CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Integrative In-Silico Exploration of Non-HLA Genetic and Molecular Determinants Underlying Celiac Disease Pathogenesis

by

Saima Bibi

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences

Department of Bioinformatics and Biosciences

2026

CERTIFICATE OF APPROVAL

Integrative In-Silico Exploration of Non-HLA Genetic and Molecular Determinants Underlying Celiac Disease Pathogenesis

by

Saima Bibi

(MBS243017)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Uzma Abdullah	AAU, Rawalpindi
(b)	Internal Examiner	Dr. Sami Ullah Jan	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
March, 2026

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. & Biosciences
March, 2026

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
March, 2026

Author's Declaration

I, **Saima Bibi** hereby state that my MS thesis titled "**Integrative In-Silico Exploration of Non-HLA Genetic and Molecular Determinants Underlying Celiac Disease Pathogenesis**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

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

(Saima Bibi)

Registration No: MBS243017

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Integrative In-Silico Exploration of Non-HLA Genetic and Molecular Determinants Underlying Celiac Disease Pathogenesis**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

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

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

(Saima Bibi)

Registration No: MBS243017

Acknowledgement

All praise and gratitude are due to **Allah Almighty**, the Most Merciful and the Most Compassionate, for granting me the strength, patience, and guidance to successfully complete this thesis. Without His countless blessings, this work would not have been possible.

I would like to express my deepest gratitude to my **supervisor, Dr. Marriam Bakhtiar**, for her valuable guidance, constant encouragement, and constructive feedback throughout the course of this research. Her expertise, support, and motivation played a vital role in shaping this thesis. I am also sincerely thankful to my **PhD fellow, Ms. Saba Farooq**, for her kind assistance, insightful suggestions, and continuous support during this research work.

I am profoundly grateful to my **family**, especially my **beloved parents**, whose prayers, sacrifices, and unwavering support have been the foundation of my academic journey. Their encouragement and belief in me gave me the strength to overcome every challenge. I am also thankful to my **brother** for his moral support and motivation throughout this period.

Lastly, I extend my appreciation to my **department and all those who directly or indirectly contributed** to the completion of this thesis.

(Saima Bibi)

Abstract

Celiac disease is a chronic autoimmune enteropathy that develops as a consequence of gluten consumption in genetically susceptible individuals. About 35-40% of genetic predisposition is accounted for by Human Leukocyte Antigen class II haplotypes DQ2 and DQ8 (HLA-DQ2/HLA-DQ8); however, there is still a large amount of heritability that is not explained. Genome-wide association studies have discovered over thirty non-HLA loci, such as Interleukin-2/Interleukin-21 (IL2/IL21), SH2B adaptor protein 3 (SH2B3), T-cell activation Rho GTPase activating protein, regulator of G-protein signaling 1 (RGS1), C-C motif chemokine receptor gene clusters and cytotoxic T-lymphocyte associated protein 4, although their functional contribution to immune system dysfunction and intestinal injury is still not well understood. To this end, an integrative in silico multi-omics approach was used, which included genomic variant discovery, functional annotation, regulatory element mapping, protein-protein interaction network modeling, pathway enrichment analysis, and network topology evaluation. Network analysis identified key convergence in cytokine signaling, specifically in the Interleukin-2/Interleukin-21 signaling axis (IL-2/IL-21 axis), and janus kinase-signal transducer and activator of transcription signaling in SH2B3. Enrichment was also found in immune cell migration via C-C motif chemokine receptors 1, 2 and 3 and T-cell activation signaling in TAGAP and the Cluster of Differentiation 28/cytotoxic T-lymphocyte-associated protein 4 at CD28,CTLA4 region. These included pathways involved in T helper type 1mediated inflammation, epithelial stress response, activation of intraepithelial lymphocytes, and increased intestinal permeability. In conclusion, this article offers a systems biology approach to non-HLA-related factors in CD and highlights regulatory nodes as potential biomarkers and therapeutic targets. Future work incorporating single-cell multi-omics and personalized genomic profiling could allow for stratified risk assessment and mechanism-based precision therapies.

Keywords: Celiac disease; non-HLA genetic loci; multi-omics integration; genome-wide association studies; cytokine signaling; protein - protein interaction network; precision medicine.

Contents

List of Figures

List of Tables

Abbreviations

ATAC-seq	Assay for Transposase-Accessible Chromatin using Sequencing
DGP	Deamidated Gliadin Peptide
EATL	Enteropathy-Associated T-Cell Lymphoma
EmA	Endomysial Antibody
Eqtl	Expression Quantitative Trait Locus
GFD	Gluten-Free Diet
IEL	Intraepithelial Lymphocyte
LPP	Lipoma Preferred Partner protein
MYO9B	Myosin IXB
NKG2D	Natural Killer Group 2 Member D (activating receptor)
PLEK	Pleckstrin gene
RCD	Refractory celiac disease
RGS21	Regulator of G-Protein Signaling 21
TAGAP	T-Cell Activation Rho GTPase-Activating Protein
TNFAIP3	Tumor Necrosis Factor Alpha–Induced Protein 3
tTG / TG2	Tissue Transglutaminase (Type 2 Transglutaminase)

Chapter 1

Introduction

Celiac disease represents lifelong immune system driven intestinal disease activated by gluten consumption in genetically susceptible persons. Gluten consists of prolamin protein components, encompassing gliadin, glutenin found in hordein, wheat present in secalin, barley occurring in rye all of which demonstrate resistance to full enzymatic breakdown owing to their elevated proline and glutamine composition [1]. In susceptible individuals these partially digested protein fragments interact with the intestinal lining, initiating an abnormal adaptive immune response that results in the destruction of villi, increased crypt cell growth, and ongoing inflammation of the small intestine. Historically, Samuel Gee first described celiac disease in 1888, identifying chronic diarrhea and nutritional deficiencies as key symptoms in children. The connection to gluten was identified in the mid-20th century, particularly during the wheat shortages of World War II, which resulted in symptom relief for those affected [2]. Over time, the understanding of CD has evolved from being viewed as a localized disorder causing malabsorption to being recognized as a multisystem autoimmune disease. It is influenced by genetic, ecological, immunological and microbial parameters affecting organ systems beyond gastrointestinal tract [3]. Celiac disease is estimated to occur one to two percent among the population in Europe and North America, although it can be affected by various considerations. These include geographical distribution, age factors, eating habits, and diagnostic approaches. According to serological studies,

it was found that prevalence is estimated to be relatively high compared to those diagnosed through biopsy [4].

It was found that research studies showed an increase in the number of diagnoses for celiac disease in areas known to contain a population of low incidence, such as the Asian regions, Middle Eastern countries. This can be explained by growing awareness for diagnosis, a considerable amount of consumption, as well as thorough genetic analysis. Studies showed that it was found to occur more among females as compared to males with a female to male ratio of 2:1. It was found that it was not restricted to a particular age group; instead, it was seen to increase among fifty to sixty-year-old adults, contrary to the fact that it was prevalent among children exclusively [5].

Although it is not a prevalent condition among children, due to these changes it is important to assess CD for symptoms that occur among adults. Studies on Celiac disease is recognized as a global public health concern. A crucial step data analysis was found that the overall global serologic prevalence indicated by positive anti-tissue transglutaminase and anti-endomysial antibodies, was approximately 1.4 percent (95% CI: 1.1-1.7 percent). In contrast prevalence confirmed through biopsy was about 0.7 percent (95% CI: 0.5-0.9 percent) among more than 275,000 individuals from different regions. The study also pointed out regional variations, with Asia showing a serologic prevalence of around 0.6% and Europe/Oceania at about 0.8% [2]. A distinct meta-analysis focusing on the Asia-Pacific region found that the overall sero-prevalence in the overall population was around 1.2%, while the prevalence confirmed by biopsy in low-risk groups was about 0.61% [4].

In Pakistan the available epidemiological data are limited and varied. A review highlighted that the prevalence of celiec disease in the overall population is not well-documented, although it is considered to be a very common disorder with estimates suggesting that as many as 90% of cases might go undiagnosed [5]. Hospital-based studies in specific sub-populations reveal higher rates: for example, among adults with iron deficiency anemia in Peshawar and Swabi, a study found a prevalence of approximately 5.4% for positive IgA tTG screening, with 3.35% confirmed on duodenal biopsy [5]. Another survey in children suspected of

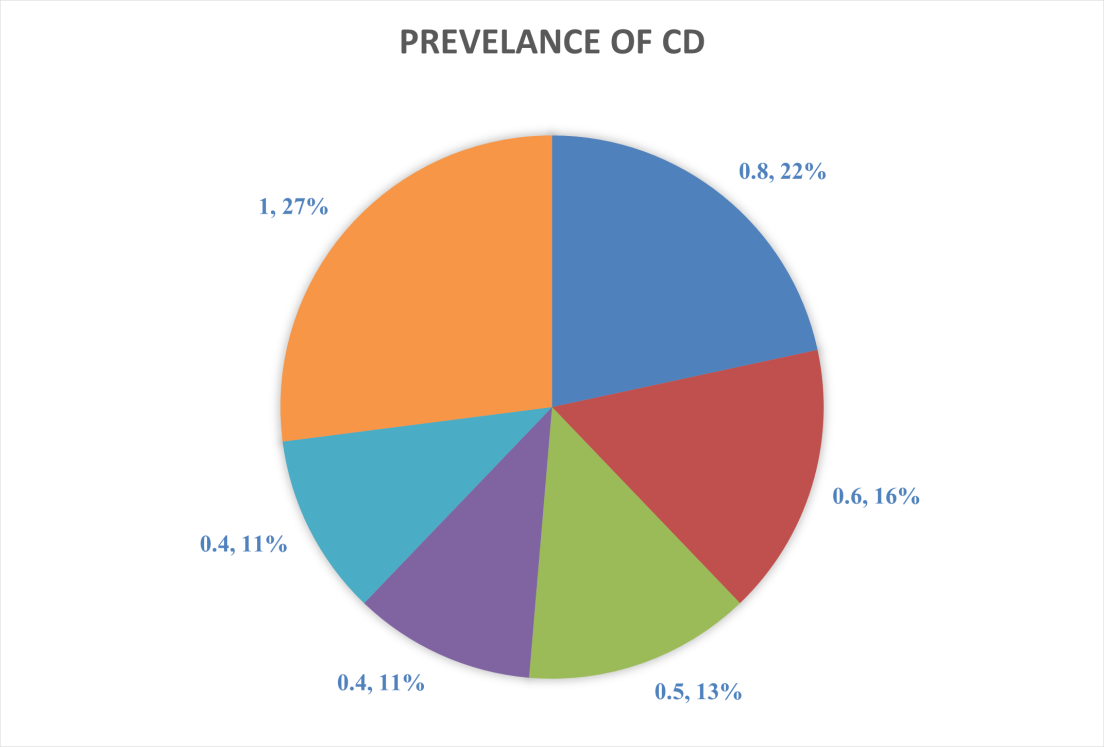


FIGURE 1.1: Prevalence of celiac disease [4, 5]

CD in Lahore reported 17.8% anti-tTG positivity [6]. These findings suggest that while true community-based prevalence remains to be established in Pakistan, the burden particularly in at-risk groups may be considerable. Hence, the clarity need for large representative population size based screening research and for raising clinical awareness, especially given the potential for serious complications in undiagnosed cases.

Hereditary parameters are crucial in causing celiac disease. Moreover, most well documented risk elements include HLA class II molecules, specifically the HLA DQ8 and HLA DQ2 heterodimers that are coated by DQB1 and DQA1gene loci located on chromosome 6p21 [7] [8]. Approximately 90-95 percent of individual with celiac disease carry HLA DQ2 while most of the other cases have HLA DQ8. These specific molecules can bind to deamidated gluten peptides, allowing them to be presented to CD4 T cells and initiating an inflammatory response in the small intestine. Although HLA DQ2/DQ8 is present in percentage of 30-40 to general population only a small percentage actually occurs the disease indicating the necessity for other genetic and environmental factors [9].

Beyond HLA alleles genetic regions also contribute to determining susceptibility and clinical variability. GWAS have identified more than 40 non-HLA genetic regions linked to CD highlighting genes vital for regulating immune responses facilitating cytokine signaling pathways and maintaining strength of epithelial barrier. Key non-HLA genes include the following

1.0.1 IL2 and IL21 Locus Chromosome 4q27

This region plays a key role in generating interleukin 2 and interleukin 21, two cytokines essential for regulating the equilibrium between regulatory and effector T cells. Any disturbance in their regulation can disrupt immune tolerance, increase effector T-cell activity and result in the production of autoantibodies.

1.0.2 SH2B3 LNK Gene Chromosome 12q24

Acts as a blocker of cytokine signaling via the JAK-STAT pathway. Variants may lower the activation threshold of T cells, resulting in increased inflammatory responses to gluten.

1.0.3 CCR Gene Cluster Chromosome 3p21

The genes for chemokine receptors CCR1, CCR2, and CCR3 are responsible for directing the movement of leukocytes. Variations in these genes can increase the tendency of lymphocytes to migrate to mucosal areas, which plays a role in villous atrophy and ongoing inflammation [10].

Beyond HLA variants other genetic factors identified including TAGAP, RGS1, IL12A, and TNFAIP3. Many of genetic variants are connected to other autoimmune disorders like type 1 diabetes, autoimmune thyroid disorders and rheumatoid arthritis.This shared genetic architecture indicates that CD may be part of a wider range of immune system disorders with common genetic factors. [11].

The combination of HLA and non-HLA genes contributes to a polygenic risk score stratify disease susceptibility. Genetics alone does not suffice to cause disease; environmental factors, the consumption of gluten, serve as triggers. The extent to which the disease manifests is affected by factors like method of delivery the length of breastfeeding, the timing of gluten introduction, early exposure to infections and composition of gut microbiota. epigenetic processes like DNA methylation and histone modifications are involved in regulating the expression of immune-related genes and may account for differences in individuals with a genetic predisposition [11] [10].

TABLE 1.1: Genetic Basis of celiac disease [10, 11]

Sr.#	**Features**	**HLA Genes**	**Non-HLA Genes**
1	Types	MHC Class II (DQ2, DQ8)	Immune & epithelial regulatory genes
2	Roles	Present gluten peptides to $CD4^+$ T cells	Amplify immune response, regulate cytokines, maintain barrier integrity
3	Main genes	DQ2 (DQA105 / DQB102) DQ8 (DQA103 / DQB103:02)	IL2, IL21, SH2B3, TAGAP, MYO9B, CTLA4, TNFAIP3, LPP
4	Effect	Initiates celiac disease	Modify disease severity and chronicity
5	Contribution to genetic risk	∼35–40%	∼60 - 65%

The gut microbiome, according to studies, is one of the crucial elements involved in the development of CD. Analyses of microbial communities in both fecal and duodenal samples show changes in composition and metabolic activity in those with CD. Key observations include reduced abundance of protective species, including genera such as Bifidobacterium and Lactobacillus, which produce SCFAs

involved in maintaining epithelial barrier integrity. Increased prevalence of pro-inflammatory species, including Bacteroides and certain Proteobacteria, which may enhance gluten immunogenicity. Microbial enzymes can break down gluten peptides, thereby either dampening or amplifying immune activation based on the species and enzymatic activity .The microbiome also interacts with innate immune pathways, influencing epithelial stress responses and cytokine production. Although gluten-free diets can help restore some microbial balance they might also decrease microbial diversity, making it challenging to determine whether dysbiosis is a cause or a result of CD [20]. Environmental factors play an essential role in CD onset and phenotypic diversity [21].

These factors include: Introduction of gluten, either early or very late in infancy, can affect oral tolerance. The nature of infant feeding, whether introducing the infant to gluten while exclusively nursing, may confer a degree of protection against CD. Delivery mode, which means an infant born through a C-section, causes microbial differences in gut microbiota.

Viral infections, specifically rotavirus and reovirus, affecting the gastrointestinal system, can cause disturbances in oral tolerance, immune system activation, and, as a potential result, CD development. Certain dietary factors, including the nature of complementary foods, nutritional content, and overall dietary exposure to gluten, can contribute to the probability of developing CD [22] [23].

More recent evidence points to thefact that epigenetics, DNA methylation, histone acetylation, non-coding RNAs, can profoundly affect HLA and non-HLA genes associated with CD [24]. Environmental factors, such as infection or diet, can be responsible for epigenetics, potentially influencing immune system mechanisms, thereby modulating CD initiation, progression, and severity [25].

Celiac disease is a complex immune-mediated disorder with a strong genetic basis and a known environmental trigger, namely gluten. It affects about 1% of the global population, with increased prevalence in first-degree relatives (2.8-22.5%) and a concordance rate of 85% in monozygotic twins, suggesting a strong genetic trait. The familial aggregation of celiac disease causes a relative risk of tenfold

susceptibility, which emphasizes the influence of genetic components along with other factors.

The mode of inheritance of this condition is non-Mendelian, suggesting a polygenic trait.The main genetic factor is found in the human leukocyte antigen HLA region on chromosome 6, specifically the class II molecules HLA-DQ2 DQA10501, DQB10201 and HLA-DQ8 DQA10301, DQB10302 which together account for about 40% of the genetic risk. Over 90% of patients express HLA-DQ2, while most of the remaining patients have HLA-DQ8 in either cis or trans configurations (such as DR3/DR5 or DR5/DR7 heterozygotes). Individuals who are homozygous for DR3-DQ2 or heterozygous for DR3-DQ2/DR7-DQ2 show increased disease susceptibility, likely due to a gene dosage effect that boosts DQ2 expression levels.

Beyond HLA other genes contribute modestly to susceptibility, including CTLA-4/CD28 and MYO9B which may influence immune regulation and intestinal barrier integrity, although their impact varies across populations. Overall, HLA genes account for nearly half of the heritable risk, while non-HLA genes provide additional but smaller contributions [26].

At the molecular level, the onset of disease occurs when gluten fragments pass through the intestinal barrier and are altered by tissue transglutaminase an enzyme that converts specific glutamine residues into glutamic acid through a process known as deamidation. This alteration enhances negative charge of gluten fragments improving to interact with DQ2 and DQ8 present on antigen-presenting cells. Once bound, fragments are identified by $CD4^+$ T cells in intestinal lamina propria initiating a Th1 skewed immune response marked by the release of interferon gamma IFN-γ [27]. Many gluten-derived epitopes those from α- and γ-gliadin requiring deamidation by tTG for immune recognition. gluten-specific T cells promote generation of anti-tTG autoantibodies through an intramolecular help mechanism enabling gliadin-reactive T cells to stimulate B cells. This mechanism explains the decrease in antibody levels when gluten is eliminated from the diet. Therefore, the molecular pathogenesis of CD is defined by a tightly linked sequence involving genetic susceptibility, gluten exposure, enzymatic modification

by tTG and abnormal immune activation, ultimately resulting in intestinal autoimmunity [28].

