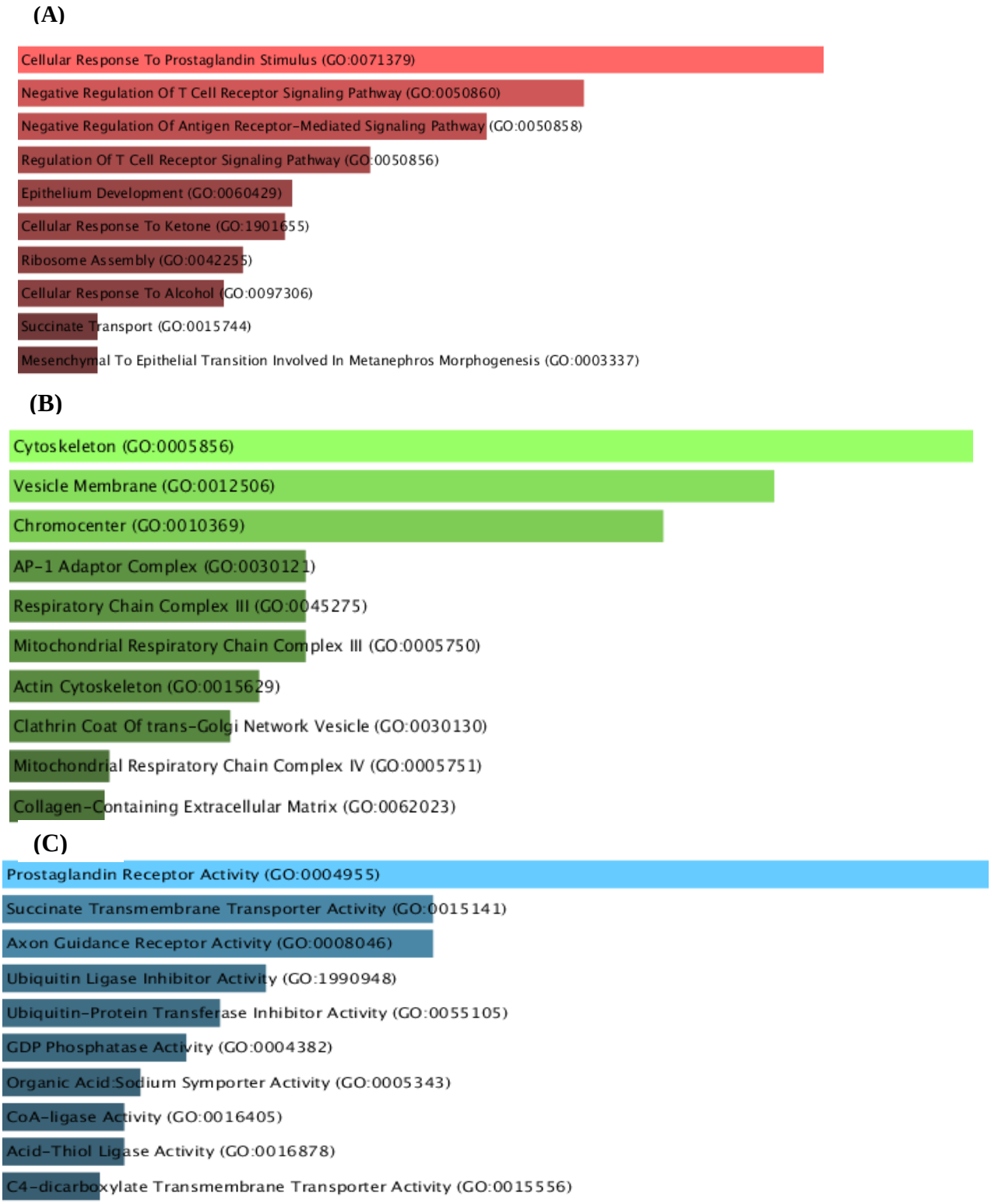
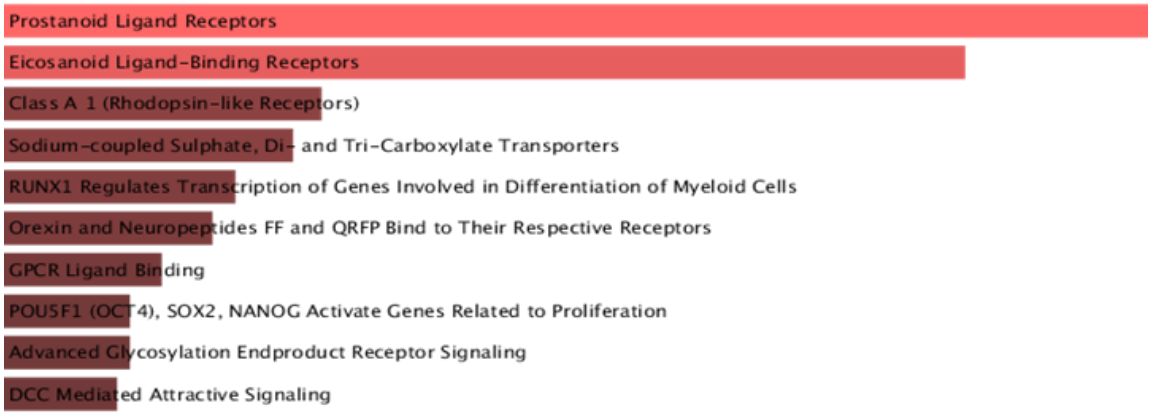
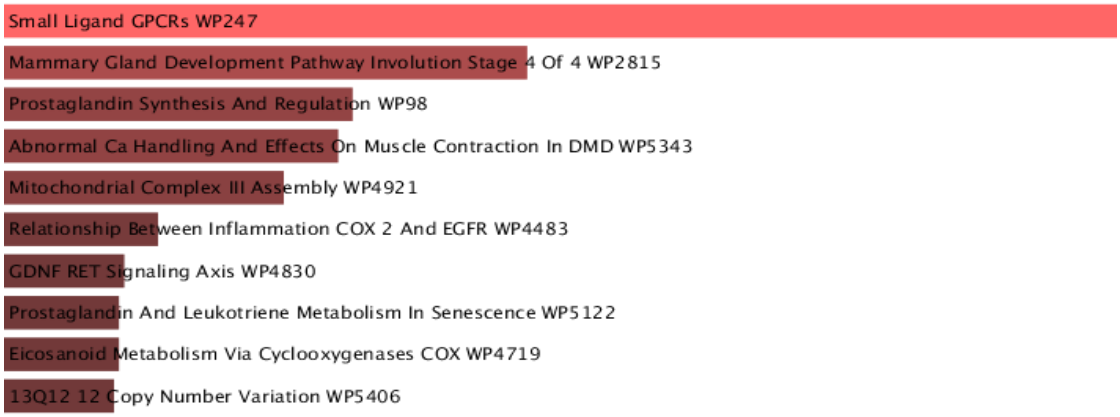