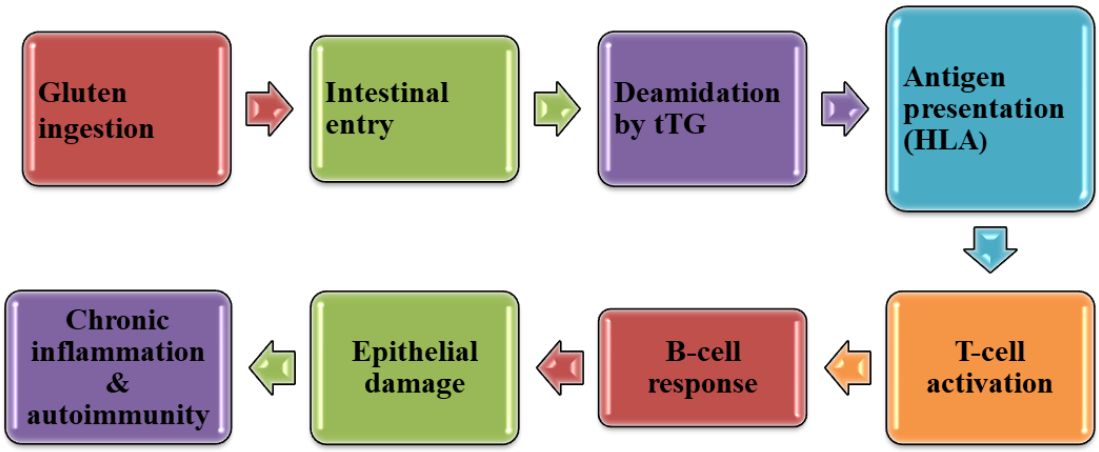


FIGURE 1.2: Overview of Immune Pathogenesis of celiac disease

Celiac disease presents a wide array of clinical manifestations, ranging from typical gastrointestinal problems to less common or even symptomless forms. Typical symptoms encompass persistent diarrhea, fatty stools, abdominal bloating, weight loss, and in children, growth delays and delayed puberty due to inadequate nutrient absorption. Frequent deficiencies in iron, folate, vitamin B12, calcium and fat-soluble vitamins can result in complications [29].

Adults frequently display non-classical or extraintestinal symptoms such as unexplained anemia, osteopenia, neuropathy, gluten ataxia, headaches and cognitive impairments. [30]. Silent or subclinical forms of disease are detected through screening in high-risk populations. Although these people do not show clear symptoms, they remain susceptible to long-term complications if the disease is left untreated [31]. If left untreated, celiac disease can cause serious complications as a result of malabsorption, which causes a continuous level of micronutrient deficiencies, anemia, osteoporosis, and fractures [32]. The risk of developing an autoimmune disease also increases [33]. In some rare cases, the intestinal inflammation can, over time, develop into serious complications, including EATL or adenocarcinoma of the small intestine. Apart from the physical burden, the disease also causes serious psychological disturbances, thereby contributing to anxiety, depression, and a poor quality of life [32] [25].

Serological evaluations using IgA anti-tissue transglutaminase (tTG) and anti-endomysial antibodies (EmA) are known for their high sensitivity and specificity. In contrast, IgG-based tests, such as those for deamidated gliadin peptides, are employed for patients who are deficient in IgA [33]. Histological analysis of duodenal biopsies continues to be the definitive method for identifying villous atrophy, crypt hyperplasia and increased intraepithelial lymphocytes, as classified by Marsh Oberhuber. According to ESPGHAN guidelines, in children, a diagnosis without biopsy is possible in certain cases where there are high tTG levels and positive EmA results. HLA typing for HLA-DQ2/DQ8 aids in diagnosis, as a negative result rules out CD although a positive result is not conclusive due to its common occurrence in the general population. New diagnostic methods, such as detecting gluten immunogenic peptides in stool or urine and T-cell-based assays are promising for monitoring adherence and identifying refractory disease though they are primarily research-focused at this stage [33] [34].

The main strategy for managing celiac disease is to strictly follow a gluten-free diet for life. This approach typically leads to symptom relief, normalization of serological markers and the restoration of the intestinal lining. Even a trace of gluten can cause the immune system to react, increase the time span of inflammation, and slow the healing of the intestine [35]. Even if the gluten-free foods are effective, they can be a daunting experience to the patient because of the widespread nature of gluten consumption and the possibility of contamination. It is important to note that many gluten-free foods are often deficient in fiber, iron, and some vitamins, which is a burden to those who are deficient [36] [37].

Because of these limitations, studies focus on complementary therapies. Gluten-degrading enzymes, latiglutenase ALV003, are created to break down harmful gluten peptides before they can stimulate the immune system. Barriers modulators, such as larazotide acetate, focus on lowering permeability and inflammation. Immune system modulators include peptide vaccines, for instance, Nexvax2, as well as monoclonal antibodies that focus on modulating the inflammation pathway, with the goal of inducing gluten tolerance. Although these are still under research, these therapies may help complement nutritional management for patients

[36] [38].Some patients continue to have symptoms even when they strictly follow a gluten-free diet, a condition known as non-responsive celiac disease. The more severe form called refractory celiac disease is categorized into type I and type II. Type II is characterized by clonal growth of abnormal intraepithelial lymphocytes and risk of developing enteropathy-associated T-cell lymphoma EATL. Treating RCD involves use of corticosteroids, immunosuppressive or cytotoxic therapies. Prompt referral to tertiary care centers is crucial for comprehensive evaluation and treatment [39].

Celiac disease is becoming a significant public health issue in Asia including Pakistan where it often goes undiagnosed due to a lack of awareness, insufficient diagnostic resources and diets rich in wheat. The delay in diagnosis can result in nutritional deficiencies and increased healthcare costs. Proper management necessitates availability of gluten-free products, precise labeling, and dietary guidance that is culturally appropriate. Education, community support and multidisciplinary care can enhance adherence and alleviate stress. There is debate over population-wide screening due to concerns about overdiagnosis and the economic impact. [40][41]. Current research efforts are combining genomics, microbiome studies and immunology to create targeted treatments. Personalized approaches may enhance quality of life. Preventive strategies such as optimizing timing of gluten introduction and decreasing gluten's immunogenic properties through food processing show potential [42].

Celiac disease serves as an example of a multifaceted autoimmune condition. Although a gluten-free diet remains the cornerstone of treatment, ongoing translational research is exploring various supplementary or alternative therapies that could potentially enhance long-term outcomes and lessen the challenges of adhering to a lifelong dietary regimen [34] .

Utilizing in silico tools is essential for revealing the genetic factors of celiac disease that remain unexplained by the well-known HLA-DQ2 and HLA-DQ8 alleles. These methods facilitate the large-scale integration of genomic, transcriptomic and proteomic data which in turn aids in identifying new susceptibility genes, molecular pathways that play a role in the disease's development. Researchers

can use these tools to predict the functional impacts of genetic variants, examine protein structures and interactions and investigate regulatory networks that lead to immune system dysregulation and intestinal barrier issues in CD. Additionally, in silico approaches are both cost-effective and time-saving, enabling the generation of hypotheses and the prioritization of potential targets for experimental validation based on data. Therefore, applying bioinformatics and computational biology to the study of non-HLA genes offers a comprehensive framework for understanding disease mechanisms and identifying new biomarkers and therapeutic targets for better diagnosis and treatment.

1.1 Problem Statement

Celiac disease is a persistent autoimmune disorder affecting the intestines triggered in individuals with a genetic predisposition when they consume gluten. While the HLA-DQ2 and HLA-DQ8 alleles contribute to about 35-40% of the genetic risk other genetic factors outside the HLA region also play a role by influencing immune regulation and disease susceptibility. Genome-wide association studies have identified several potential loci such as IL2/IL21, SH2B3 and CCR gene clusters but their exact functional roles and the molecular mechanisms involved remain largely unclear. The absence of comprehensive in silico analyses that connect these non-HLA variants to biological pathways hinders our understanding of the pathogenesis of celiac disease and obstructs the identification of new biomarkers and therapeutic targets for accurate diagnosis and treatment..

1.2 Research Gap

Previous research has primarily focused on HLA-related genetic determinants of celiac disease, leaving the non-HLA component underexplored. Although GWAS have identified several non-HLA risk loci, these findings remain fragmented, underanalyzed, and poorly integrated across molecular data types.

There is currently no comprehensive in silico framework that integrates genomic, transcriptomic, and protein-interaction data to understand how non-HLA variants contribute to immune dysregulation and intestinal damage and there is no study have been integrated genomic and protein interaction data of non-HLA genes in celiac disease. This gap limits our ability to translate genetic associations into meaningful biological and clinical insights.

1.3 Research Questions

i. Which non-HLA genetic variants and genes are significantly associated with celiac disease according to genomic and literature data?

ii. What are the functional and regulatory roles of these genes in immune signaling and intestinal epithelial health?

iii. How can an integrative in silico multi-omics approach uncover novel molecular pathways and biomarkers involved in celiac disease pathogenesis?

1.4 Scope of the Study

This study will be conducted entirely in silico using publicly available genomic and molecular datasets. The research will include:

i. Data mining and literature-based identification of non-HLA variants and genes associated with celiac disease.

ii. Functional and pathway annotation to predict gene functions, expression impacts, and regulatory roles.

iii. Integration of multi-omics datasets (genomic, transcriptomic, proteomic, and interactomic) to develop a comprehensive molecular network.

iv. Bioinformatics and network analysis to identify key genes, regulatory hubs, and pathways contributing to disease mechanisms.

Experimental validation will be beyond the scope of this study but can be pursued in future research based on computational findings.

1.5 Aim

To interpret non-HLA genetic and molecular mechanisms underlying celiac disease through integrative in silico analyses supporting discovery of novel biomarkers, pathways and potential therapeutic targets

1.6 Objectives

i. To identify and prioritize non-HLA genes associated with celiac disease through data mining and genomic analysis.

ii. To perform functional and regulatory annotation of these genes to understand their biological significance and interactions in immune pathways.

iii. To integrate multi-omics datasets to uncover novel molecular networks and key determinants driving celiac disease pathogenesis.

1.7 Impact on Society

This study will greatly improve the comprehension of genetic intricacies in celiac disease and benefit public health by discovery of **new molecular insights** into disease mechanisms beyond traditional HLA genetics. Identification of **potential diagnostic biomarkers** for early detection and better disease management. Laying the foundation for **personalized therapeutic strategies** based on an individual's genetic and molecular profile. Advancing knowledge applicable to other **autoimmune disorders** with shared genetic pathways. Supporting **public health awareness** and improved diagnostic accuracy, reducing the burden of undiagnosed and mismanaged celiac disease cases globally.

Chapter 2

Literature Review

2.1 Celiac Disease

Celiac disease is an autoimmune enteropathy that affects the intestines in individuals with a genetic predisposition, most commonly in individuals with the HLA-DQ2 or HLA-DQ8 haplotype. It is triggered by gluten ingestion, a protein complex in wheat, barley, and rye. Following gluten ingestion, an abnormal immune response is activated where the immune cells attack the epithelial cells of the small intestine, resulting in damage to these cells through both antibody and T-cell mediated mechanisms [35]. Immune damage results in villous atrophy, crypt hyperplasia, and mucosal inflammation that is then quantified according to the degree of severity based on bothMarsh (I-III) and Oberhuber (I-IV) systems of classification. It is estimated that in the world, celiac disease affects about 1% of all populations with a higher prevalence among women (1.4%) compared to that observed in men (0.7%). It has also been estimated to affect a total of 17.4 per 100,000 person-years in women and 7.8 per 100,000 in men [36]. It has also been observed to manifest with a wide range of symptoms that range from general diarrhea, oily droppings, and inability to absorb nutrients to abdominal intestinal pains in women including anemia due to decreased bone density, reproductive abnormalities also.

An important milestone in understanding celiac disease was the identification of tissue transglutaminase as the predominant autoantigen, which established its autoimmune etiology. Apart from genetic predisposition, several environmental factors like viral infections, infancy exposure to gluten, and changes in microbiota patterns within the gastrointestinal tract are considered to trigger the disease. The diagnosis of this condition is made through a set of serologic assays such as anti-tTG, anti-endomysial, and anti-deamidated gliadin peptide antibodies, supplemented by histopathological examination of biopsy samples of the small intestine, which remains the standard for diagnosis of this condition to date. Currently, the only proven management strategy for this condition is a strict, lifelong gluten-free diet that helps in mucosal healing, managing symptoms, and preventing potential post-disease complications such as enteropathy-associated lymphoid tissue lymphoma, ulcerative jejunoileitis, and others. However, this poses potential nutritional deficits due to restricted dietary practices, warranting comprehensive nutritional advice by healthcare professionals [38].

The rise in gluten-free diets among individuals who do not have CD, largely due to concerns or possible gluten sensitivity, draws attention to the need for clearer discrimination between CD, gluten-related disorders, and non-CD conditions. A balanced gluten-free diet is crucial for optimal health for individuals with this condition to prevent secondary deficiency symptoms related to malnutrition. Malnutrition during infancy can lead to conditions such as stunted growth or failure to thrive [39].

2.2 Normal Immune Responses in the Gastrointestinal Tract

An understanding of the normal immunological functions in the gut is important for the discovery of the mechanism of celiac disease occurrence. Gluten, the major environmental factor triggering the immunopathology, is composed of storage proteins referred to as gliadins and glutenins. These proteins are highly proline and

glutamine rich, making them not easily susceptible to degradation by enzymes. Subsequently, partially degraded peptides of gluten are persistently observed in the lumen of the gut, giving rise to the generation of several immunogenic peptides responsible for the induction of the immunologic response in the gut [3]. When the immunogenic peptides interact with the gut-associated lymphoid tissue, an important constituent of the immunologic framework of the gut, an immunologic response is induced, culminating in the characteristic damage observed in celiac diseases [38].

In the normal state, the immune system maintains a delicate balance between immune activation and tolerance. This aspect is of particular interest when considering the GALT, since the lining of the intestines is continuously challenged with antigenic substances from certain foods and a sea of resident microbes. The immune system needs to be able to distinguish and destroy pathogenic microbes while tolerant of other antigenic substances like foods. The process of immune tolerance is maintained by regulatory T cells (Tregs), and the condition of T cell anergy. This process of immune tolerance at the mucosal lining of the gut is maintained through the help of antigen dosages. Higher dosages lead to the development of allergy, while the process of immune tolerance leads to the differentiation of Tregs. The regulatory process of the immune system at the mucosal lining of the gut is facilitated by the action of regulatory T cells with the help of factors like transforming growth factor-β (TGF-β), and retinoic acids. The process of immune tolerance of the mucosal lining of the gut is further maintained by the action of the nuclear factor of activated T cells (NFAT), whose function maintains the lack of responsiveness of the orally[40].

2.3 Morbidity

Celiac patients are predisposed to various other autoimmune disorders. The propensity for these disorders has been mainly attributed to their genetic link and immune system dysfunction due to exposure to wheat proteins. Some commonly established associations with celiac disease are type 1 diabetes mellitus;

the incidence of CD has been shown to range between 5.1% and 6.0% in type 1 diabetes mellitus patients. A plethora of research carried out on a well-established base of data has revealed positive associations between CD and various other autoimmune disorders such as autoimmune thyroiditis, juvenile idiopathic arthritis, rheumatoid arthritis, psoriasis, and inflammatory bowel disease.

However, it is still unclear whether having undiagnosed or untreated celiac disease raises the likelihood of developing additional autoimmune conditions. If this connection is confirmed it would highlight the crucial need for early diagnosis and the prompt start of a gluten-free diet. Nonetheless, the current epidemiological evidence regarding this relationship is inconsistent. Celiac disease has been linked to various liver disorders, including autoimmune hepatitis, primary biliary cholangitis, and primary sclerosing cholangitis. Furthermore, individuals with CD seem to have a higher risk of metabolic dysfunction associated fatty liver disease a condition that may be partly affected by gluten-free diets which are often higher in sugar and fat content [47].

A meta-analysis that included cross-sectional, case control and prospective cohort studies found that 4.6% of individuals with cryptogenic cirrhosis have celiac disease a prevalence notably higher than in the general population. In addition to liver-related symptoms CD is associated with a range of systemic comorbidities, though a direct causal link is still uncertain. Studies based on population data have shown that people with CD have an increased risk of cardiovascular disease and a higher mortality rate from cardiovascular causes compared to matched controls. Furthermore, older adults with CD seem to be more prone to end-stage renal disease and frailty which adds to the overall disease burden [48].

2.4 Cancer and Mortality

Research in epidemiology suggests that people with celiac disease face a slightly elevated overall risk of developing cancer compared to the general population, with large national cohort studies reporting a hazard ratio of about 1:11. A particularly

strong link has been identified between CD and non-Hodgkin's lymphoma especially enteropathy-associated T-cell lymphoma. A nationwide case-control study in France found that individuals with CD were nearly four times more likely to develop NHL than those without the condition. Importantly, the risk of lymphoproliferative cancers is notably higher in patients who continue to show villous atrophy in follow-up biopsies, as opposed to those who experience mucosal healing after adopting a gluten-free diet.

Other cancers associated with CD include those of the small intestine, hepatobiliary system, pancreas, esophagus, stomach and non-melanoma skin, while several studies have noted a decreased incidence of breast cancer among those affected.

Individuals with CD face a slightly elevated risk of death mainly due to cancer. A comprehensive Swedish cohort study revealed a hazard ratio of 1:21 for overall mortality in celiac patients compared to matched controls, with increased mortality rates linked to cancer, cardiovascular issues and respiratory diseases. Similarly, a meta-analysis of 25 studies examining mortality in CD indicated a heightened risk of death from all causes and cancer, especially non-Hodgkin's lymphoma and respiratory illnesses. Positively, recent studies indicate that this mortality risk has diminished over time, likely as a result of earlier diagnosis, better management and more widespread use of gluten-free diets [49].

2.5 Types of Celiac Disease

Celiac disease also occurs with varying clinical manifestations, meaning there are differences regarding their manifestation, chronicity, and histopathologic findings. This is an index of the complex interplay among genetic predisposition, immunological activation, as well as gluten exposure [50].

The main types of CD include classical, non-classical, silent, latent, and refractory celiac disease. An understanding of these types is very important for accurate treatment because they all have different clinical and histopathologic manifestations [51].

TABLE 2.1: Types of Celiac Disease and Their Key Characteristics [52, 53]

Type	**Main Features**	**Common Symptoms**	**Histology**	**Diagnosis**	**Treatment / Prognosis**	**Ref.**
1. Classical (Typical)	Clear malabsorption with small intestinal mucosal damage.	Chronic diarrhea, steatorrhea, weight loss, growth failure (children), anemia.	Villous atrophy, crypt hyperplasia,elevated IELs.	Positive anti-tTG / EMA; biopsy confirmation.	Gluten-free diet (GFD), excellent recovery.	[52]
2. Non-Classical (Atypical)	Minimal or no GI symptoms; extra-intestinal features.	Fatigue, anemia, osteoporosis, infertility, neuropathy, dermatitis herpetiformis.	Variable villous atrophy.	Serology (anti-tTG, EMA) + biopsy.	GFD effective; often underdiagnosed.	[52]
3. Silent CD	Asymptomatic but abnormal serology and biopsy.	None; risk of osteoporosis, infertility, malignancy.	Villous atrophy with inflammation.	Positive serology + biopsy.	GFD to prevent complications.	[52]
4. Latent CD	Genetic predisposition (HLA-DQ2/DQ8); normal mucosa initially.	Usually asymptomatic; may develop later.	Initially normal; may progress to atrophy.	HLA typing, serologic follow-up.	Monitor; start GFD if active.	
5. Refractory CD (RCD)	Persistent symptoms despite $\geq$12 mo GFD.	Chronic diarrhea, weight loss, malnutrition.	Persistent villous atrophy; Type I normal IELs, Type II clonal	Exclude gluten contamination	Immuno suppressants	[52] [53]

2.6 Pathogenesis

Celiac disease is an inflammatory condition of the human small intestine due to an immune response involving both the innate immune system and the adaptive immune systems that causes persistent damage to it that lasts for life [53]. The immune response in celiac disease is initiated by certain peptides of gliadin that have crossed the epithelial barrier of the intestine and have been acted upon by an enzyme known as transglutaminase in the body to form glutamic acid by changing glutamine molecules [54]. This process involves changing glutamine molecules present in gliadin peptides to glutamic acids by the actions of an enzyme known as transglutaminase in celiac disease; then, this complex adheres firmly to HLA-DQ2/HLA-DQ8 molecules present in antigen-presenting cells; hence, the initiation of an immune response in celiac disease [54]. The complexes are then recognized by CD4 T-cells that are activated by an immune response in celiac disease; hence, an immune response that leads to an influx of T-cells that eventually multiply in huge numbers [2]. The T-cells activate an immune response by releasing cytokines such as interferon gamma and IL-15 that cause an immune response that leans towards T1 cells in celiac disease [2].

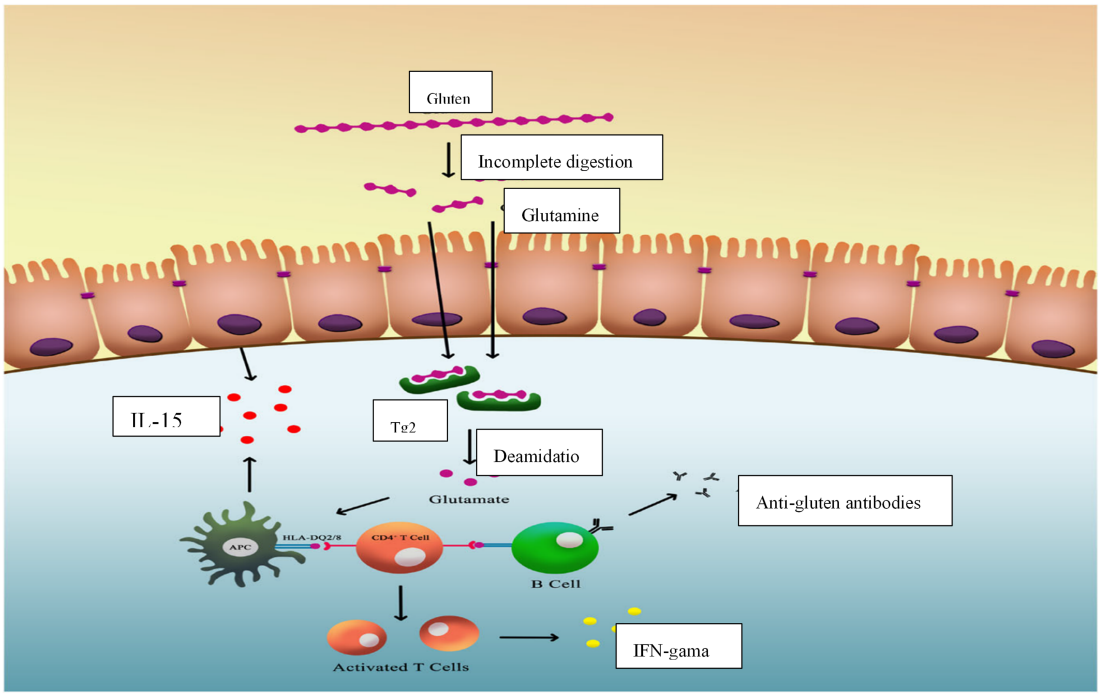


FIGURE 2.1: Pathogenesis of celeiec disease [31]

At the same time, activated $CD4^+$ T cells prompt B lymphocytes to generate autoantibodies specific to the disease, such as anti-tissue transglutaminase and anti-deamidated gliadin peptide antibodies. These antibodies form the humoral aspect of the immune response in celiac disease and are crucial diagnostic markers. Importantly, anti-tTG antibodies can interfere with transforming growth factor-beta (TGF-β) signaling, a pathway essential for preserving villous structure and aiding epithelial renewal, thus worsening villous atrophy. In the intestinal lining, gliadin peptides that resist enzymatic breakdown cross the epithelial barrier due to heightened intestinal permeability and activate innate immune cells like natural killer cells and intraepithelial lymphocytes with $\gamma\delta$ T-cell receptors. These cytotoxic cells trigger enterocyte apoptosis through NKG2D-MIC-A interactions, which ultimately results in epithelial damage, crypt hyperplasia and mucosal remodeling [15] [12] [55] [40].

While having a genetic predisposition particularly the HLA-DQ2 or HLA-DQ8 alleles is necessary for celiac disease to develop it is not enough on its own to trigger the condition. These alleles are present in about 40% of the general population, yet only around 2-5% of those with the alleles actually develop the disease. This highlights the significant roles that environmental factors and gut microbiota play in the pathogenesis of CD [14] [15]. The hygiene hypothesis posits that insufficient exposure to microbes during early childhood might impede the development of immune tolerance, making genetically predisposed individuals more likely to have an abnormal immune reaction when exposed to gluten [55] [40] [15].

2.7 Genetic Factors

The major genetic factor for the occurrence of celiac disease is the presence of the HLA DQ2 and HLA DQ8 antigen. The gene for this antigen is present in the major histocompatibility complex of the MHC class II region. This antigen produces around 35%-40% of the genetic predisposition for the disease [8]. More than 95% of the patients with celiac disease have the gene for the HLA DQ2 antigen. The remaining patients have the gene for the HLA DQ8. Nonetheless, since about

one-third of the general population carries these alleles and only 1-3% actually develop the disease, these genes are considered necessary but not sufficient for the disease to occur. Evidence from twin studies further supports the role of additional genetic contributors, as concordance rates remain below 100%. Genome-wide association studies (GWAS) have identified over 39 non-HLA loci associated with CD, encompassing genes involved in immune regulation and intestinal barrier function. Notable examples include MYO9B, which may influence epithelial tight junction integrity and the response to a gluten-free diet, and immune-related genes such as *IL2, IL21, SH2B3, CTLA4*, and *TAGAP*, which modulate cytokine signaling, T-cell activation, and immune tolerance. Both HLA-DQ2 and HLA-DQ8 heterodimers, expressed on antigen-presenting cells, present deamidated gluten peptides modified by tissue transglutaminase (tTG2) to $CD4^+$ T cells, leading to cytokine production (e.g., IFN-γ, IL-15) and subsequent intestinal inflammation. [8] [9] [7].

Genetic susceptibility involving HLA-DQ2 and HLA-DQ8 alleles, together with multiple non-HLA immune-regulatory genes, forms a critical foundation for the development of celiac disease (CD).

However, the clinical manifestation of the disorder typically depends on the interaction between these genetic factors and environmental influences such as dietary gluten exposure, intestinal infections, and gut microbiota dysbiosis, which act as essential cofactors in disease initiation and progression [18].

2.7.1 HLA Genes

The majority of CD patients exhibit the DRB10301 - DQA10501 - DQB1*0201 haplotype, also known as DR3 - DQ2, or they are heterozygous carriers of DRB111 / 12-DQA10505 - DQB1*0301 and DRB107 - DQA10201 - DQB1 * 0202 (collectively referred to as DR5-DQ7 / DR7-DQ2). The α-chain encoded by DQA10501* and DQA10505* differ by only a single amino acid residue in the leader peptide, while the β-chain encoded by DQB10201* and DQB10202* differs by one amino

acid in the membrane-proximal domain changes that are unlikely to have any functional effect. Therefore, individuals possessing these haplotypes express the same functional DQ molecule, whether the genes occur in cis (on the same chromosome) or trans (on opposite chromosomes).

Recombination within the HLA region plays a critical role in generating these haplotypes. Comparative analyses have revealed that DR3-DQ2, DR7-DQ2, and DR5-DQ7 share a close evolutionary relationship, with DNA fragments around the DQA1 and DQB1 genes being exchanged among them through recombination events.

The combination of DQA1*05 and DQB1*02 alleles regardless of their chromosomal arrangement is believed to be the main factor conferring susceptibility to CD, as they encode the DQ2 heterodimer responsible for gluten peptide presentation.

People who lack the combination of both DQA1*05 and DQB1*02 will have the DRB104 - DQA10301 - DQB1*0302 gene complex, known as DR4-DQ8. In this kind of association, the DQ8 antigen is considered the major contributing factor for susceptibility [57].

2.7.1.1 Interaction between HLA and Gluten

Recent investigations have elucidated the interplay between genetic susceptibility and environmental triggers in celiac disease. The HLA-DQ2 and HLA-DQ8 molecules, which confer risk and are encoded by particular alleles, play a key role due to their ability to bind gluten-derived peptides and present them to T lymphocytes in the intestinal epithelium. Interestingly, T cells from individuals with celiac disease recognize these gluten peptides only when they are presented by HLA-DQ2 or HLA-DQ8, indicating an unusually specific immunological interaction that forms the basis of the disease mechanism .58. Immune recognition targets gluten peptides that have been deamidated-a process catalyzed by tissue transglutaminase, in which glutamine residues are converted to glutamic acid. This modification enhances the peptides' affinity for the binding grooves of HLA-DQ2 and HLA-DQ8 molecules, which preferentially bind negatively charged residues.

Located subjacent to the intestinal epithelium, tTG catalyzes not only deamidation but also can couple gluten peptides to lysine residues in juxtaposed proteins, including itself. The creation of these tTG-gluten complexes allows gluten-reactive T cells to stimulate tTG-specific B cells, resulting in the generation of anti-tTG autoantibodies. This mechanism explains the gluten-dependent development of anti-tTG antibodies a hallmark of celiac disease serology [31].

2.7.2 CTLA4 and CD28 Genes

Among non-HLA genes, the cytotoxic T-lymphocyte-associated antigen 4 gene on chromosome 2q33 exhibits the most consistent association with CD. This gene contains a single nucleotide polymorphism (SNP) at position +49 (A/G) in exon 1, with the A allele reported to be more prevalent among CD patients in French and Scandinavian populations. Additional studies in Finnish and Swedish/Norwegian cohorts found supportive linkage near CTLA4, although research on Italian and Tunisian patients yielded conflicting results.

CTLA4 plays a critical immunoregulatory role by inhibiting overactive T-cell responses. Variants carrying the G allele of the +49 SNP have been associated with reduced suppression of T-cell proliferation, which may contribute to the autoimmune activation observed in CD. It is worth noting that the CD28 and ICOS genes key regulators of T and B cell activation are located in close proximity to CTLA4, suggesting a possible combined genetic effect within this chromosomal region [24].

2.7.3 Key Non-HLA Genes and Their Functional Roles

Although H LA-DQ2 and HLA-DQ8 are the primary genetic determinants of celiac disease (CD), they explain only part of the heritability. Non-HLA loci have been identified as critical modifiers of disease susceptibility, progression, and variability in clinical presentation. Large-scale genome-wide association studies have revealed multiple non-HLA genes that regulate immune function and intestinal homeostasis. Each genetic locus contributes a small risk individually, with odds ratios

generally between 1.1 and 1.5. Nonetheless, the overall effect of having multiple risk alleles is substantial; people with thirteen or more non-HLA risk variants face about a sixfold higher risk of developing CD. When these non-HLA variants are combined with HLA genotyping, it improves the identification of individuals at high risk, especially among those who are HLA-DQ2-positive facilitating early monitoring and targeted preventive measures [59].

Among the key non-HLA genes, the IL2/IL21 locus plays a crucial role in controlling T-cell growth and B-cell maturation while SH2B3 (LNK) acts as a negative modulator of cytokine signaling, reducing the activation threshold for T cells and thereby exacerbating immune-driven mucosal damage. The CCR gene cluster, which includes CCR1, CCR2 and CCR3, is responsible for directing chemokine-driven migration of immune cells, leading to inflammatory infiltration in the intestinal lining and worsening villous atrophy.

Other genes such as TAGAP, IL18R1, RGS21, PLEK, CCR9, CTLA4 and LPP also influence immune responses. TAGAP is linked to Rho GTPase signaling and T-cell activation, whereas IL18R1 is involved in the production of interferon-gamma, connecting it to mucosal inflammation. Importantly, CTLA4 and LPP are related to both early celiac autoimmunity and clinically diagnosed celiac disease, while RGS21, IL18R1, PLEK, CCR9 and TAGAP have stronger ties to established disease [60] [61].

2.8 Evidence from Prospective Studies and Gene-Environment Interactions

Prospective cohort studies, notably the TEDDY study, have shed light on the timing of non-HLA genes' impact on disease development. In these cohorts, the hazard ratios for individual non-HLA single nucleotide polymorphisms tend to be higher than the odds ratios found in retrospective case-control studies, highlighting their significance in the early stages of autoimmunity and disease progression. Furthermore, disparities in the incidence of CD in different nations, for instance,

higher rates in Swedish children than in other participants in TEDDY, suggest that environmental factors play an important part in shaping the manifestation of HLA and non-HLA risk gene variants [62].

Various non-HLA loci specific to certain regions have also been identified, including chromosome 8q21.1, which is proximate to the gene for protein kinase A, insulin, and chromosome 10p15, which comprises PFKFB3 and PRKCQ, regions that show a very strong linkage to CD within Swedish populations.

The role of non-HLA regions in celiac disease can also be explained, as they are supposed to influence intracellular signaling as well as T-cell activation, indicating that certain environmental factors within regions could have a role in modulating disease through non-HLA genes.

The findings from prospective studies also establish that non-HLA regions, aside from modulating disease susceptibility, also modulate disease expression since early events for seroconversion, as tested by anti-tissue transglutaminase antibodies, have been identified [63].

2.9 Implications for Risk Prediction and Autoimmune Overlap

Other genetic factors outside of the HLA genes also have an important role in susceptibility to CD. While each factor has a limited role, together these factors substantially aid in the predictive potential of susceptibility with the inclusion of HLA genotyping. Most of these genetic factors are also linked with other autoimmune diseases, including type-1 diabetes, rheumatoid arthritis, etc.

This suggests a common path mechanism in T-cell modulation and innate immunity.. By incorporating non-HLA genes into risk models, it becomes possible to identify at-risk individuals early, allowing for preventive actions such as serologic monitoring and early dietary changes.

The ongoing identification of additional non-HLA loci will improve genetic risk prediction, deepen the understanding of molecular mechanisms and inform early detection strategies for CD and potentially other autoimmune diseases [64].

2.10 Environmental Factors

Gluten is the environmental factor most closely linked to celiac disease . Patients who completely eliminate gluten from their diet generally achieve full clinical remission, whereas reintroducing gluten results in the return of symptoms. These findings clearly identify gluten as the main environmental trigger for the onset and progression of the disease.

Nonetheless, the possibility of other environmental factors contributing to the disease remains an area of active research. Among the potential factors being studied are intestinal infections, with adenovirus 12 once considered a candidate due to its partial amino acid sequence resemblance to α-gliadin. Although this theory initially seemed credible, the current understanding of CD immunopathogenesis provides limited evidence for a causal link, and existing epidemiological data are still inconclusive [65].

Celiac disease has been reported worldwide including Europe, North and South America, Asia, Africa, and Oceania though prevalence rates vary considerably between regions. The disorder appears relatively uncommon in sub-Saharan Africa and Southeast Asia, but recent studies from China suggest it may be underdiagnosed rather than rare. A meta-analysis estimated the global seroprevalence of CeD at 1.4%, with a biopsy-confirmed prevalence of 0.7%. Data from the U.S. National Health and Nutrition Examination Survey (NHANES) revealed a similar seroprevalence of 0.7%, emphasizing that a large number of cases remain undiagnosed in the general population [29]. For this reason, active case finding is considered a better approach than passive diagnosis. The prevalence of CeD depends on several factors, including sex, age, geographic distribution, and genetic predisposition. The two main determinants of disease frequency across populations

are the presence of HLA risk haplotypes and the per-capita wheat consumption. The disease is slightly more common in females, and familial clustering is frequent about 10% of first-degree relatives of affected individuals are also diagnosed with CeD. Concordance studies demonstrate that monozygotic twins show an 80% concordance rate, compared with about 30% in HLA-identical siblings and 10% in dizygotic twins, confirming the combined influence of genetic and environmental factors in CeD development [54] [3].

Regions where wheat and gluten intake is high tend to show greater CeD prevalence, underscoring the dietary component of the disease. Nevertheless, gluten exposure alone cannot explain why some genetically predisposed individuals never develop CeD or why symptoms can emerge later in life despite decades of gluten consumption. Significant differences between genetically similar populations with comparable wheat intake such as those in Finland and Russian Karelia suggest that non-genetic environmental variables also play a decisive role. Various observational studies have suggested that several additional factors, such as infant feeding habits, viral infections, antibiotic use, and delivery method, may play a role, although the evidence is varied and sometimes contradictory. The way infants are fed could potentially affect their risk of developing diseases [32].