Figure 4.6 bar graph shows the ontological analysis of common genes between DM and GERD (A) biological process, (B) cellular component, and (C) molecular function.

(A)



(B)



(C)

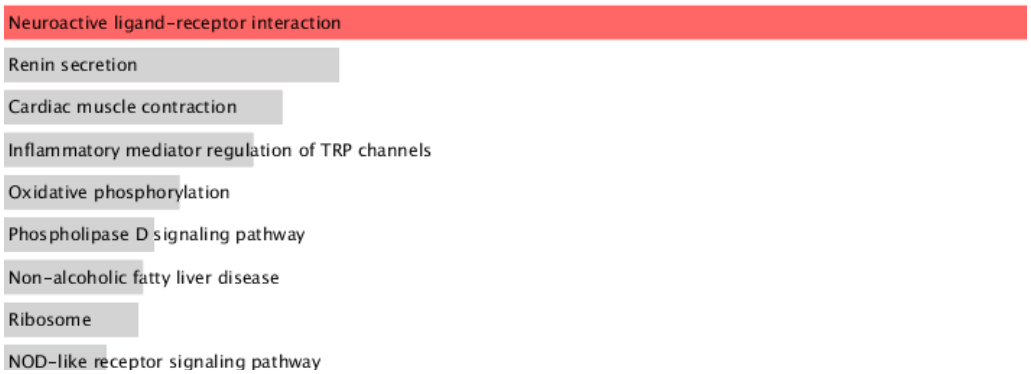


Figure 4.7 bar graphs shows the pathway analysis of common genes between Diabetes Mellitus, and Gastroesophageal Reflux Disease: (**A**) KEGG pathways, (**B**) Reactome (**C**) WikiPathways 2024 Humans.

4.5 PPI Network Analysis and Gene Co-expression

To study the interactions and common pathways of the identified common genes, we examined the protein interaction network using STRING database. The produced network comprised 40 nodes and 95 links, and the enrichment p-value was highly significant (less than 1.0e-16) as shown in Figure 4.9. In order to gain more insight into the functional role of these genes, we built a complete network of interactions of these genes using GeneMANIA database (Figure 4.10). With this network, 44 percent of the interactions were physical, 28.07 percent co-expression, 22.13 percent predicted associations, and 3.38 percent co-occurrence. The gene co expression network showed that the genes were mostly associated to prostanoid receptor activity, amino sugar catabolic processes, cellular response to fatty acids, and ERK1/ERK2 signaling cascade. The results indicate that the characterized genes can be biological markers and present viable targets in establishing therapeutic approaches to GERD and DM.