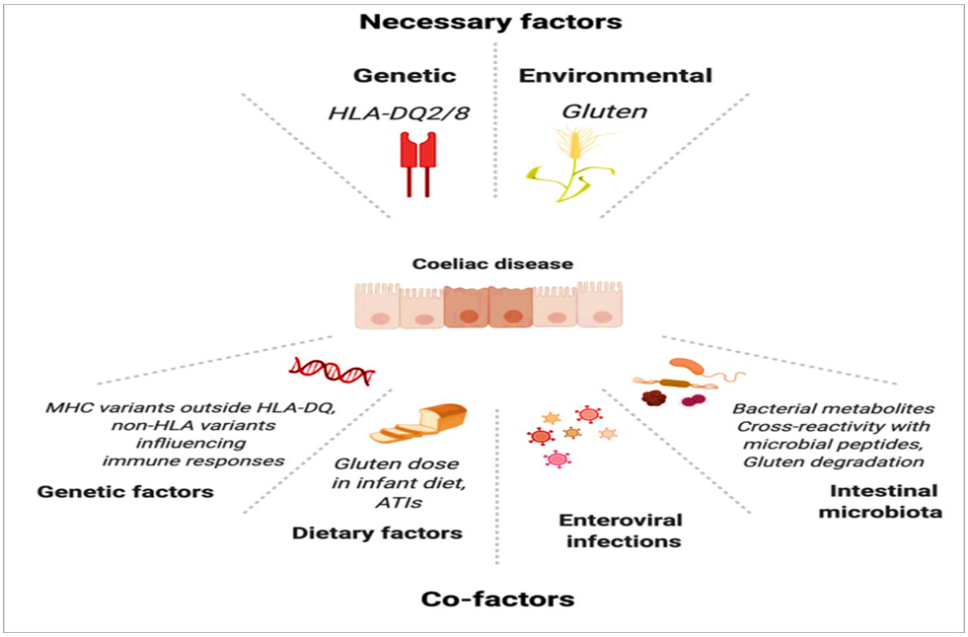


FIGURE 2.2: Genetic and environmental factors in coeliac disease [66]

Factors such as the length of breastfeeding, the timing of introducing gluten, and early exposure to cow's milk proteins are thought to influence immune tolerance. Notably, rotavirus infection has been linked to an increased risk of developing CeD autoimmunity, with prospective studies indicating that a higher frequency of rotavirus infections is associated with a greater risk of the disease. Gluten is a complex protein mixture, primarily composed of prolamins (gliadins in wheat, secalins in rye and hordeins in barley and glutenins. Research suggests that consuming as little as 50 mg of gluten daily can cause noticeable damage to the small-intestinal lining in susceptible individuals. Additionally, changes in the gut microbiota such as decreased Bifidobacteria and increased Gram-negative bacteria have been observed in CeD patients implying that microbial imbalance might play a role in disease manifestation. Other potential environmental factors include exposure to heavy metals and bacterial transglutaminase found in processed foods although their precise roles are still being studied [21].

2.11 Immunological and Non-Genetic Factors

The immune system plays role in onset of celiac disease by triggering an abnormal inflammatory reaction to gluten. When deamidated gluten peptides are presented by HLA-DQ2 or HLA-DQ8 molecules, $CD4^+$ T helper cells in the lamina propria are activated and release pro-inflammatory cytokines interferon-gamma (IFN-γ) and interleukin-15 (IL-15).

These cytokines stimulate cytotoxic $CD8^+$ intraepithelial lymphocytes leading to villous atrophy and damage to mucosa. The enzyme tissue transglutaminase acts both as an autoantigen and a modulator of gluten peptides, which is the basis of the autoimmune pathophysiology of the condition. In a susceptible individual, the immune reaction to gluten can cross-react with tTG, leading to the production of anti-tTG antibodies, which is a diagnostic feature of celiac disease [14].

This autoimmune reaction can continue without the presence of gluten, leading to persistent inflammation and damage to the lining of the intestine. An impaired

or "leaky" gut, which is the compromised integrity of the intestinal barrier, allows the passage of gluten peptides and bacterial antigens into the lamina propria of the intestine, exaggerating the reaction. The increased production of zonulin, a modulator of the permeability of the tight junctions, has been implicated in the pathophysiology of celiac disease, suggesting that the impairment and integrity of the epidermis layer is both a consequence and a cause of the condition [24].

Environmental factors such as infections, food components, and other external stimuli can result in epigenetics that provide a plausible explanation of why some individuals who are genetically predisposed may develop the condition. It can therefore be argued that the condition of celiac disease is a reflection[50].

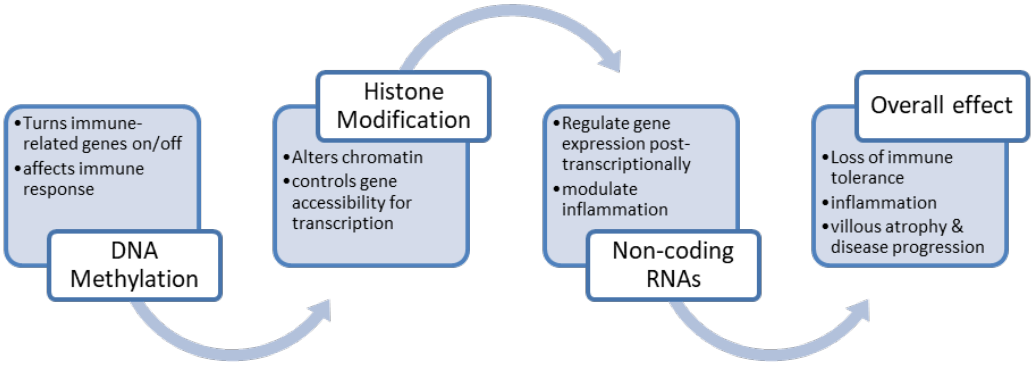


FIGURE 2.3: Epigenetic Modifications in CD [17]

2.11.1 Adaptive Immune Response to Gluten

Tissue transglutaminase plays role in triggering adaptive immune response in celiac disease by deamidating gliadin peptides derived from gluten. This process transforms glutamine residues into glutamic acid, introducing negative charges. Additionally, TG2 is the main autoantigen targeted by disease-specific antibodies, making it a significant diagnostic marker for celiac disease. Under normal conditions, TG2 is found in subepithelial and connective tissues, where it aids in apoptosis, wound healing, and the organization of the extracellular matrix. TG2 assists in the transcytosis of gliadin peptides across the intestinal epithelium through interactions with CD71 and secretory IgA (sIgA) and inhibiting TG2 prevents the uptake of immunogenic peptides like p31-49 [67]. TG2-mediated

deamidation occurs preferentially in slightly acidic environments (pH ¡ 7.0), especially when glutamine donors outnumber lysine acceptors. Since gliadin is rich in glutamine, it is highly prone to TG2 modification, which enhances peptide binding to HLA-DQ2 pockets (P4, P6, P7) and increases immunogenicity, allowing multiple immunogenic epitopes to activate distinct pathogenic pathways [30].

This alteration results in enhanced T-cell activation and the release of inflammatory cytokines. Furthermore, TG2 has the ability to cross-link gluten peptides with self-proteins including itself, creating complexes that trigger the production of anti-TG2 autoantibodies, a characteristic feature of CD.

Although deamidation is not absolutely necessary to start immune activation, it greatly exacerbates inflammation once the process is underway. The activation of TG2 can also be heightened by factors such as infections, damage to the epithelium or existing mucosal inflammation which may disrupt immune tolerance and encourage the continuation of autoimmune processes in celiac disease [68].

2.11.1.1 T-Cell Response to Gluten

In celiac disease, the adaptive immune response is mainly directed by T lymphocytes in the lamina propria which depend on antigen-presenting cells that display HLA-DQ2 or HLA-DQ8 molecules for antigen presentation. In a healthy duodenum, APCs are mainly macrophages (≈80%, $CD163^+$ $CD11c^-$) and tolerogenic dendritic cells (DCs, $CD103^+$ $CD11c^+$).

In contrast, celiac disease involves the recruitment of peripheral monocytes that differentiate into proinflammatory DCs ($CD14^+$ $CD11c^+$), accompanied by a decline in tolerogenic macrophages and DCs. Increased levels of interferon-alpha (IFN-α) within the intestinal mucosa promote a pro-inflammatory environment. This association is further supported by reports of celiac disease emerging in hepatitis C patients undergoing IFN-α therapy, as well as by its higher occurrence in individuals with Down syndrome, in whom the IFN-α receptor gene is located on chromosome 21.

The HLA-DQ2 and HLA-DQ8 molecules present gliadin peptides and TG2–gluten complexes to $CD4^+$ T cells in the lamina propria, leading to activation of a Th1-type immune response characterized by increased production of IFN-γ, TNF-α, IL-18, and IL-21, along with a concurrent decrease in the regulatory cytokines IL-10 and TGF-β. This imbalance in cytokines initiates the breakdown of the matrix through metalloproteinases, resulting in damage to the mucosa [62]. Under normal circumstances, TGF-β1 plays a crucial role in preserving epithelial integrity, but its decrease in celiac tissue is associated with crypt damage and the infiltration of immune cells. IFN-α promotes the differentiation of Th1 cells and maintains inflammation by inhibiting T-cell apoptosis and increasing the expression of co-stimulatory molecules on APCs. Additionally, IL-18 which is produced by macrophages, DCs and epithelial cells works in conjunction with IL-12 and IFN-α to boost IFN-γ production. The mature, active form of IL-18 is significantly elevated in celiac lesions, likely due to its proteolytic activation by IL-1β–converting enzyme [69].

2.11.1.2 B-Cell Response to Gluten

Celiac disease is characterized by the generation of various autoantibodies that target both self and foreign antigens. In 1997, tissue transglutaminase was discovered to be the main self-antigen targeted by anti-endomysial antibodies. In the duodenum's lamina propria, plasma cells produce anti-TG2 IgA antibodies, with concentrations two to three times higher during active disease than in remission. Importantly, TG2-specific IgA deposits are detectable at all disease stages, even before clinical symptoms or visible mucosal damage appear. Under normal conditions, the intestinal mucosa has relatively few naïve or memory B cells, with most cells being plasma cells or plasmablasts that express low levels of HLA class II molecules.

In celiac disease, B cells likely function as antigen-presenting cells in mesenteric lymph nodes, thereby boosting gluten-specific T-cell responses. Although TG2-specific T cells have not been definitively identified, gluten-reactive $CD4^+$ T cells

can promote the differentiation of B cells into plasma cells that produce anti-TG2 IgA and IgG antibodies, both of which decrease when following a gluten-free diet [14]. One suggested mechanism suggests that B cells take in TG2 gluten complexes and present them through HLA-DQ molecules to gluten-specific T cells, which then provide the co-stimulatory signals required for antibody production.

Anti-TG2 antibodies might further worsen intestinal inflammation by increasing gluten uptake and activating Fc receptors on local granulocytes. In addition to TG2, autoantibodies have been found against actin, various collagen isoforms other members of the transglutaminase family (TG3, TG6) and clotting factor XIII. The presence of IgA-TG3 complexes in the skin is linked to dermatitis herpetiformis while antibodies targeting neuronal TG6 are associated with gluten ataxia. Together, these findings underscore the role of antibody-mediated mechanisms in driving the extraintestinal manifestations of celiac disease [70].

2.11.2 Innate Immune Response to Gluten

In addition to adaptive mechanisms certain gliadin peptides have the ability to directly stimulate innate immune responses in intestinal epithelial cells and dendritic cells (DCs). These peptides, such as α-gliadin fragments p31-43 and p31-49 are not detected by gluten-specific $CD4^+$ T cells. Instead, they induce stress responses in epithelial cells and modify intracellular signaling pathways that remain partially understood.

Ex vivo research using intestinal biopsies from individuals with celiac disease has shown an increase in the expression of IL-15, cyclooxygenase-2 (COX-2) and activation markers like CD25 and CD83 in lamina propria mononuclear cells after exposure to gliadin. Concurrently, epithelial cells exhibit heightened expression of MHC class I related chain (MIC) molecules which can activate intraepithelial lymphocytes [67].

Some gliadin peptides can imitate the activity of epidermal growth factor by preventing the endocytosis of the EGF receptor (EGFR) which extends receptor

signaling and encourages the proliferation of epithelial cells. Even though EGFR and its signaling pathway are already heightened in celiac patients regardless of gluten consumption, this could account for the epithelial sensitivity and damage observed in the disease.

Furthermore, gliadin peptides might interact with Toll-like receptor 4 on dendritic cells, stabilize non-classical MHC molecules like HLA-E, or engage with the chemokine receptor CXCR3 thereby enhancing gut permeability and disrupting the tight junctions between enterocytes.

Together, these innate processes intensify local inflammation, weaken barrier integrity and sustain the intestinal damage typical of celiac disease [67] [30].

2.11.2.1 Role of the Intraepithelial Lymphocytes

Intraepithelial lymphocytes (IELs) are a diverse group of immune cells located in the basolateral area of intestinal epithelial cells, with their quantity and distribution differing across various sections of the intestine.

They are generally categorized into natural IELs (TCR $\alpha\beta$ and TCR $\gamma\delta$) and induced IELs (TCR $\alpha\beta$ $CD4^+$ and TCR $\alpha\beta$ $CD8\alpha\beta^+$), playing crucial roles in defending the host, promoting immune tolerance and preserving the integrity of the epithelium [71].

In celiac disease intraepithelial lymphocytes play a role in damaging the mucosa with a rise in TCR $\alpha\beta^+$ T cells being linked to villous atrophy. Additionally, IELs undergo a shift towards cytotoxicity, and the persistence of high levels of TCR $\gamma\delta^+$ IELs even on a gluten-free diet is a characteristic feature of CD. Inflammatory conditions cause natural IELs to become activated locally by proinflammatory molecules, leading them to develop autoreactive properties.

Cytotoxic IELs identify epithelial cells by using NKG2D and CD94 receptors to bind to MICA and HLA-E ligands on intestinal epithelial cells resulting in intraepithelial lymphocytosis, epithelial damage, villous atrophy and crypt hyperplasia [72].

Other immune cell populations implicated in CD include natural killer and natural killer T (NKT) cells. NK cells are primarily involved in eliminating virus-infected and tumor cells, independent of MHC molecules and antibodies. Studies have reported a marked decrease in natural killer (NK) cell numbers in patients with active celiac disease compared to individuals adhering to a gluten-free diet or healthy controls. NKT cells, on the other hand, represent a mixed population that expresses both T-cell (TCR, CD3) and immunoglobulin receptors.

Some subsets express a semi-invariant TCR (such as TCR Vα24β11) and can be activated via TCR signaling, independently of MHC, to stimulate IL-10 secretion by epithelial cells. However, the precise role of NKT cells in CD remains unclear, as these cells can produce a variety of cytokines, including both proinflammatory and regulatory types [73].

2.11.2.2 Role of Interleukin IL-15 and IL-21

Interleukin-15 (IL-15) is a key cytokine mediating the gluten-driven innate immune response in the intestinal mucosa. It binds to its high-affinity receptor, IL-15Rα, which is structurally related to the IL-2 receptor. The interaction between IL-15 and IL-15Rα occurs before the receptor is expressed on the cell surface and is essential for the cytokine's biological function. In celiac disease, IL-15 is overexpressed by intestinal epithelial cells in response to gluten exposure, as well as by macrophages, dendritic cells (DCs), and other mononuclear cells in the lamina propria. This overexpression of IL-15 contributes to IEL reprogramming, upregulation of MICA stress ligands on enterocytes, and activation of DCs. It also encourages the production of IL-21 another cytokine that plays a crucial role in the development of CD. Gliadin peptides can trigger the release of IL-15 from the intestinal lining in both individuals with and without celiac disease however, only those with celiac disease show increased levels of IL-15Rα expression which might reduce their activation threshold and increase their sensitivity to IL-15 [73].

Research in genetics has identified a link between the IL2/IL21 gene region and the risk of developing celiac disease, highlighting the essential role of IL-21 in both

the onset and continuation of the disease. Elevated IL-21 expression has been observed in intestinal biopsies from patients with active CD, with its production localized to lymphocytes in both the epithelium and lamina propria. IL-21 frequently coexists with IFN-γ and can also be secreted by natural killer T (NKT) cells. Induced by IL-15, IL-21 contributes to the maintenance of mucosal inflammation by disrupting immune regulatory mechanisms. Although IL-21 is primarily produced by Th17 cells, other Th17-associated cytokines generally do not show significant upregulation in most CD patients, except in a subset of adults. IL-15 and IL-21 can act synergistically through multiple signaling pathways to enhance the resistance of $CD4^+$ T cells to suppression by regulatory T cells (Tregs) within the intestinal mucosa.IL-15 interferes with anti-inflammatory signaling pathways, including TGF-β1/Smad3 and PI3K, thereby sustaining inflammation, while the exact mechanisms of IL-21 remain under investigation [73] [74].

Moreover, IL-15 has been implicated in the pathogenesis of refractory celiac disease (RCD) and enteropathy-associated T-cell lymphoma (EATL). By promoting the proliferation and survival of cytotoxic intraepithelial lymphocytes (IELs), IL-15 contributes to their resistance to apoptosis and facilitates progression toward lymphomagenesis [74].

2.12 Gut Microbiota

There is a general awareness that an imbalance in gut microbiota is a very important factor influencing the intestinal environment in celiac disease patients. Ongoing imbalance has also been observed in the intestinal environment of CD patients in whom a gluten-free diet is practiced meticulously, which provides a clear indication that intestinal imbalance in microbes may be associated with chronic intestinal symptoms in CD patients [33]. The altered microbial environment in CD is very significant in the regulation of immunity across these patients. Microbes in CD influence immune reactions, inflammation, or allergic reactions in this manner. Some microbes were observed to ensure increased immune activation. Moreover, probiotics possess protective and anti-inflammatory properties.

Microbiota in the intestinal region prime/introduce intestinal immune cells mostly to Toll-like receptors (TLRs), which results in the release of antimicrobial peptides or chemokines that modulate local immunity in these regions. Some deleterious microbes like P. aeruginosa were observed to enhance immunogenicity of gluten by processing its peptides with enzymes. However, probiotics in CD work to counter this mechanism by degrading immunogenic peptides of gluten and preventing T-cell activation in immunity response. This clearly demonstrates that gut microbiota has very complex functions in modulating gluten's immunogenicity in intestinal[70].