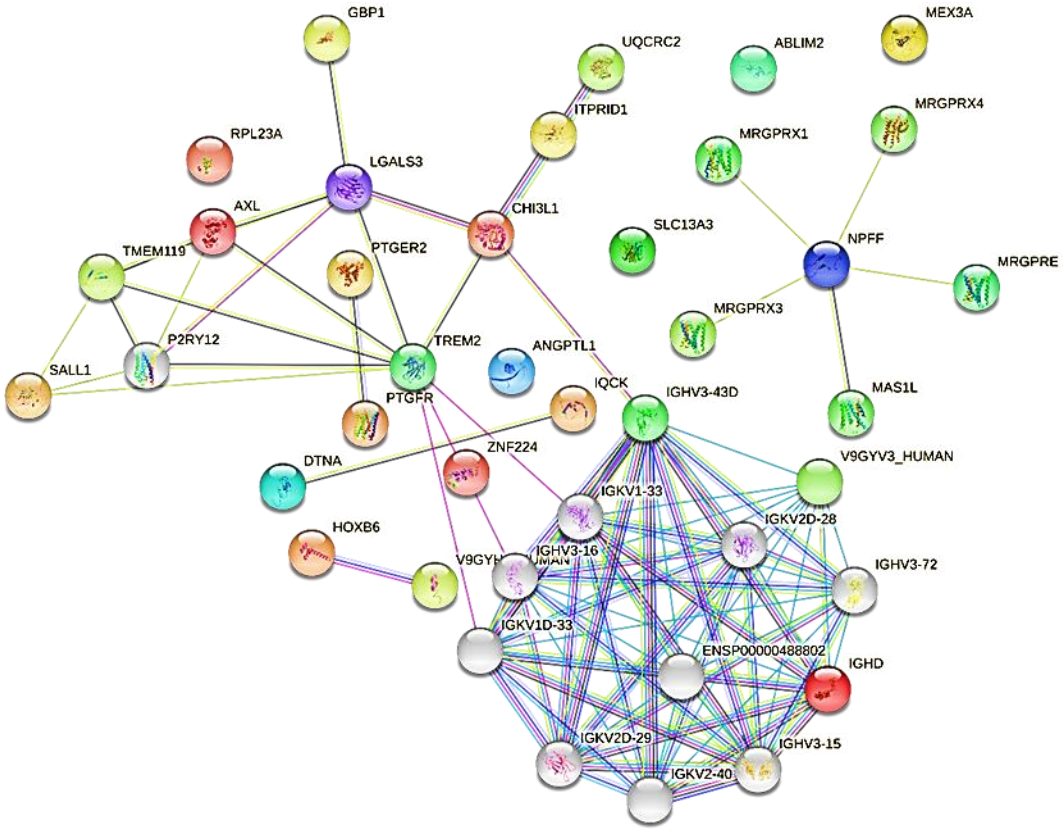


Figure 4.8 PPI Network of Associated Genes between DM and GERD.

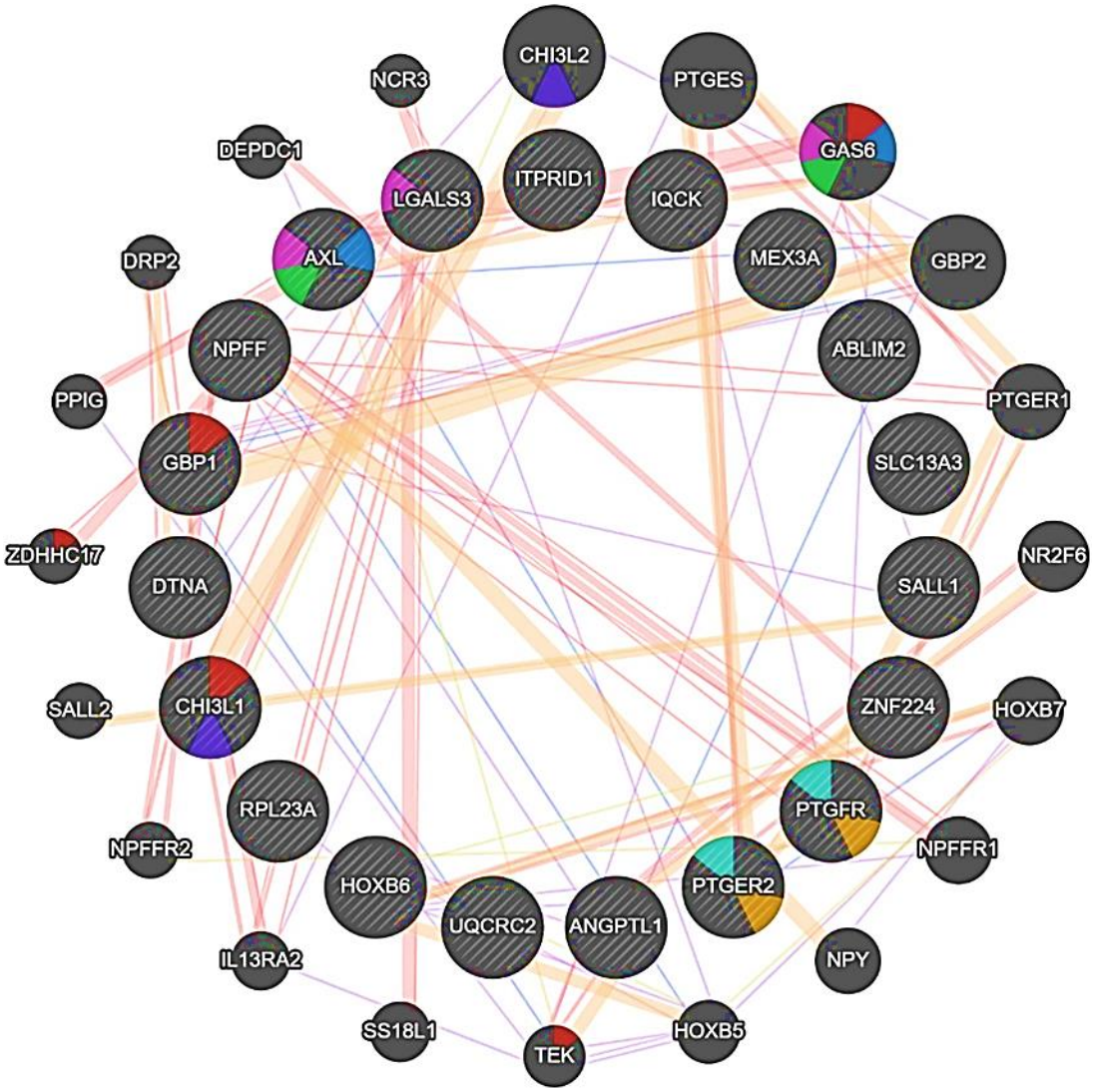


Figure 4.9. GeneMANIA-generated gene co-expression network showing the common genes and their functionally associated co-expressed partners linked to both GERD and DM.

4.6 Gene Disease Association

In this study, the genes identified through association rule mining were further evaluated for their biological relevance to DM and GERD using web-based resources, including the CTD, DisGeNet, and PubMed. We individually assessed each gene involved in the association rules by selecting "Diabetes Mellitus" and "GERD" from the disease category tab in these databases. CTD provided association data for all the identified genes (AXL, UQCRC2, GBP1, ABLIM2, NPFF, PTGER2, MEX3A, DTNA, LGALS3, CHI3L1, RANBP3L, ITPRID1, PTGFR, SALL1, SLC13A3, GCSIR, RPL23A, ZNF224, IGHD, and IQCK) based on inference and reference scores. However, DisGeNet reported associations for only six genes, while PubMed retrieved relevant literature for seven genes with both diseases. To refine the results, we selected genes commonly identified in at least two of the three databases. Notably, PubMed and DisGeNet both included most of the genes, except that NPFF and UQCRC2 appeared only in PubMed, while SALL1 appeared only in DisGeNet. Among them, NPFF showed an association with DM only, while UQCRC2 and SALL1 showed an association with GERD. The genes AXL, GBP1, PTGER2, LGALS3, and CHI3L1 emerged as common candidates, showing strong associations with both DM and GERD across all three databases. These findings highlight a set of key genes potentially involved in both diseases, providing a foundation for further functional and clinical investigations.

4.7 Discussion

DM and GERD are common chronic conditions that often occur together, with studies showing a strong connection between them. Diabetes can lead to complications affecting the autonomic nervous system, resulting in nerve dysfunction, slow stomach emptying, and issues with esophageal movement, can contribute to acid reflux and GERD symptoms. Additionally, factors like obesity and lifestyle habits may further increase the risk. Understanding the connection between these two conditions can help develop better treatment strategies and improve patient care. This study investigate the potential relationship between GERD and DM through association rule mining and identify common genes that are involved in both conditions. This study introduce innovative approach by applying association rule mining in R to examine the genomic relationship between GERD and DM. Recent studies increasingly recognized the association between GERD and DM, showing that individuals with DM face a higher risk of GERD symptoms and

complications. Poor glycemic control and autonomic neuropathy significantly contribute to these conditions. Meta-analysis studies reported that DM patients exhibit a 61% higher risk of GERD, with delayed gastric emptying proposed as a possible link. Furthermore, GERD may worsen metabolic dysregulation in DM patients, suggesting a bidirectional relationship (19). Recent advancements in genomic research have highlighted shared molecular mechanisms underlying GERD and DM. Mendelian randomization and genome-wide association studies (GWAS) have revealed overlapping genetic pathways, including those involved in inflammation, cytokine signaling, and metabolic regulation (37).