After the gut microbiota is established shortly after birth, its composition and stability are influenced by various external influences. Studies have shown that the timing of gluten introduction during infancy represents a critical turning point in determining the risk of CD. The ESPGHAN recommends avoiding too-early or too-late gluten introduction before 4 and after 7 months of age during breastfeeding. Instead, this should take place during a window period moderately and gradually to promote mucosal immune tolerance. Other factors besides the timing of gluten exposure that have been claimed to influence microbiota development are the type of infant feeding, delivery method, early infections, and antibiotic use. Microbial imbalances have been consistently noticed in the oral cavity, intestinal tract, and feces in several studies conducted on CD patients. Interestingly, though a GFD is essential for the treatment of CD, most studies report that the complete restoration of gut microbiota after the treatment rarely occurs. According to Marasco et al. (2016), unintentionally, a GFD might reduce the intake of polysaccharides, which are the main energy source for gut microbes, hence keeping dysbiosis. This suggests that without dietary gluten, gut microbiota composition in CD patients may remain altered in the long term[76].

2.13 Signs and Symptoms

Celiac disease is often missed when the symptoms of the disease are nonspecific. Earlier, the disease was known for malabsorption problems in children. More

recently, though, a pattern of more subtle or atypical symptoms presenting in multiple systems in people of all ages has been recognized. While symptoms of the gastrointestinal type, including diarrhea, abdominal pain, bloating, and gas, have not decreased in frequency, most people experience only mild symptoms or no symptoms at all. Advances in diagnostic tools and greater familiarity with the disease on the part of physicians have made it possible to recognize the conditions of a more nonspecific presentation [54].

Many symptoms of CD extend beyond the intestines, illustrating the wide-ranging immunological effects of the disease. Dermatitis herpetiformis is a well-recognized extraintestinal symptom that affects up to 10% of adult patients and is characterized by intensely pruritic vesicular rashes located on the elbows, knees, buttocks, and scalp among other areas [55]. Other extraintestinal symptoms include arthralgias or joint inflammation, neurological problems such as peripheral neuropathy, and hematological problems such as anemia, all of which make the diagnosis more difficult. Due to the broad range of clinical presentations and lack of familiarity among health care providers, these conditions have been known to be diagnosed as long as ten years later in resource-rich nations, and longer when resources are scarcer [3]. As such, it remains critical to identify both common and uncommon presentations of CD as early as possible. In addition, asymptomatic CD may be uncovered through targeted screenings among at-risk populations such as first-degree relatives of those with the disease and other auto-immune conditions like type 1 diabetes mellitus. In this regard, it becomes important that asymptomatic disease be found in such populations for the prevention of long-term illness and improvement in overall quality of life [77] [78].

2.14 Diagnosis

2.14.1 Histopathological Analysis

The mucosal damage in celiac disease typically takes years to develop; indeed, pronounced villous atrophy is observed in most cases. The classical histological

changes have long been categorized using grouped systems, but these systems have demonstrated poor reproducibility, even in cases that are defined as borderlines by these criteria. Indeed, an accurate assessment can be achieved by sophisticated morphometric analysis; this assessment measures villus height to crypt depth ratios and numbers of intraepithelial lymphocytes in each case individually [37].

2.14.2 Intestinal Biopsy

CD can only be properly diagnosed by histological proof of involvement, demonstrating by biopsy the features of intraepithelial lymphocytosis, T cells, and plasma cells infiltrating the lamina propria, villous atrophy, and crypt hyperplasia. The classification described by Marsh et al. has widely been adopted to grade these lesions. Though previous criteria required at least partial villous atrophy (Marsh 3A), it has recently become accepted that lesions of Marsh 1 indicative of active disease could correctly be identified when assessed in the context of clinical and serological criteria [54] [3] [34].

Biopsy of the duodenum, usually carried out by gastrodudenoscopy, represents the gold standard. However, sampling errors may occur due to the patchy lesions; therefore, it is essential to obtain at least four properly oriented biopsy samples. Biopsy interpretation represents an important aspect, given that similar villous lesions may be caused by other underlying processes, such as infections, drugs, or autoimmune enteropathy.

Though immunohistological studies of $\gamma\delta$ T-cell expression may add to enhanced sensitivity, it is rarely done. Villous atrophy signifies the advanced stage of mucosal damage, but the pathogenesis of CD often starts years before these morphological changes are evident. Patients with positive CD-specific antibodies but normal histology are considered to have potential CD, and their management, whether through observation or a gluten-free diet, depends on the severity of symptoms and clinical context. Since lesions can be localized, especially in the duodenal bulb the biopsy site and specimen quality are vital. Misinterpretation can occur due to poor orientation or the inclusion of Brunner's glands. Therefore, obtaining

high-quality, well-processed samples is crucial for dependable histopathological assessment [79].

2.14.3 Serological and Other Biomarker Candidates

Serological testing is the initial step in diagnosing CD. Initially, antibody tests focused on reticulin (ARA), endomysium (EmA), and gliadin (AGA) using indirect immunofluorescence, which was labor-intensive. After tissue transglutaminase 2 (TG2) was identified as the main autoantigen, more standardized and objective serological tests were created. These TG2-based tests are now the preferred diagnostic method, though their specificity can be lower than EmA tests, especially in those with other autoimmune diseases. Research indicates that serum antibodies in CD patients recognize specific conformational TG2 epitopes not targeted by antibodies from people with other autoimmune conditions. Another significant advancement in CD diagnosis is detecting antibodies against deamidated gliadin peptides (DGP). However, its accuracy may be affected in individuals with selective IgA deficiency. To address this, newer POC tests measuring combined IgA and IgG antibodies to DGP have been developed and are recommended for professional use [80] [81].

Recent research has uncovered new potential biomarkers. One such candidate is intestinal fatty acid-binding protein (I-FABP), which is quickly released into the bloodstream after mucosal damage and can signal harm to the intestinal lining. However, increased levels of I-FABP may also be present in other inflammatory bowel conditions like Crohn's disease. Another promising marker is related to cytochrome P450 3A4 (CYP3A4) activity, an enzyme predominantly found in villous epithelial cells but significantly diminished in cases of villous atrophy. As a result, individuals with active CD might show elevated systemic levels of orally taken simvastatin due to decreased metabolism by CYP3A4. Additionally, the combination of gut microbiota profiling and microRNA (miRNA) signatures is being investigated as future biomarkers, though research is ongoing to pinpoint the specific miRNAs involved. Currently, the combination of serological testing and small

intestinal biopsy is the diagnostic gold standard for CD. Among serological tests, EmA and TG2 antibody (TG2-Ab) assays offer the highest sensitivity (90–100%) and nearly perfect specificity. Although EmA testing is the reference method, it is labor-intensive and subjective, whereas enzyme-linked immunosorbent assay (ELISA) and radiobinding assays for TG2-Ab provide automated, reproducible alternatives. However, the performance of these assays can vary based on the conformation and quality of the TG2 antigen used [82].

First-generation anti-gliadin antibody tests are no longer advised due to their lack of accuracy. Contemporary tests that focus on DGP-specific antibodies can identify CD cases that TG2 or EmA assays might miss, although they are not yet commonly used in clinical practice. For patients with selective IgA deficiency, IgG-based antibody tests are the preferred option. It is crucial to recognize that around 10% of CD patients do not show serological markers, making a biopsy-based diagnosis necessary. In these instances, histological improvement after adopting a gluten-free diet confirms the diagnosis. Rapid POC tests for detecting anti-DGP and TG2-Abs have shown promise for use in the field, especially in areas with limited laboratory infrastructure, but further validation studies are needed before they can be widely adopted [83].

2.14.4 Antibody Detection

Detecting antibodies through serological methods remains crucial for diagnosing CD. Patients with untreated CD often show increased levels of IgA and IgG antibodies against gluten and TG2, which decrease once a gluten-free diet is started. In the past, the endomysial antibody (EmA) test was conducted using immunofluorescence but it has mostly been replaced by ELISA-based assays that target recombinant TG2, providing enhanced accuracy and automation. IgA antibodies are more specific than IgG, so IgG tests are mainly beneficial for patients with IgA deficiency, a condition more common in those with CD. Some individuals who are HLA-DQ2 or HLA-DQ8 positive produce anti-TG2 antibodies and show gluten-related symptoms even without clear mucosal changes leading to discussions about

whether biopsy should continue to be the definitive diagnostic method. TG2 directly contributes to CD pathogenesis by deamidating gluten peptides, increasing their binding affinity to HLA-DQ2/DQ8 molecules and enhancing T-cell recognition. Deamidated gliadin peptide (DGP) antibody tests demonstrate similar specificity and sensitivity to TG2 assays, and combining both slightly improves detection rates. Innovative diagnostic techniques include (1) rapid blood-based POC TG2 tests, (2) detection of mucosal IgA-TG2 deposits, and (3) ex vivo gluten challenge assays that measure IgA-TG2 secretion by intestinal B cells though the latter are not yet widely adopted in clinical practice [84].

2.14.5 Detection of Gluten-Specific T Cells After Gluten Challenge

Diagnosing CD in patients already on a gluten-free diet can be difficult since serology and histology normalize upon dietary restriction. Traditionally, diagnosis requires a prolonged gluten challenge, but recent studies suggest that a short-term (3-day) gluten challenge can elicit detectable gluten-specific T cells in peripheral blood, identifiable using ELISPOT or HLA-DQ2–gliadin tetramer staining. These methods offer promise for future diagnostic use but require further clinical validation [82].

2.14.6 HLA Typing

More than 95% of CD patients carry HLA-DQ2.5 (DQA105 / DQB102), while most others possess HLA-DQ8. A minority may express variants such as DQ2.2 (DQA10201 / DQB102) or DQ7 (DQA105 / DQB10301). The absence of DQB1*02 effectively rules out CD, making HLA typing a valuable exclusion tool. However, since HLA-DQ2 and DQ8 alleles are also common in the general population, their presence alone does not confirm disease but rather indicates genetic susceptibility [85].

2.15 Additional Diagnostic Tools

In some diagnostically uncertain situations such as patients who are seronegative or exhibit mild villous abnormalities non-traditional investigative methods are often required for accurate identification of celiac disease.

HLA typing serves as an important exclusion test, since individuals who do not carry HLA-DQ2 or HLA-DQ8 alleles have an extremely low likelihood of developing the disease.

Evaluating the inflammatory cell population within small intestinal tissue can also support diagnosis. Although an increase in $CD3^+$ intraepithelial lymphocytes (IELs) alone is not specific to celiac disease, quantifying their distribution particularly the presence of $\gamma\delta^+$ IELs concentrated at villous tips can provide valuable diagnostic insight in borderline cases [86] [79] [3].

Detection of tissue transglutaminase-2 specific IgA deposits in frozen intestinal biopsy samples further strengthens diagnostic accuracy, especially in ambiguous presentations. Moreover, detecting circulating gluten-reactive T cells presents a possible diagnostic method for those who have already decreased their gluten consumption.

A brief three-day gluten challenge can activate these memory T cells, which can then be identified using IFN-γ ELISPOT assays.

Although this test demonstrates high specificity its sensitivity is still limited. Alternatively, flow cytometric detection with HLA-DQ gluten tetramers enables the identification of gluten-specific T cells with high accuracy, regardless of the patient's adherence to a gluten-free diet.

At present, HLA typing, immunohistochemical evaluation of IEL subsets, and occasionally the detection of intestinal IgA deposits remain the primary supplementary tools used outside of research environments [84].

2.16 Non-Celiac Gluten Sensitivity

The symptoms of celiac disease are frequently vague and can resemble those of conditions like irritable bowel syndrome or cereal allergies. Notably, many people find relief from symptoms by removing wheat from their diet, even if they haven't been diagnosed with celiac disease. Controlled studies have shown that some of these individuals have non-celiac gluten sensitivity where gluten consumption causes symptoms without autoimmune or allergic reactions. According to epidemiological studies, non-celiac gluten sensitivity affects approximately 2-5% of people making it more common than celiac disease.

At present, there is no specific biomarker for NCGS so diagnosing it requires a comprehensive exclusion of celiac disease and wheat allergy. Typically, patients exhibit normal intestinal histology and test negative for antibodies specific to celiac disease. Diagnosis is confirmed through a double-blind gluten challenge to verify that symptoms are indeed related to gluten consumption a procedure that is both time-consuming and complex. Recent studies also propose that other wheat components, like fructans, might contribute to symptom development suggesting that gluten might not be the sole cause [87].

2.17 Treatment

2.17.1 Gluten-Free Diet and Nutritional Management

The only established treatment for celiac disease is a lifelong adherence to a strict gluten-free diet, which requires the complete elimination of wheat, barley, rye and their derivatives [50]. Most people experience improvement in a few weeks to months after beginning the diet. The effectiveness of dietotherapy largely depends on education and ongoing dietary counseling. Patients with celiac disease are encouraged to avoid foods that contain gluten, such as bread and pasta, or processed meats and even certain medications or supplements. The problem of cross-contamination in common kitchens can become a serious issue and may arise due

to the use of common cooking utensils, toasters, or contaminated cutting boards [87] [88]. Patients with celiac disease may experience nutritional deficiencies due to malabsorption for an extended period after being newly diagnosed with celiac disease. The common deficiencies include iron, folic acid, vitamin B12, vitamin D, calcium, zinc, and fat-soluble vitamins such as A, E, and K. It is important to note that there is a high need for nutritional assessment and supplements during the early treatment phases. Weight and blood level monitoring is crucial for long-term health stability. For children with celiac disease, a gluten-free diet can aid in growth and development, and for those with this condition who mature into adults with this disease, the diet can decrease the potential complications thereof, which include osteoporosis, infertility, or intestinal cancer. Therefore, in patients with celiac disease, a gluten-free diet acts both therapeutically and preventively.

2.17.2 Monitoring and Refractory Celiac Disease

Long-term management of celiac disease extends beyond merely eliminating gluten from one's diet it also necessitates regular follow-up assessments to evaluate disease control and patient adherence. Clinical evaluations, along with serological tests like anti-tissue transglutaminase and anti-endomysial antibodies are useful for assessing dietary compliance and mucosal healing. Typically, antibody levels decline after several months of strict adherence, but persistently elevated antibodies or recurring symptoms may suggest ongoing gluten exposure or the development of complications [91] [53]. Even when patients strictly adhere to a gluten-free diet some may still experience symptoms. In such cases, it is crucial to explore other possible causes, such as lactose intolerance, small intestinal bacterial overgrowth, microscopic colitis or irritable bowel syndrome. A small subset of patients may develop refractory celiac disease a severe condition characterized by persistent villous atrophy and malabsorption despite following the diet for at least 12 months. RCD is divided into Type I and Type II based on the immunophenotype of intraepithelial lymphocytes. Type I RCD features a normal IEL phenotype while Type II involves clonal or abnormal IELs and has a higher risk of progressing to

enteropathy-associated T-cell lymphoma a rare but severe complication. Managing RCD is intricate and necessitates expert care, frequently involving the use of corticosteroids such as budesonide or prednisone to alleviate inflammation and relieve symptoms. In more persistent cases, immunosuppressive drugs like azathioprine, cyclosporine, or cladribine may be utilized to control abnormal immune reactions. Supportive therapies, including nutritional supplements, parenteral nutrition, or enteral feeding, might be essential to avert severe malnutrition and promote intestinal recovery. Consistent monitoring through histological evaluations, imaging, and serological markers is vital to identify complications early and maintain ongoing remission [53].

2.17.3 Emerging and Experimental Therapies

While a gluten-free diet is recognized for its considerable health advantages, strictly following it can lead to social and financial challenges. As a result, scientists are investigating alternative or complementary treatments to either improve or eventually replace the gluten-free diet as the main therapy. Recent advancements in understanding the molecular mechanisms of gluten-induced autoimmunity have identified new therapeutic targets at different stages of disease progression. One promising strategy is oral enzyme therapy, which seeks to break down gluten peptides into non-immunogenic fragments before they can initiate an immune response. For instance, ALV003 (latiglutenase), a combination of glutamine-specific cysteine protease (EP-B2) and proline-specific endoprotease, has proven effective in degrading gluten peptides in the stomach. Similarly, a prolyl endoprotease from Aspergillus niger has shown gastric activity that enhances gluten digestion. These enzyme therapies are currently in phase II and III clinical trials and may serve as protective measures against accidental gluten exposure [92]. Another potential therapeutic strategy involves strengthening the intestinal epithelial barrier, which is compromised in celiac disease. Researchers are evaluating the zonulin inhibitor AT1001 (larazotide acetate) for its ability to reduce intestinal permeability, thereby limiting the movement of gluten peptides and subsequent inflammation. Other approaches include tissue transglutaminase (TG2) inhibitors, which prevent

gluten deamidation and its presentation by HLA-DQ2/DQ8 molecules. Moreover, the development of HLA-blocking agents and peptide-based vaccines aims at immune tolerance-by selective desensitization of gluten-reactive T cells.

Researchers are also working on gene modification and plant breeding for wheat varieties that do not express harmful gluten peptides. This would likely provide a long-term preventive approach against the disease; however, the approach is practically burdened by the extensive complexity of gluten genes.

Other investigational treatments include helminth (hookworm) therapy aimed at changing gut immune responses and dampening gluten hypersensitivity.

Finally, biologic treatments, including monoclonal antibodies targeting cytokines like IL-15, are in development for treatment-resistant celiac disease and may be used for uncomplicated disease in the future. Overall, these novel therapeutic approaches constitute a step toward comprehensive management of the disease, offering new possibilities both for patients with poor dietary adherence and for those with severe TRCD [93].

2.17.4 Novel Therapeutic Approaches

To ensure precision in chromosomal positioning, reference and alternate alleles and the related gene or intergenic region each SNP was validated using databases such as dbSNP and Ensembl. The final dataset included only human SNPs with verified genomic information and strong evidence linking them to celiac disease for further analysis.

Additionally peptide-based vaccines designed to induce immune tolerance against dominant gluten epitopes are in development. These vaccines aim to desensitize the immune system reducing its pathogenic response to gluten exposure.Other experimental therapies include controlled hookworm infections hypothesized to shift local immune balance toward tolerance and cytokine-targeted treatments. Among these, anti-IL-15 therapy has shown particular promise for refractory celiac

disease with the potential for broader application in uncomplicated cases if proven safe and effective [95] [96].

2.18 Recent Studies on Celiac Disease using Bioinformatics

Recent population-based and genomic studies confirm that celiac disease (CeD) risk extends beyond classical HLA-DQ2/DQ8 alleles, with polygenic models capturing substantial variation and robustly stratifying risk in both diagnosed and screen-detected cases, implying comparable genetic architectures and a major influence of non-genetic modifiers [97] [98]. Modern polygenic risk score (PRS) workflows built on PLINK-based quality control, GWAS summary statistics, and LD-aware algorithms such as LDpred2, PRS-CS, and SBayesR are increasingly applied to large cohorts to quantify genetic liability and assess cross-population portability [99] [100]. Complementary HLA analyses have progressed from classical imputation frameworks (HIBAG, SNP2HLA) to high-resolution allele calling from whole-exome and whole-genome sequencing, improving precision for diverse populations and enabling dissection of non-HLA immune loci that interact with HLA risk [101] [102].