The analysis identified 47 DEGs from GERD dataset and 17 DEGs from DM datasets by applying p-value threshold to 0.002. We applied association rule mining to identify gene expression patterns linked to GERD and DM using the apriori algorithm and Jaccard similarity to assess the similar expression genes, with a minimum support of 0.2 and confidence value of 0.8. From an initial 1282 rules, filtering based on confidence resulted in 88 unique rules after removing redundancies. The extracted rules revealed strong co-expression patterns, highlighting genes like AXL, PTGER2, LGALS3, GBP1 and CHI3L1 as key regulators in immune response, inflammation, and cellular structure regulation. Notably, these five genes frequently appeared in the highest number of rules, emphasizing their central role in the association. All 20 genes (AXL, UQCRC2, GBP1, ABLIM2, NPFF, PTGER2, MEX3A, DTNA, LGALS3, CHI3L1, RANBP3L, ITPRID1, PTGFR, SALL1, SLC13A3, GCSIR, RPL23A, ZNF224, IGHD, and IQCK) undergo further evaluation to check their role in various pathways and disease databases that linked with both conditions. Five of them (AXL, PTGER2, LGALS3, GBP1 and CHI3L1) have been involved in the pathogenic mechanism of both diseases. GERD and DM share molecular similarities, with several key genes implicated in both diseases. Through its control of metabolic and inflammatory signalling pathways, Studies have recently stated that AXL has a critical function in the pathogenesis of both DM and GERD, hence could specify the molecular link. Endothelial dysfunction, impaired angiogenesis, and increased inflammation play a role in vascular complications in DM due to disruption of Gas6/AXL/Akt signaling pathway induced by hyperglycemia. In GERD, AXL contributes to tumor progression, chemo resistance and autophagy through the ROS–AMPK–ULK1 axis in the context of esophageal adenocarcinoma (EAC),

pointing to common mechanisms inflammatory and metabolic (38, 39). GBP1 contributes to disease progression in both conditions. In esophageal squamous cell carcinoma (ESCC) associated with GERD, GBP1 enhances the lymphatic metastasis and tumor invasiveness. It is involved in DM as a player in diabetic retinopathy through the processes of pyroptosis and vascular damage, tying inflammation to metabolic dysfunction (40, 41). Receptor PTGER2 for prostaglandin E2 (PGE2) has dual functions. In DM, its activation prevents β-cell loss and preserves function and insulin secretion, while its inhibition aggravates β-cell loss. In contrast, in GERD, PTGER2 enhances esophageal epithelial inflammation and remodeling by inducing cytokines (e.g., IL-6, CXCL-8) and increases mucosal barrier dysfunction (42, 43). LGALS3 (Galectin-3) is elevated in both diseases. In DM, it is involved in the endothelial dysfunction, insulin resistance and atherosclerosis. Elevated serum levels correlate with mucosal inflammation in GERD and therefore have a potential use as a diagnostic test and may play a role in immune modulation (44, 45). CHI3L1 (Chitinase-3-like protein 1) is associated with chronic inflammation and tissue remodeling in both conditions, In DM, it promotes nephropathy and retinopathy and in GERD associated ESCC, it promotes tumor progression by recruiting macrophages and immune suppression (46, 47). However, gene–disease association analysis from CTD, PubMed, and DisGeNet databases show that UQCRC2 and SALL1 are only associated with GERD, and NPFF with DM. The remaining 14 genes have no or little known association with any of the two diseases in the databases referenced.

Pathway enrichment analysis identified prostaglandin receptor signalling and COX inflammation as a shared pathogenic and inflammatory pathway between GERD and DM. In diabetes mellitus, the upregulation of COX-2 and prostaglandin receptors causes disruption of stomach motility and subsequent gastroparesis, the inhibition of COX-2 restores pacemaker activity. Several studies report that reflux exposure activates the COX-2 and PGES isoforms leading to increased PGE2 synthesis, which triggers mucosal inflammation and epithelial destruction in GERD. Additionally, PGES isoforms shows differential expression in GERD pathophysiology across normal, squamous, and adenocarcinoma cells. These findings establish Prostaglandin signalling as a shared inflammatory and functional mechanism in both disorders (48, 49).