Concurrently, high-throughput and single-cell technologies coupled with advanced bioinformatic pipelines have refined mechanistic insight into CeD pathogenesis. Single-cell RNA-seq, single-cell TCR-seq, and spatial transcriptomics analyzed using tools such as Seurat, Scanpy, and Cell Ranger now map immune–epithelial niches and disease-specific T-cell clonotypes, while integrative bulk RNA-seq, ATAC-seq, eQTL/TWAS, and proteomic analyses reveal regulatory effects of non-HLA loci [103]. Microbiome studies using 16S and shotgun metagenomics, functional profiling, and metabolomics continue to uncover microbial contributions to CeD expression, immune modulation, and variable response to a gluten-free diet [104]. Machine-learning and deep-learning models including random forests, gradient boosting, and convolutional neural networks are increasingly applied for early

prediction, PRS refinement, and automated histopathological diagnosis, while network and pathway analyses enable prioritization of therapeutic targets and interpretation of multi-omic data.

Although numerous studies have characterized the HLA contribution, my work focuses on non-HLA factors because they remain comparatively underexplored yet hold promise for revealing novel regulatory variants influencing immune tolerance, intestinal homeostasis, and gluten-driven inflammation offering opportunities for precision diagnostics and new therapeutic interventions beyond the canonical HLA pathways [97] [104].

Chapter 3

Methodology

3.1 Identification of Celiac Disease - Associated SNPs

The study focused on identifying genes and single nucleotide polymorphisms associated with celiac disease. Genes linked to celiac diseases were collected from publicly available databases and previously published genome-wide association studies (GWAS).

These genes were selected based on their known roles in immune regulation, intestinal barrier function, and inflammatory responseSNPs linked to celiac disease were sourced from GWAS data and scientific publications.

To ensure precision in chromosomal positioning, reference and alternate alleles, and the related gene or intergenic region each SNP was validated using databases such as dbSNP and Ensembl. The final dataset included only human SNPs with verified genomic information and strong evidence linking them to celiac disease for further analysis.

3.2 Identification and Analysis of Celiac Disease Associated SNPs

After defining the gene set, single nucleotide polymorphisms SNPs associated with these genes were identified using genomic databases such as dbSNP and HaploReg v4.2. Each variant was examined to determine its chromosomal position, reference and alternative alleles and whether it was located in a coding, intronic or intergenic region. The information was further corroborated by consulting existing literature and the UCSC Genome Browser. For each SNP the corresponding amino acid sequence was retrieved to facilitate later functional analysis and prediction of variant impacts although specific sequence findings were elaborated in the results section.

3.3 Prediction of Potential Deleterious SNPs

The identified SNPs were then evaluated for their potential detrimental effects using computational prediction tools such as SIFT, PolyPhen-2, MutationTaster and the Likelihood Ratio Test . These algorithms assess the probability that a particular amino acid substitution or nucleotide change affects protein function by analyzing sequence conservation, physicochemical properties and evolutionary context. SNPs classified as damaging or possibly damaging by multiple tools were prioritized for further investigation.

3.4 Functional Analysis of Celiac Disease Associated SNPs

To better comprehend their biological importance a functional analysis was performed on the selected SNPs. Each variant was annotated according to its genomic

position, whether exonic, intronic or intergenic and evaluated for possible regulatory roles such as transcription factor binding or enhancer activity. RegulomeDB and HaploReg were utilized to assess the regulatory potential by incorporating data on epigenetic marks, DNase hypersensitivity sites and histone modifications. This approach provided insights into how non-coding variants might affect gene expression changes linked to celiac disease.

3.5 Structural Stability Analysis of Protein Variants

In addition to functional annotation, an examination of nonsynonymous SNPs nsSNPs was performed to evaluate their impact on protein structural stability. The tools I-Mutant 3.0 and SNPs3D were employed to assess the variations in Gibbs free energy △△G between the wild-type and mutant proteins. A negative △△G value indicated reduced stability suggesting that the amino acid substitution could destabilize the protein and impair its normal function. This structural stability analysis was crucial in identifying the SNPs most likely to cause detrimental conformational changes in proteins linked to diseases.

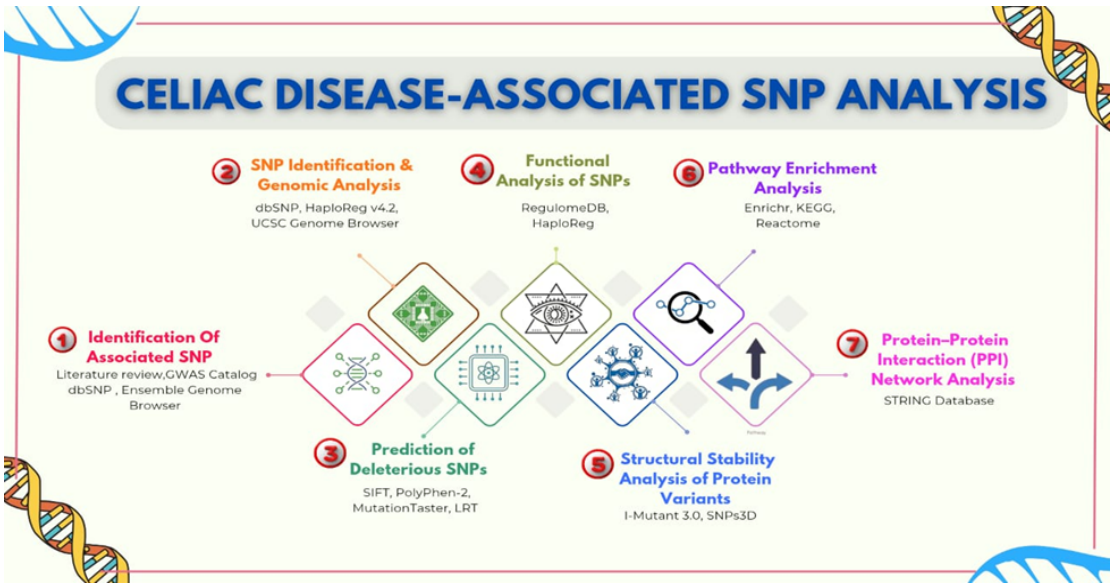


FIGURE 3.1: Methodology Overview

3.6 Pathway Enrichment Analysis

To explore the overall biological significance of the identified genes, a pathway enrichment analysis was conducted using Enrichr. The pathways enriched among genes with detrimental SNPs were pinpointed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases. This analysis aimed to determine if these genes were clustered within particular biological pathways particularly those associated with immune signaling cytokine activity or apoptosis processes which are frequently dysregulated in celiac disease.

3.7 Protein-Protein Interaction Network Analysis

The final phase involved examining the protein protein interaction network to reveal potential molecular interactions among proteins encoded by the chosen genes. Utilizing the STRING database an interaction network was constructed with a minimum confidence score of 0.7 to ensure its dependability. This visualization of the network allowed for the identification of key hub proteins and functional modules that might play a significant role in disease progression. This comprehensive systems biology approach provided an in-depth understanding of how genetic variants collectively influence molecular and cellular processes in celiac disease.

Chapter 4

Results

4.1 Identification of Celiac Disease - Associated Genes

A comprehensive compilation of genes associated with celiac disease was created by examining literature and exploring databases. Genes previously recognized for their roles in immune response, intestinal barrier regulation and inflammation were gathered from genome-wide association studies and academic publications. To verify these genes and their functional significance, resources such as the GWAS Catalog, UCSC Genome Browser and HaploReg v4.2 were employed. As a result a group of genes related to immunity and signaling including IL2, SOCS1, UBE2L3, PFKFB3, DAD1, PKIA, ABCC9 and KCNK10 was identified for further single nucleotide polymorphism SNP analysis.

Table 4.1: List of SNPs associated with celiac disease from HaploReg v4.2 and UCSC Genome browser.

rsID	**Position**	**Gene**	**Ref/Alt**	**Region**
rs117128341	chr8:79449773	PKIA	G/A	Intronic
rs73687528	chr8:79430748	PKIA	A/G	Intronic
rs79215674	chr8:79479883	PKIA	C/T	Intronic
rs74450608	chr8:79587012	IL7	T/C	Intronic
rs79374792	chr8:79590478	IL7	G/A	Intronic

Table 4.1 continued from previous page

rsID	Position	Gene	Ref/Alt	Region
rs117139146	chr10:6200555	PFKFB3	C/T	Intronic
rs17810546	chr3:159665049	Il12A/SCHIP1	A/G	Intergenic
rs183665868	chr18:8103224	PTPRM	G/A	Intronic
rs189838725	chr2:216021219	ATIC	G/T	Intergenic
rs192900921	chr6:40402691	LRFN2	G/A	Intronic
rs32723	chr5:3451905	LINC01019	T/G	Intronic
rs32726	chr5:3452577	LINC01019	T/C	Intronic
rs32727	chr5:3452885	LINC01019	G/C	Intronic
rs3748816	chr1:2526745	MMEL1	A/G	Missense
rs576626084	chr2:54974118	EML6	C/G	Intergenic
rs761616279	chr12:22005548	ABCC9	T/C	Intergenic
rs769377780	chr14:88777583	KCNK10	C/T	Intergenic
rs5979785	chrX:12971523	TLR7	C/T	Intergenic
rs653178	chr12:112007755	SH2B3	C/T	Intronic
rs13010713	chr2:181996044	ITGA4	A/G	Intergenic
rs13151961	chr4:123115501	IL2	A/G	Intronic
rs11712165	chr3:119118795	CD80	T/G	Intronic
rs10936599	chr3:169492100	MYNN	C/T	synonymous
rs13098911	chr3:46235200	CCR1,CCR3	C/T	Intergenic
rs17810546	chr3:159665049	Il12A/SCHIP1	A/G	Intergenic
rs2327832	chr6:137973067	OLIG3	A/G	Intergenic
rs2298428	chr22:21982891	UBE2L3	C/T	Missense
rs3184504	chr12:111884607	SH2B3	T/C	Missense
rs76830965	chr3:159637677	IL12A/SCHIP1	C/A	Intergenic
rs7616215	chr3:46205685	CCR1,CCR3	C/T	Intergenic
rs17264332	chr6:138005514	OLIG3	A/G	Intergenic
rs243323	chr16:11361201	SOCS1	A/G	Intergenic
rs13132308	chr4:123551113	IL2	A/G	Intergenic
rs990171	chr2:103086769	IL18R1, IL18RAP	A/C	Intergenic
rs551170288	chr2:60664485	MIR4432HG, BCL11A	C/G	Intergenic
rs780153546	chr14:22910739	LOC105370401, DAD1	C/A	Intergenic
rs137888770	chr15:46555534	LOC105370802, SEMA6D	G/A	Intergenic
rs1002929661	chr15:55560198	RAB27A	C/T	Intergenic
rs894868996	chr15:55868822	PYGO1	G/A	Intergenic
rs945505625	chr17:60075777	MED13	T/C	Intergenic

4.2 Identification and Analysis of Celiac Disease Associated SNPs

Following the identification of the genes, SNPs linked to the selected genes associated with celiac disease were collected through a combination of literature based evidence and bioinformatics databases. The literature review ensured the inclusion of variants already acknowledged as clinically or functionally important, while additional variants were obtained from databases such as HaploReg v4.2 and dbSNP to ensure a comprehensive and current collection. This thorough approach allowed for the inclusion of both experimentally validated and computationally predicted variants. A total of 41 SNPs were identified across various genes. Each SNP was associated with its chromosome, genomic region, and corresponding allele variants. Most were located in intronic or intergenic regions, suggesting potential regulatory roles rather than direct coding changes. A smaller subset of variants, including rs3748816 MMEL1, rs2298428 UBE2L3 and rs3184504 SH2B3 were found within exonic or missense regions, indicating possible structural impacts. Table 4.1 provides the complete dataset of these SNPs including chromosomal coordinates, gene associations and variant types, offering an overview of the genomic architecture of the analyzed variants and serving as a foundation for further functional, structural and pathway-based evaluations.

RegulomeDB-based graphical view of SNP effects on transcription factor binding and chromatin accessibility. SNPs (k) rs243323 (1f) and (ke) rs11712165, (g) rs13098911, (ij) rs2298428, (jl) rs3184504 (3a) show the highest regulatory potential while (bj) rs32726, (de) rs13151961 and (lo) rs13132308 (6) exhibit the weakest evidence.

4.3 Prediction of Potential Deleterious SNPs

To predict the functional importance of the identified SNPs, multiple in-silico tools were used. RegulomeDB and RegPotential scoring revealed several variants with high regulatory potential. SNPs with low RegulomeDB scores (1f-3a)

demonstrated strong evidence of transcription factor binding or enhancer activity. Notably, rs243323 SOCS1, rs11712165 CD80, rs13098911 CCR1/CCR3 rs2298428 UBE2L3 and rs3184504 SH2B3 were predicted to play key regulatory roles. Regulatory potential and conservation metrics are detailed in Table 4.2, with Figure 7 illustrating the graphical effects of SNPs on transcription factor binding and chromatin accessibility. These results reveal that specific variants likely affect gene expression by modifying promoter accessibility and transcriptional regulation in immune related genes.

TABLE 4.2: Functional analysis of celiac disease associated SNPs

Sr.	**SNP ID**	**RegPotential***	**Conservation****	**Regulome DB Score*****
1	rs117128341	-	-	No Data
2	rs73687528	-	-	No Data
3	rs79215674	-	-	No Data
4	rs74450608	-	-	No Data
5	rs79374792	-	-	6
6	rs117139146	-	-	No Data
7	rs17810546	0	0	4
8	rs183665868	-	-	No Data
9	rs189838725	-	-	No Data
10	rs192900921	-	-	5
11	rs32723	0.001756	0	5
12	rs32726	0	0	6
13	rs32727	0	0	No Data
14	rs3748816	0.206919	0	4
15	rs576626084	-	-	-
16	rs761616279	-	-	-
17	rs769377780	-	-	-
18	rs5979785	0	0	No Data
19	rs653178	0	0	5
20	rs13010713	0	0	5
21	rs13151961	0	0.001	6
22	rs11712165	0.049487	0.847	3a
23	rs10936599	0.272578	1	5
24	rs13098911	NA	0.001	3a
25	rs17810546	-	-	4
26	rs2327832	0	0.066	6

Table 4.2 continued from previous page

Sr.	SNP ID	RegPotential*	Conservation**	Regulome DB Score***
27	rs2298428	0.366729	1	4
28	rs3184504	0.287954	0.005	3a
29	rs76830965	-	-	5
30	rs7616215	NA	0	4
31	rs17264332	0	0	6
32	rs243323	0	0.086	1f
33	rs13132308	0	0.02	6
34	rs990171	NA	0.199	5
35	rs551170288	-	-	-
36	rs780153546	-	-	-
37	rs137888770	-	-	6
38	rs1002929661	-	-	-
39	rs894868996	-	-	-
40	rs945505625	-	-	-
41	rs117128341	-	-	No Data

4.4 Functional Analysis of Possibly Deleterious SNPs

Functional characterization revealed variants influence celiac disease. Variants impacted critical pathways including cytokine signaling, apoptosis and immune regulation.

Regulatory SNPs modulate transcription factor activity while coding variants likely impair protein function or signaling. Table 4.3 details significant SNPs with their predicted structural/regulatory impacts.

rs243323 displayed both strong regulatory effects and modest destabilization $\triangle\triangle G$ = -1.04 while rs192900921 (LRFN2) represented the most destabilizing variant $\triangle\triangle G$ = -5.22.

TABLE 4.3: Summary of high impact SNPs associated with celiac disease

SNP ID	**Gene**	**Region**	**DB Score**	$\Delta\Delta G$	**Functional Category**	**Predicted Impact**
rs 243323	SOCS1	Intergenic	1f	-1.04	Regulatory / Structural	Highest regulatory potential; may influence TF binding and immune-gene regulation
rs 11712165	CD80	Intronic	3a	-0.63	Regulatory	Moderate enhancer activity; possible effect on T-cell co-stimulation
rs 13098911	CCR1 / CCR3	Intergenic	3a	-0.33	Regulatory	May alter promoter accessibility in immune-cell loci
rs 2298428	UBE2L3	Missense	4	-1.12	Structural / Regulatory	Deleterious coding variant; potential effect on ubiquitin-pathway signaling
rs 3184504	SH2B3	Missense	3a	1.11	Regulatory	Reported autoimmune-risk variant; alters cytokine signaling
rs 192900921	LRFN2	Intronic	5	-5.22	Structural	Strongly destabilizing; may disrupt protein folding
rs 769377780	KCNK10	Intergenic	5	-3.46	Structural	Predicted to decrease protein stability; affects ion-channel regulation
rs 894868996	PYGO1	Intergenic	-	-1.87	Structural	Reduces stability of Wnt-pathway co-activator
rs 780153546	DAD1	Intergenic	-	-1.83	Structural	Destabilizing variant; may affect apoptosis-related protein
rs 137888770	SEMA6D	Intergenic	6	-1.32	Structural	Mild destabilization; may influence neuronal/immune signaling
rs 551170288	BCL11A	Intergenic	-	-1.44	Structural	Slightly destabilizing; may affect transcriptional regulation
rs 13151961	IL2	Intronic	6	-1.04	Structural / Regulatory	Weak regulatory evidence but minor destabilizing effect on cytokine structure

4.5 Structural Analysis of Protein Variants Based on Stability

I Mutant 3.0 analysis evaluated how amino acid changes affect protein stability. Several non-synonymous variants showed negative $\Delta\Delta$G values, indicating conformational destabilization; those with $\Delta\Delta$G < -1.0 were considered unstable. Table

4.4 details these predictions.

The most destabilizing substitutions were rs192900921 in LRFN2 $\Delta\Delta G$ = -5.22 rs769377780 in KCNK10 $\Delta\Delta G$ = -3.46 and rs894868996 in PYGO1 $\Delta\Delta G$ = -1.87.

Variants such as rs551170288 BCL11A and rs780153546 DAD1 also displayed substantial negative energy shifts, implying disrupted folding or degradation vulnerability. This evidence links protein destabilization to disease risk.

TABLE 4.4: SNPs involved in destabilizing protein structure.

rsID	**Position**	**Gene/ Region**	**Ref/ Alt**	**AA**	**Pos**	**I - MUTANT**	**3D SNP Score**
rs117128341	chr8:79449773	PKIA	G/A	S/N	59	-1.22	2.38
rs73687528	chr8:79430748	PKIA	A/G	S/N	59	-1.22	8.69
rs79215674	chr8:79479883	PKIA	C/T	S/N	59	-1.22	1.55
rs117139146	chr10:6200555	PFKFB3	C/T	A/T	115	-1.33	6.24
rs183665868	chr18:8103224	PTPRM	G/A	A/T	75	-1.08	1.03
rs192900921	chr6:40402691	LRFN2	G/A	S/G	68	-5.22	1.7
rs761616279	chr12:22005548	ABCC9	T/C	N/D	8	-1.42	2.4
rs769377780	chr14:88777583	KCNK10	C/T	T/A	4	-3.46	2.98
rs13151961	chr4:123115501	IL2	A/G	T/A	23	-1.04	1.24
rs2298428	chr22:21982891	UBE2L3, YDJC	C/T	K/E	20	-1.12	63.92
rs243323	chr16:11361201	SOCS1	A/G	V/I	13	-1.04	7.76
rs13132308	chr4:123551113	IL2	A/G	T/A	23	-1.04	1.84
rs551170288	chr2:60664485	MIR4432HG BCL11A	C/G	A/G	129	-1.44	9.48
rs780153546	chr14:22910739	LOC105370401 DAD1	C/A	I/T	20	-1.83	0.71
rs137888770	chr15:46555534	LOC105370802 SEMA6D	G/A	V/F	3	-1.32	1.12
rs1002929661	chr15:55560198	RAB27A	C/T	S/N	56	-1.79	39.91
rs894868996	chr15:55868822	PYGO1	G/A	S/A	11	-1.87	5.8

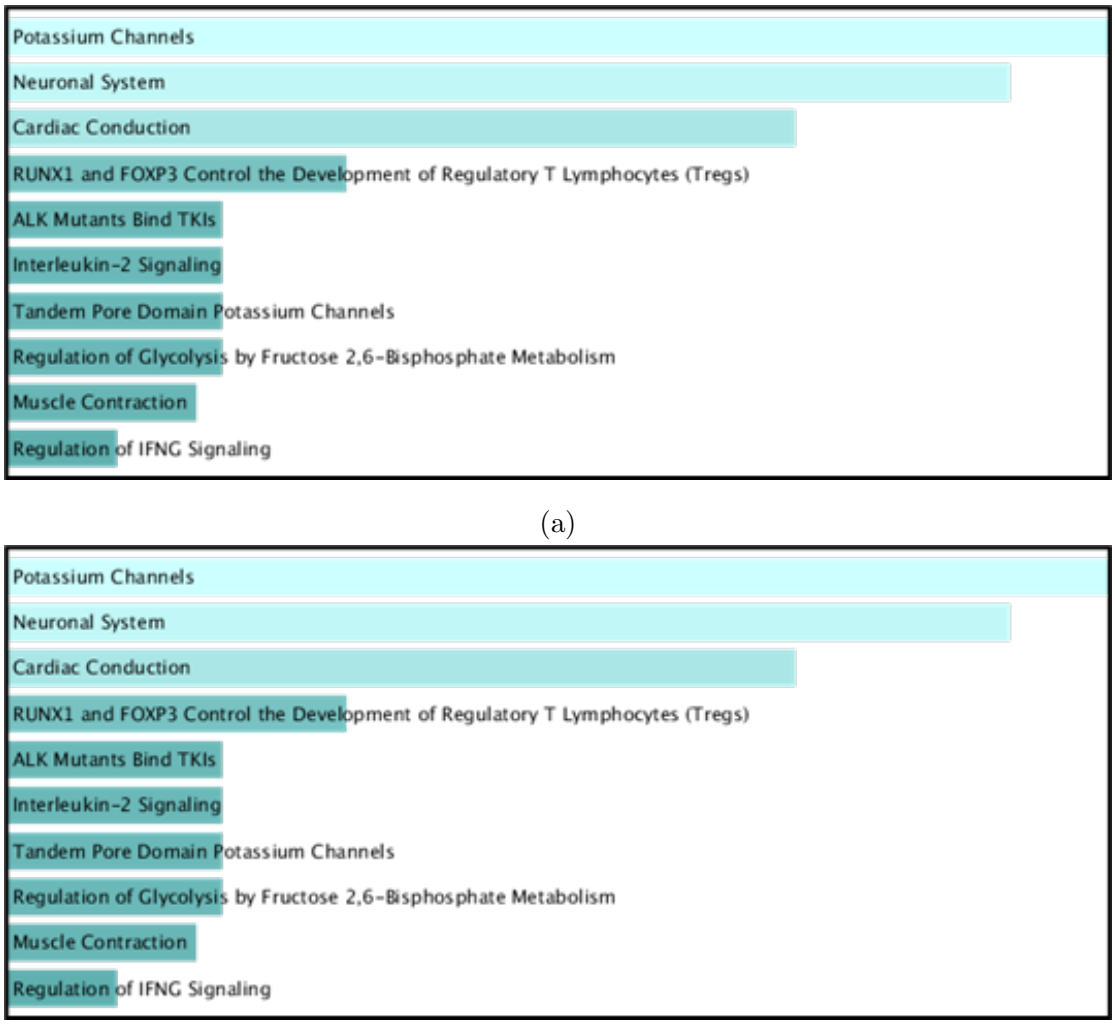
TABLE 4.5: Genes list from enricher kegg pathway

Gene list
Suppressor of Cytokine Signaling 1 (SOCS1)
Ubiquitin Conjugating Enzyme E2 L3 (UBE2L3)
Interleukin 2 (IL2)
6-Phosphofructose-2-Kinase/Fructose-2,6-Bisphosphate 3 (PFKFB3)
Defender Against Apoptotic Death 1 (DAD1)
ATP Binding Cassette Subfamily C Member 9 (ABCC9)
Protein Tyrosine Phosphate, Receptor Type M)PTPRM)
Potassium Two Pore Domain Channel Subfamily K Member 10 (KCNK10)
Semaphoring 6D (SEMA6D)
Protein Kinase Inhibitor Alpha (PKIA)

4.6 Pathway Enrichment Analysis

Pathway enrichment analysis via Enrichr emphasized major biological networks influenced by the genes of interest. Using KEGG and Reactome databases, we identified significant enrichment in immune related pathways notably IL2 signaling and cytokine cytokine receptor interactions among SNP harboring genes. Reactome (Figure 8A) and KEGG (Fig 8B) enrichment maps illustrate clusters involving immune, signaling and apoptotic pathways. Table 4.5 details the enriched KEGG pathway genes Suppressor of Cytokine Signaling 1, Ubiquitin Conjugating Enzyme E2 L3, Interleukin 2, 6-Phosphofructo-2-Kinase / Fructose-2,6-Bisphosphatase 3, Defender Against Apoptotic Death 1, TP Binding Cassette Subfamily C Member 9, Protein Tyrosine Phosphatase, Receptor Type M, Potassium Two Pore Domain Channel Subfamily K Member 10, Semaphorin 6D

Table: 4.6 provides a comprehensive breakdown of the IL2 signaling pathway featuring cytokine IL2 receptors interleukin receptor 2(alpha,beta and gamma), janus kinases 1,3 transducers, signal transducer and activator of transcription proteins (3,5,alpha ,beta) transcription factors and SOCS1 negative regulation. As shown in Fig 9 this pathway emphasizes ligand-receptor interactions and downstream mechanisms governing T cell production.



(a)

(b)

FIGURE 4.1: a) Reactome pathway from enrichr b) Kegg Pathway from enricher.

TABLE 4.6: Genes involved in IL2 signaling pathway identified from enrichment and KEGG analysis

Functional Role	**Gene Symbol**	**Description**
Cytokine (ligand)	IL2	Activates proliferation and differentiation of T cells
Receptor subunits	IL2RA, IL2RB, IL2RG	Form the IL2 receptor complex
Signal transducers	JAK1, JAK3	Kinases that activate STAT proteins
Transcription factors	STAT3, STAT5A, STAT5B	Induce expression of IL2-responsive genes
Negative regulator	SOCS1	Feedback inhibitor of IL2–JAK–STAT signaling

First he KEGG pathway diagram shows key cytokine families and receptor complexes. IL2 highlighted in red binds its receptor subunits Interleukin 2 receptor subunit alpha, interleukin 2 receptor subunit beta, interleukin 2 receptor subunit gamma to form the high-affinity IL2 receptor complex initiating downstream immune signaling by the interleukin 2 Janus kinases signal transducer and activator of transcription proteins pathway that regulates T cell proliferation and immune responses. Secondly magnified view of the interleukin 2, interleukin 2 receptor complex illustrating ligand receptor binding within the class one helical cytokine γ chain family.

4.7 Interaction Network, Chemical Association Analysis of Protein

Using STRING v12.0 we generated a protein protein interaction network to examine connections among prioritized genes. As shown in Fig 10 the network reveals robust interactions within the central cluster of SOCS1, IL2, DAD1 and UBE2L3 connecting immune signaling to apoptotic regulation. However, the genes on the periphery includes PFKFB3, KCNK10, PKIA and PTPRM, exhibit limited direct links but contribute to secondary pathway functions.

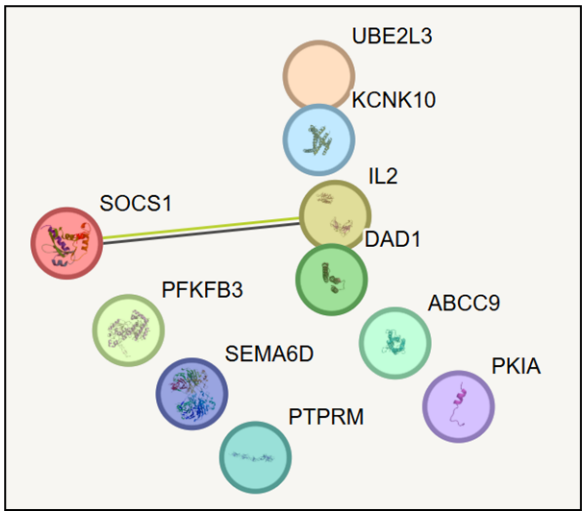


FIGURE 4.2: PPI network illustrating interactions among proteins from prioritized genes, constructed via the STRING v12.0 database.

It highlights high-confidence associations score ≥ 0.7 for proteins encoded by deleterious SNP-containing genes linked to celiac disease. Proteins are depicted as nodes, with edges representing functional links supported by experimental and predictive data. The central cluster formed by SOCS1, IL2, DAD1, and UBE2L3 indicates strong regulatory connectivity within immune and apoptotic pathways, while peripheral nodes (PFKFB3, KCNK10, ABCC9, PTPRM, SEMA6D, and PKIA) suggest gene-specific or indirect involvement in immune regulation and signaling processes.

Chapter 5

Discussion

In this study, in silico approach was used to examine non-HLA genetic factors in celiac disease (CD) susceptibility by utilizing literature-based selection of genes and SNP annotation, functional and regulatory forecasting, structural stability estimation, pathway enrichment, and protein-protein interaction evaluation. Although HLA-DQ2 and HLA-DQ8 haplotypes are the most powerful inherited determinants of CD, there is growing evidence that non-HLA loci are instrumental in the regulation of immune dysregulation, heterogeneity of the disease, and clinical expression. This opinion is justified by the current results which prove that non-HLA risk is mediated by a set of regulatory and structural and network-level processes that influence immune signaling, epithelial homeostasis, apoptosis, and metabolic processes. Past genome-wide association investigations have revealed over 40 non-HLA-linked loci linked to the occurrence of CD, and most of those changes are in the intronic or intergenic regions, and not within the protein-coding sequence. This distribution indicates that regulation difference, and not an explicit change in protein structure, is a significant constituent of non-HLA genetic vulnerability. In line with this paradigm, the current analysis laid more emphasis on variants that are enriched with regulatory factors like enhancers, promoters, and transcription factor binding sites. The loci that were well developed, such as the IL2IL21 region, SH2B3, CCR family of genes, and UBE2L3, were duplicated and further contextualized in the immune regulatory pathways. The fact

that a number of the prioritized loci overlap with the ones reported in other autoimmune disorders such as type 1 diabetes, rheumatoid arthritis that indicates the pleiotropy of immune regulatory genes and reinforces the idea of common pathogenic processes between autoimmune diseases.

The genes that have been given priority in this research paper are going to focus on biologically consistent pathways at the centre of CD pathophysiology. Cytokine signaling, activation of T cell, and tolerance(immune). One of the fundamental functions of core immune regulators, such as IL2, SOCS1, and SH2B3 is to regulate the JAK-STAT pathway. The disturbance in the functioning of these processes is a characteristic feature of gluten-mediated inflammation of the intestines. On the same note, genes in the co-stimulatory signaling and leukocyte recruitment like CD80 and CCR family members also indicate the significance of aberrant immune activation and cellular movements into the intestinal mucosa. Notably, the fact that regulatory forms of the genes influence them most supports the hypothesis that rather than structural gene disruptions, altered thresholds of gene expression influence much of the non-HLA genetic risk of CD.

In addition to regulation of immunity, the results also refer to the role of metabolic, epithelial, and stress-response pathways. PFKFB3, KCNK10, and DAD1 are some of the genes indicating that metabolic reprogramming, ion transport imbalance, and apoptosis could be among the causes of epithelial barrier dysfunction and severity of the disease. These observations are in line with developing evidence that the pathogenesis of CD cannot be entirely attributed to the immune mechanism, but its occurs because of complex interplay between the immune cells and intestinal epithelium. A combination of metabolic and epithelial stress pathways and immune signaling offers more detailed models of disease progression and development.

The structural stability analysis determined a wide range of variants that are predicted to destabilize protein structure, meaning that protein instability and dysfunctional molecular interactions could be an alternate, although less frequent, pathway of non-HLA genetic risk. Although these predictions must be interpreted carefully, since in silico predictions of $\triangle\triangle$G do not always relate to defective

functioning in vivo, their occurrence in genes associated with epithelial integrity, apoptosis and signaling networks could have biological significance. It is interesting to note that the discovery of destabilizing variants in genes which have not been viewed as inherent CD loci expands the breadth of the genetic contributors and utilizes the importance of structural analyses in tandem with regulatory annotation.

The analysis of pathway enrichment and protein-protein interaction further supported the systems-level of non-HLA genetic risk of CD. The abundance of cytokine signaling and immune interaction pathways is indicative of well-developed disease pathways, and the crosstalk between immune signaling, ubiquitination, and metabolic regulation is suggestive of the interrelation of these processes in chronic intestinal inflammation. The analysis of a central cluster of interactions between IL2, SOCS1, UBE2L3, and DAD1 confirms the notion that CD is due to the combined effect of several variants with small individual effects but not because of the malfunctioning of one gene. More precise fine-tuning of disease susceptibility and severity may be accomplished by peripheral network components, such as signaling modulation genes and epithelial activity genes.

The integrative approach used in this study is a strength of the research paper because it goes beyond the traditional association-related interpretations, where the other variants are prioritized on the basis of their functional relevance, structural impact, and network connectivity. However, there are also a number of limitations that should be admitted. These datasets might be biased to the population as most of the GWAS data are collected on people of European descent, and thus it may not be generally applicable. Furthermore, regulatory annotations are tissue- and context-specific and can have divergent functional impacts of prioritized variants across cell types and disease stages. This lack of experimental support also limits the possibility of causality inference and it is important to highlight the requirement of functional experiments to test predicted effects.

Overall, the present study suggests that the non-HLA genetic risk of celiac disease represents a multifactorial process where regulatory processes that mediate non-HLA genetic risk through changes in immune thresholds, epithelial functions,

and metabolic homeostasis have a significant role, and structurally destabilizing variants influence the process to a minor degree. Combining regulatory, structural, and network-based methods to analyze the data, the results will offer a narrower perspective on the non-HLA genetic makeup of CD and the relevance of systems-level analyses to the interpretation of complex autoimmune diseases. It will be necessary to apply the insights in better risk prediction and therapeutic approaches, which future studies combining functional validation with transcriptomic and epigenomic profiling and gene-environment interactions will be necessary.

Chapter 6

Conclusion

This study aimed to identify and prioritize non-HLA genetic features responsible for celiac disease (CD) susceptibility through a systematic in silico framework integrating literature-based gene selection, SNP annotation, functional and regulatory prediction, structural stability analysis, pathway enrichment, and protein-protein interaction mapping. This approach allowed for deeper understanding of genetic risk. The results indicate that celiac diseas, non-HLA genetic risk is mainly driven by regulatory variants with most disease-linked SNPs found in intronic or intergenic regions. The key genes identified are involved in cytokine signaling, immune regulation, epithelial homeostasis, apoptosis, metabolism, and ion transport, emphasizing the complex nature of CD pathogenesis. Network and pathway analyses uncovered coordinated involvement of immune signaling pathways ubiquitin mediated regulation, metabolic reprogramming and epithelial stress responses.

Beyond regulatory effects, structural analyses revealed potentially destabilizing variants in several genes, changes in protein stability might also play a role in disease susceptibility. these findings support a systems-level model where CD results from the combined effects of multiple non-HLA variants impacting immune, epithelial, and metabolic pathways, rather than the disruption of a single gene or process. Although this study offers a computational framework for non-HLA

genetic contributors to CD. Future research should include functional assays, transcriptomic and epigenomic data, gene environment interactions, and polygenic risk modeling to validate and expand these findings. This work enhances the understanding of the non-HLA genetic architecture in celiac disease and highlights the regulatory and network-based mechanisms in determining disease susceptibility.

Bibliography

[1] R. Parra-Medina and A. C. Cherñavsky, "Celiac disease," in *Autoimmunity: From Bench to Bedside*, J. M. Anaya, Y. S. A. R. e. a., Eds. Bogotá, Colombia: El Rosario University Press, 2013.

[2] A. Di Sabatino and G. R. Corazza, "Celiac disease: From pathophysiology to novel therapies," *Foods*, vol. 14, no. 6, 2024.

[3] A. Fasano and C. Catassi, "Clinical practice. Celiac disease," *N. Engl. J. Med.*, vol. 367, no. 25, pp. 2419–2426, 2012.

[4] A. Rubio-Tapia, I. D. Hill, C. P. Kelly, A. H. Calderwood, and J. A. Murray, "ACG clinical guidelines: Diagnosis and management of celiac disease," *Am. J. Gastroenterol.*, vol. 108, no. 5, pp. 656–676, 2013.

[5] A. Saviano *et al.*, "Gut microbiota alteration and its modulation with probiotics in celiac disease," *Biomedicines*, vol. 11, no. 10, p. 2638, 2023.

[6] P. Singh, A. Arora, T. A. Strand, D. A. Leffler, C. Catassi, P. H.. R. Green, C. P. Kelly, V. Ahuja, and G. K. Makharia, "Global prevalence of celiac disease: Systematic review and meta-analysis," *Clin. Gastroenterol. Hepatol.*, vol. 16, no. 6, pp. 823–836, 2018.