Additionally, we have done network analysis using STRING, revealing 40 nodes and 95 edges with a highly significant PPI enrichment value (<1.0e-16), indicating strong interactions between these genes. GeneMANIA analysis further confirmed functional connections among these genes through co-expression (28.07%), predicted interactions (22.13%), physical interactions (44%), and co-localization (3.38%). These findings suggest that shared molecular mechanisms, including prostanoid receptor activity, amino sugar catabolic processes, and the ERK1/ERK2 cascade, may contribute to both GERD and DM pathogenesis. The identification of these common genetic markers not only strengthens the evidence of a molecular link between GERD and DM but also suggests that these genes could serve as diagnostic candidate for targeted therapy. Future studies should focus on validating these findings through experimental approaches such as gene knockdown studies, functional assays, and animal models to dissect the molecular interactions between GERD and DM and explore potential therapeutic strategies for managing these interconnected conditions.

Chapter 5

References

1. P. Gharahkhani et al., "Multitrait genetic association analysis identifies 50 new risk loci for gastro-esophageal reflux disease," Gut, vol. 71, no. 6, pp. 1053–1060, 2022. doi.org/10.1136/gutjnl-2020-323906.
2. D. A. Katzka and P. J. Kahrilas, "Advances in the diagnosis and management of gastroesophageal reflux disease," BMJ, vol. 371, 2020. doi.org/10.1136/bmj.m3786
3. J. Dent, H. B. El-Serag, M. A. Wallander and S. Johansson, "Epidemiology of gastro-oesophageal reflux disease: a systematic review," Gut, vol. 54, no. 5, pp. 710–717, May 2005, doi: 10.1136/gut.2004.051821.
4. T. Okimoto et al., "Increasing prevalence of gastroesophageal reflux disease among adults in Japan: A cross-sectional study," J. Gastroenterol., vol. 58, no. 4, pp. 234–242, 2023.
5. J. Ahn et al., "Gastro-esophageal reflux disease in primary care practice: a narrative review," Ann. Esophagus, vol. 6, no. 1, Article 5, 2023, doi: 10.21037/aoe-21-62.
6. M. Z. Banday, A. S. Sameer, and S. Nissar, "Pathophysiology of diabetes: An overview," 2020 doi: 10.4103/ajm.ajm_53_20.
7. S. Bolen, L. Feldman, J. Vassy et al., "Systematic review: comparative effectiveness and safety of oral medications for type 2 diabetes mellitus," Ann. Intern. Med., vol. 147, no. 6, pp. 386–399, 2007 ,doi.org/10.7326/0003-4819-147-6-200709180-00178.
8. S. E. Inzucchi, R. M. Bergenstal, J. B. Buse et al., "Management of Hyperglycemia in Type 2 Diabetes: A Patient-Centered Approach," Diabetes Care, vol. 38, no. 1, pp. 140–149, 2015, doi.org/10.2337/dc14-2441.
9. A. Ramachandran, "Know the signs and symptoms of diabetes," Indian J. Med. Res., vol. 140, no. 5, pp. 579–581, Nov. 2014.
10. P. T. Kröner, P. Cortés and F. J. Lukens, "The medical management of gastroesophageal reflux disease: A narrative review," J. Prim. Care Community Health, vol. 12, pp. 1–7, 2021, doi: 10.1177/21501327211046736.
11. A. Sonnenberg and H. B. El-Serag, "Clinical epidemiology and natural history of gastroesophageal reflux disease," Yale J. Biol. Med., vol. 72, no. 2–3, pp. 81–92, 1999.
12. A. M. Ahmed, "History of diabetes mellitus," Saudi Med. J., vol. 23, no. 4, pp. 373–378, Apr. 2002.

13. J. S. Nirwan, S. S. Hasan, Z. U. D. Babar et al., "Global prevalence and risk factors of gastro-oesophageal reflux disease (GORD): Systematic review with meta-analysis," Sci. Rep., vol. 10, no. 5814, 2020, doi: 10.1038/s41598-020-62795-1.
14. P. Zimmet, K. G. M. M. Alberti, and J. Shaw, "Global and societal implications of the diabetes epidemic," Nature, vol. 414, no. 6865, pp. 782–787, 2001.
15. K. L. Ong et al., "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," Lancet, vol. 402, no. 10397, pp. 203–234, 2023, doi: 10.1016/S0140-6736(23)01301-6.
16. F. Huerta-Iga, M. V. Bielsa-Fernández, J. M. Remes-Troche, M. A. Valdovinos-Díaz and J. L. Tamayo-de la Cuesta, "Diagnosis and treatment of gastroesophageal reflux disease: Recommendations of the Asociación Mexicana de Gastroenterología," Rev. Gastroenterol. Méx. (Engl. Ed.), vol. 81, no. 4, pp. 208–222, 2016, doi: 10.1016/j.rgmxen.2016.09.002.
17. American Diabetes Association, "Diagnosis and classification of diabetes mellitus," Diabetes Care, vol. 33, Suppl. 1, pp. S62–S69, Jan. 2010, doi: 10.2337/dc10-S062.
18. H. Sun, L. Yi, P. Wu, Y. Li, B. Luo and S. Xu, "Prevalence of Gastroesophageal Reflux Disease in Type II Diabetes Mellitus," World J. Gastroenterol., vol. 20, no. 40, pp. 15000–15006, 2014, doi: 10.3748/wjg.v20.i40.15000.
19. X.-M. Sun et al., "Association between diabetes mellitus and gastroesophageal reflux disease: a meta-analysis," World J. Gastroenterol., vol. 21, no. 10, pp. 3085–3096, 2015.
20. M. Anandhavalli, M. K. Ghose and K. Gauthaman, "Association rule mining in genomics," Int. J. Comput. Theory Eng., vol. 2, no. 2, pp. 269–273, 2010, doi: 10.7763/IJCTE.2010.V2.162.
21. J. E. Lockley, "A Comparative Study of Cluster Analysis and MANCOVA in the Analysis of Mathematics Achievement Data," in Contributions to Probability and Statistics, L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson, Eds., New York, NY, USA: Springer, 1989. doi: 10.1007/978-1-4612-3678-8_17.
22. G. Natalini, A. Singh and J. Carter, "Diabetes mellitus as an independent risk factor for GERD in African-Americans," Dis. Esophagus, vol. 27, no. 6, pp. 541–546, 2014.

23. C. H. Chang, C. H. Lin and T. H. Tsai, "Risk factors for developing diabetes mellitus in patients with gastroesophageal reflux disease: A population-based study," J. Diabetes Metab. Disord., vol. 20, no. 1, pp. 120–128, 2021.
24. L. Lin, J. Zhang and H. Xu, "Silent GERD in type 2 diabetes patients: A cross-sectional endoscopic study," BMC Gastroenterol., vol. 17, no. 1, pp. 42–49, 2017.
25. Y. Fujiwara, Y. Kohata and M. Shiba, "Gastroesophageal reflux disease in patients with diabetes: Preliminary endoscopic study," J. Gastroenterol. Hepatol., vol. 30, no. 7, pp. 1120–1125, 2015.
26. F. Lorentzen, N. Hovdenak and J. K. Hertel, "Erosive esophagitis and GERD symptoms in obese patients with and without type 2 diabetes," Surg. Obes. Relat. Dis., vol. 16, no. 9, pp. 1285–1291, 2020.
27. R. Kumar, S. Gupta and A. Taneja, "Interactions between gastroesophageal reflux disease and diabetes mellitus: A systematic review of pathophysiological insights and clinical management strategies," Nat. Rev. Endocrinol., vol. 20, no. 2, pp. 87–96, 2024.
28. B. Dixon, T. S. Davis and J. A. Hall, "Association between diabetes and esophageal cancer independent of obesity in a veteran population," Dis. Esophagus, vol. 28, no. 6, pp. 524–531, 2015.
29. M. Shuai and S. C. Larsson, "Genetic and lifestyle contributions to GERD risk: A Mendelian randomization study," Int. J. Epidemiol., vol. 51, no. 3, pp. 895–904, 2022.
30. C. J. Creighton and S. M. Hanash, "Mining gene expression data by interpreting association rules," Bioinformatics, vol. 19, no. 1, pp. 79–86, 2003.
31. M. Khalid, S. Khan, J. Ahmad and M. Shaheryar, "Identification of self-regulatory network motifs in reverse engineering gene regulatory networks using microarray gene expression data," IET Syst. Biol., vol. 13, no. 2, pp. 55–68, 2019, doi: 10.1049/iet-syb.2018.5001.
32. M. Khalid, S. Khan, J. Ahmad and M. Shaheryar, "Multivariate Covariance using Principal Component Analysis for Reconstruction of Bidirected Gene Regulatory Networks," in Proc. Int. Conf. Front. Inf. Technol. (FIT), Islamabad, Pakistan, 2017, pp. 229–234, doi: 10.1109/FIT.2017.00048.

33. A. K. Chandanan and M. K. Shukla, "Removal of Duplicate Rules for Association Rule Mining from Multilevel Dataset," Procedia Comput. Sci., vol. 45, pp. 659–666, 2015, doi: 10.1016/j.procs.2015.03.106.
34. C. K. Kwoh and X.-L. Li, "Recent advances in network-based methods for disease gene prediction," Brief. Bioinform., vol. 22, no. 4, Jul. 2021, Art. no. bbaa303, doi: 10.1093/bib/bbaa303.
35. T. C. Wiegers, A. P. Davis, K. B. Cohen, et al., "Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD)," BMC Bioinformatics, vol. 10, p. 326, 2009, doi: 10.1186/1471-2105-10-326.
36. J. Piñero, J. M. Ramírez-Anguita, J. Saüch-Pitarch, et al., "The DisGeNET knowledge platform for disease genomics: 2019 update," Nucleic Acids Res., vol. 48, no. D1, pp. D845–D855, 2020, doi: 10.1093/nar/gkz1021.
37. J. Chen et al., "Gastrointestinal consequences of type 2 diabetes mellitus and impaired glycemic homeostasis: A Mendelian randomization study," Diabetes Care, vol. 46, pp. 828–838, 2023, doi: 10.2337/dc22-1791.
38. C. H. Lee et al., "High glucose induces human endothelial dysfunction through an AXL-dependent mechanism," Cardiovasc. Diabetol., vol. 13, p. 53, 2014, doi: 10.1186/1475-2840-13-53.
39. J. Hong, S. Maacha and A. Belkhiri, "Transcriptional upregulation of c-MYC by AXL confers epirubicin resistance in esophageal adenocarcinoma," Mol. Oncol., vol. 12, pp. 2191–2208, 2018, doi: 10.1002/1878-0261.12395.
40. L. Li, G. Ma, C. Jing and Z. Liu, "Guanylate-binding protein 1 (GBP1) promotes lymph node metastasis in human esophageal squamous cell carcinoma," Discov. Med., vol. 20, pp. 369–378, 2015. PMID: 26760981.
41. N. Wang et al., "Molecular investigation of candidate genes for pyroptosis-induced inflammation in diabetic retinopathy," Front. Endocrinol., vol. 13, p. 918605, 2022, doi: 10.3389/fendo.2022.918605.
42. A. Vennemann et al., "PTGS-2–PTGER2/4 signaling pathway partially protects from diabetogenic toxicity of streptozotocin in mice," Diabetes, vol. 61, pp. 1879–1887, 2012, doi: 10.2337/db11-1396.

43. L. Yu et al., "E series of prostaglandin receptor 2-mediated activation of extracellular signal-regulated kinase/activator protein-1 signaling is required for the mitogenic action of prostaglandin E2 in esophageal squamous-cell carcinoma," J. Pharmacol. Exp. Ther., vol. 327, pp. 258–267, 2008, doi: 10.1124/jpet.108.141275.
44. H. S. H. H. Al-Khalidy, W. H. Salih and B. M. Mahdi, "Galectins: A new frontier in gastroesophageal reflux disease research," Arch. Med. Res., vol. 56, p. 103195, 2025, doi: 10.1016/j.arcmed.2025.103195
45. J. Milosevic et al., "Potential protective role of Galectin-3 in patients with gonarthrosis and diabetes mellitus: A cross-sectional study," Int. J. Environ. Res. Public Health, vol. 19, p. 11480, 2022, doi: 10.3390/ijerph191811480.
46. M. Di Rosa and L. Malaguarnera, "Chitinase 3-like-1: An emerging molecule involved in diabetes and diabetic complications," Pathobiology, vol. 83, pp. 228–242, 2016, doi: 10.1159/000444855.
47. J. Huang et al., "CHI3L1 (Chitinase 3 Like 1) upregulation is associated with macrophage signatures in esophageal cancer," Bioengineered, vol. 12, pp. 7882–7892, 2021, doi: 10.1080/21655979.2021.1974654.
48. P. J. Blair et al., "The role of prostaglandins in disrupted gastric motor activity associated with type 2 diabetes," Diabetes, vol. 68, pp. 637–647, 2019, doi: 10.2337/db18-1064.
49. T. Soma et al., "Induction of prostaglandin E synthase by gastroesophageal reflux contents in normal esophageal epithelial cells and esophageal cancer cells," Dis. Esophagus, vol. 20, pp. 123–129, 2007, doi: 10.1111/j.1442-2050.2007.00657.
50. E. Y. Chen et al., "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update," Nucleic Acids Res., vol. 44, pp. W90–W97, 2016, doi: 10.1093/nar/gkw377.
51. D. Szklarczyk et al., "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," Nucleic Acids Res., vol. 47, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.
52. M. Franz et al., "GeneMANIA update 2018," Nucleic Acids Res., vol. 46, pp. W60–W64, 2018, doi: 10.1093/nar/gky3.