[7] S. Ashtari *et al.*, "Prevalence of celiac disease in low and high risk population in Asia-Pacific region: A systematic review and meta-analysis," *Sci. Rep.*, vol. 11, no. 1, 2021.

[8] M. Rashid and A. G. Khan, "Celiac disease in Pakistan: Challenges and opportunities," *J. Ayub Med. Coll.*, vol. 21, no. 3, 2009.

[9] M. A. Khan, H. Jan, M. O. Khan, M. Hanif, T. Hassan, and L. Khurshid, "Prevalence of celiac disease in adult patients with iron deficiency anemia," *Pak. J. Med. Health Sci.*, vol. 16, no. 5, 2022.

[10] Hassan and M. Asim, "High prevalence of celiac disease among pediatric population; A cross sectional survey," *Riphah J. Allied Health Sci.*, vol. 1, no. 1, 2022.

[11] S. Aboulaghras *et al.*, "Meta-analysis and systematic review of HLA DQ2/DQ8 in adults with celiac disease," *Int. J. Mol. Sci.*, vol. 24, no. 2, p. 1188, 2023.

[12] F. Megiorni and A. Pizzuti, "HLA-DQA1 and HLA-DQB1 in celiac disease predisposition: Practical implications of the HLA molecular typing," *J. Biomed. Sci.*, vol. 19, no. 1, p. 88, 2012.

[13] F. Biagi *et al.*, "Influence of HLA-DQ2 and DQ8 on severity in celiac disease," *J. Clin. Gastroenterol.*, vol. 46, no. 1, pp. 46–50, 2011.

[14] A. Sharma *et al.*, "Identification of non-HLA genes associated with celiac disease and country-specific differences in a large, international pediatric cohort," *PLoS ONE*, vol. 11, no. 3, 2016.

[15] M. Kumar *et al.*, "Autoimmune disease: Genetic susceptibility, environmental triggers, and immune dysregulation. Where can we develop therapies," *Front. Immunol.*, vol. 16, pp. 1–13, 2026.

[16] A. K. Akobeng, "Role of the gut microbiota in the pathogenesis of coeliac disease and potential therapeutic implications," *Eur. J. Nutr.*, vol. 59, pp. 3369–3390, 2020.

[17] E. Lionetti, F. Gatti, S. Pulvirenti, and M. Catassi, "The role of environmental factors in the development of celiac disease," *Nutrients*, vol. 7, no. 10, pp. 4403–4423, 2015.

[18] B. Lebwohl, J. A. Murray, E. F. Verdú, and C. Khosla, "Gluten introduction, breastfeeding, and celiac disease: Back to the drawing board," *Am. J. Gastroenterol.*, vol. 111, no. 1, pp. 12–22, 2016.

[19] M. Mladenović *et al.*, "Influence of breastfeeding and timing of gluten introduction on the onset of celiac disease in infants," *Srp. Arh. Celok. Lek.*, vol. 147, no. 11, pp. 683–693, 2019.

[20] E. Gnodi, R. Meneveri, and D. Barisani, "Celiac disease: From genetics to epigenetics," *World J. Gastroenterol.*, vol. 28, no. 4, pp. 449–463, 2022.

[21] M. Baran *et al.*, "Epigenetic mechanisms of genes influencing immune response in patients with celiac disease," *J. Pediatr. Res.*, vol. 8, p. 08456, 2024.

[22] A. J. Monsuur and C. Wijmenga, "Understanding the molecular basis of celiac disease: What genetic studies reveal," *Ann. Med.*, vol. 38, no. 8, 2006.

[23] D. Stepniak, "Celiac disease—Sandwiched between innate and adaptive immunity," *Hum. Immunol.*, vol. 67, no. 6, pp. 460–468, 2006.

[24] J. F. Ludvigsson *et al.*, "Outcomes of patients with celiac disease: A systematic review and meta-analysis," *JAMA*, vol. 321, no. 13, pp. 1258–1268, 2019.

[25] N. R. Lewis and G. K. T. Holmes, "Risk of morbidity in contemporary celiac disease," *Expert Rev. Gastroenterol. Hepatol.*, vol. 4, no. 6, pp. 767–780, 2014.

[26] S. A. Larson *et al.*, "Prevalence and morbidity of undiagnosed celiac disease from a community-based study," *Gastroenterology*, vol. 153, no. 4, pp. 1–9, 2017.

[27] "Mortality and malignancy in celiac disease," *Gastrointest. Endosc. Clin. N. Am.*, vol. 22, no. 4, pp. 705–722, 2012.

[28] D. Schuppan and K.-P. Zimmer, "The diagnosis and treatment of celiac disease," *Dtsch. Arztebl. Int.*, vol. 110, no. 49, pp. 835–846, 2013.

[29] G. Malamut *et al.*, "Presentation and long-term follow-up of refractory celiac disease: Comparison of type I with type II," *Gastroenterology*, vol. 136, no. 1, pp. 81–90, 2009.

[30] V. A. Ionescu *et al.*, "Celiac disease: Diagnostic advances, differential challenges, and interface with non-celiac gluten sensitivity," *Gastrointest. Disord.*, vol. 7, no. 4, 2026.

[31] C. Gianfrani, S. Auricchio, and R. Troncone, "Adaptive and innate immune responses in celiac disease," *Immunol. Lett.*, vol. 99, no. 2, pp. 141–145, 2005.

[32] J. Voisine and V. Abadie, "Interplay between gluten, HLA, innate and adaptive immunity orchestrates the development of coeliac disease," *Front. Immunol.*, vol. 12, p. 674313, 2021.

[33] C. Catassi, J. R. Lionetti, and A. Fasano, "Celiac disease," *Lancet*, vol. 399, no. 10344, pp. 2413–2426, 2022.

[34] A. Levescot, G. Malamut, and N. Cerf-Bensussan, "Immunopathogenesis and environmental triggers in coeliac disease," *Gut*, vol. 71, no. 11, pp. 2337–2349, 2022.

[35] E. Lionetti *et al.*, "Introduction of gluten, HLA status, and the risk of celiac disease in children," *N. Engl. J. Med.*, vol. 371, no. 1, pp. 42–49, 2014.

[36] S. Martina *et al.*, "Genetic susceptibility and celiac disease: What role do HLA haplotypes play?," *Acta Biomed.*, vol. 89, no. 9, pp. 17–21, 2018.

[37] A. Sharma *et al.*, "Identification of non-HLA genes associated with celiac disease and country-specific differences in a large, international pediatric cohort," *PLoS ONE*, vol. 11, no. 3, 2016.

[38] "A major non-HLA locus in celiac disease maps to chromosome 19," *Gastroenterology*, vol. 125, no. 4, pp. 1032–1041, 2003.

[39] V. Izzo *et al.*, "Improving the estimation of celiac disease sibling risk by non-HLA genes," *PLoS ONE*, vol. 6, no. 12, 2011.

[40] L. M. Sollid *et al.*, "Genes and environment in celiac disease," *Acta Odontol. Scand.*, vol. 59, no. 3, pp. 183–186, 2001.

[41] D. Schuppan, Y. Junker, and D. Barisani, "Celiac disease: From pathogenesis to novel therapies," *Gastroenterology*, vol. 137, no. 6, pp. 1912–1933, 2009.

[42] D. Schuppan, M. D. Dennis, and C. P. Kelly, "Celiac disease: Epidemiology, pathogenesis, diagnosis, and nutritional management," *Nutr. Clin. Care*, vol. 8, no. 2, pp. 54–69, 2005.

[43] H. J. Freeman, "Adult celiac disease and its malignant complications," *Can. J. Gastroenterol.*, vol. 23, no. 8, pp. 643–657, 2009.

[44] A. Levescot, G. Malamut, and N. Cerf-Bensussan, "Immunopathogenesis and environmental triggers in coeliac disease," *Gut*, vol. 71, pp. 2337–2349, 2022.

[45] S.-W. Qiao, R. Iversen, M. Ráki, and L. M. Sollid, "The adaptive immune response in celiac disease," *Semin. Immunopathol.*, vol. 34, pp. 523–540, 2012.

[46] S. Martucciello *et al.*, "Interplay between type 2 transglutaminase (TG2), gliadin peptide 31-43 and anti-TG2 antibodies in celiac disease," *Int. J. Mol. Sci.*, vol. 21, no. 10, 2020.

[47] M. Fleur du and L. M. Sollid, "T-cell and B-cell immunity in celiac disease," *Best Pract. Res. Clin. Gastroenterol.*, vol. 29, no. 3, pp. 413–423, 2015.

[48] T. Lejeune, C. Meyer, and V. Abadie, "B lymphocytes contribute to celiac disease pathogenesis," *Gastroenterology*, vol. 160, no. 7, pp. 2608–2610, 2021.

[49] B. Meresse, G. Malamut, and N. Cerf-Bensussan, "Celiac disease: An immunological jigsaw," *Immunity*, vol. 36, no. 6, pp. 907–919, 2012.

[50] V. Abadie, V. Discepolo, and B. Jabri, "Intraepithelial lymphocytes in celiac disease immunopathology," *Semin. Immunopathol.*, vol. 34, pp. 551–566, 2012.

[51] "Interleukin-15-dependent T-cell-like innate intraepithelial lymphocytes develop in the intestine and transform into lymphomas in celiac disease," *Immunity*, vol. 45, no. 3, pp. 610–625, 2016.

[52] T. Hisamatsu, U. Erben, and A. A. Kühl, "The role of T-cell subsets in chronic inflammation in celiac disease and inflammatory bowel disease patients: More common mechanisms or more differences," *Inflamm. Intest. Dis.*, vol. 1, no. 2, pp. 52–62, 2016.

[53] B. Lebwohl *et al.*, "Diagnosis of celiac disease," *Gastrointest. Endosc. Clin. N. Am.*, vol. 22, no. 4, pp. 661–677, 2012.

[54] T. D. Pelkowski, "Celiac disease: Diagnosis and management," *Am. Fam. Physician*, vol. 89, no. 2, pp. 99–105, 2014.

[55] R. A. Shannahan and M. E. Leffler, "Diagnosis and updates in celiac disease," *Gastroenterol. Hepatol.*, vol. 17, no. 1, pp. 23–30, 2021.

[56] A. Rostom *et al.*, "The diagnostic accuracy of serologic tests for celiac disease: A systematic review," *Gastroenterology*, vol. 128, no. 4, pp. S38–S46, 2005.

[57] D. A. W. M. van der Windt, P. Jellema, and C. J. Mulder, "Diagnostic testing for celiac disease among patients with abdominal symptoms," *JAMA*, vol. 303, no. 17, pp. 1738–1746, 2010.

[58] C. P. Kelly *et al.*, "Advances in diagnosis and management of celiac disease," *Gastroenterology*, vol. 148, no. 6, pp. 1175–1186, 2015.

[59] B. Lebwohl *et al.*, "Diagnosis of celiac disease," *Gastrointest. Endosc. Clin. N. Am.*, vol. 22, no. 4, pp. 661–677, 2012.

[60] B. Iwańczak, K. Matusiewicz, and F. Iwańczak, "Clinical picture of classical, atypical and silent celiac disease in children and adolescents," *Adv. Clin. Exp. Med.*, vol. 22, no. 5, pp. 667–673, 2013.

[61] K. Kaukinen and K. Lindfors, "Novel treatments for celiac disease: Glutenases and beyond," *Dig. Dis.*, vol. 33, no. 2, pp. 277–281, 2015.

[62] M. V. Machado, "New developments in celiac disease treatment," *Int. J. Mol. Sci.*, vol. 24, no. 2, 2023.

[63] M. S. Alam *et al.*, "Genetic differences between diagnosed and undiagnosed celiac disease: A population-based study," *Hum. Genet.*, vol. 24, no. 1, 2026.

[64] L. Tsali, "Elucidating the non-genetic risk factors for celiac disease: An umbrella review of meta-analyses," *Eur. J. Gastroenterol. Hepatol.*, vol. 36, no. 9, 2024.

[65] S. Moreno-Grau *et al.*, "Polygenic risk score portability for common diseases across genetically diverse populations," *Hum. Genomics*, vol. 18, 2024.

[66] K. Sandås *et al.*, "Using LDpred2 to adapt polygenic risk score techniques for disease prediction," *Bioinformatics*, vol. 41, no. 7, 2026.

[67] G. Butler-Laporte *et al.*, "HLA allele-calling using multi-ancestry whole-exome data," *Commun. Biol.*, vol. 6, 2023.

[68] T. Naito *et al.*, "HLA imputation and its application to genetic and molecular studies," *Front. Genet.*, vol. 12, 2021.

[69] B. Zhao *et al.*, "Integration of multi-omic data to decipher immune-genetic networks in celiac disease," *BMC Genomics*, vol. 24, 2023.

[70] V. C. C. Luz *et al.*, "Celiac disease gut microbiome studies in the third millennium," *Front. Microbiol.*, vol. 15, 2024.

[71] G. A. Heap and D. A. van Heel, "Genetics and pathogenesis of coeliac disease," *Semin. Immunol.*, vol. 21, no. 6, pp. 346–354, 2009.

[72] P. Brar, A. R. Lee, S. K. Lewis, G. Bhagat, and P. H. R. Green, "Celiac disease in African-Americans," *Dig. Dis. Sci.*, vol. 51, no. 5, pp. 1012–1015, 2006.

[73] P. Singh *et al.*, "Global prevalence of celiac disease: Systematic review and meta-analysis," *Clin. Gastroenterol. Hepatol.*, vol. 16, no. 6, pp. 823–836, 2018.

[74] K. D. Katz *et al.*, "Screening for celiac disease in a North American population: Sequential serology and gastrointestinal symptoms," *Am. J. Gastroenterol.*, vol. 106, no. 7, pp. 1333–1339, 2011.

[75] S. M. Kim, T. Mayassi, and B. Jabri, "Innate immunity: Actuating the gears of celiac disease pathogenesis," *Best Pract. Res. Clin. Gastroenterol.*, vol. 29, no. 3, pp. 425–435, 2015.

[76] S. S. Kupfer and B. Jabri, "Pathophysiology of celiac disease," *Gastrointest. Endosc. Clin. N. Am.*, vol. 22, no. 4, pp. 639–660, 2012.

[77] R. P. Anderson, "Innate and adaptive immunity in celiac disease," *Curr. Opin. Gastroenterol.*, vol. 36, no. 6, pp. 470–478, 2020.

[78] M. I. San-Martín *et al.*, "Microbiome markers in gastrointestinal disorders: Inflammatory bowel disease, colorectal cancer, and celiac disease," *Int. J. Mol. Sci.*, vol. 26, no. 10, p. 4818, 2026.

[79] R. Fernández-Pinilla, C. Sousa, and I. Comino, "Celiac disease and gut microbiota: Role of diet and probiotics," *Front. Nutr.*, vol. 9, 2022.

[80] J. A. Ludvigsson, E. B. Murray, and C. Catassi, "Celiac disease and celiac gluten sensitivity," *Curr. Opin. Gastroenterol.*, vol. 25, no. 9, pp. 116–122, 2009.

[81] P. de Souza, M. Oliveira, R. Fernandes, and F. Santos, "Celiac disease and gluten intolerance: An integrative literature review," *Res. Soc. Dev.*, vol. 11, no. 15, 2022.

[82] A. Therrien, C. P. Kelly, and J. A. Silvester, "Celiac disease: Extraintestinal manifestations and associated conditions," *J. Clin. Gastroenterol.*, vol. 54, no. 1, pp. 8–21, 2020.

[83] P. Laurikka *et al.*, "Review article: Systemic consequences of coeliac disease," *Aliment. Pharmacol. Ther.*, vol. 56, no. 14, pp. S64–S72, 2022.

[84] N. Gujral, D. R. Freeman, A. Thomson *et al.*, "Celiac disease: Prevalence, diagnosis, pathogenesis and treatment," *World J. Gastroenterol.*, vol. 18, no. 6, pp. 6496–6519, 2012.

[85] G. Caio, U. Volta, A. Sapone, D. A. Leffler, R. De Giorgio, C. Catassi, and A. Fasano, "Celiac disease: A comprehensive current review," *BMC Med.*, vol. 17, 2019.

[86] D. Schuppan, "Celiac disease: Diagnosis and treatment," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 10, no. 4, pp. 134–150, 2013.

[87] C. Meijer *et al.*, "Celiac disease prevention," *Front. Pediatr.*, vol. 6, no. 1, pp. 19–29, 2018.

[88] P. H. R. Green and C. Cellier, "Celiac disease," *N. Engl. J. Med.*, vol. 357, no. 17, pp. 1731–1743, 2007.

[89] W. Holtmeier and W. F. Caspary, "Celiac disease," *Orphanet J. Rare Dis.*, vol. 1, no. 3, 2006.

[90] G. Caio *et al.*, "Celiac disease: A comprehensive current review," *BMC Med.*, vol. 17, no. 142, 2020.

[91] J. E. Rubin and S. E. Crowe, "Celiac disease," *Ann. Intern. Med.*, vol. 172, no. 1, 2020.

[92] L. Rodrigo, "Celiac disease," *World J. Gastroenterol.*, vol. 12, no. 41, pp. 6577–6584, 2006.

[93] K. Karell *et al.*, "HLA types in celiac disease patients not carrying the DQA105–DQB102 (DQ2) heterodimer: Results from a Swedish prospective study," *Am. J. Gastroenterol.*, vol. 98, no. 4, pp. 945–951, 2003.

[94] J. D. Godfrey *et al.*, "Morbidity and mortality among older individuals with undiagnosed celiac disease," *Gastroenterology*, vol. 139, no. 4, pp. 1–9, 2010.

[95] M. Sallese *et al.*, "Beyond the HLA genes in gluten-related disorders," *Front. Nutr.*, vol. 7, no. 2, pp. 14–24, 2020.

[96] W. Vader, D. Stepniak, Y. Kooy, and F. Koning, "The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses," *PNAS*, vol. 100, no. 21, pp. 12390–12395, 2003.

[97] J. A. Ellis, A. S. Kemp, and A. L. Ponsbony, "Gene–environment interaction in autoimmune disease," *Expert Rev. Mol. Med.*, vol. 16, no. 4, 2014.

[98] R. Uibo, Z. Tian, and M. E. Gershwin, "Celiac disease: A model disease for gene–environment interaction," *Cell. Mol. Immunol.*, vol. 8, no. 1, pp. 93–103, 2011.

[99] S. Krishnareddy, "The microbiome in celiac disease," *Gastroenterol. Clin.*, vol. 48, no. 1, pp. 115–126, 2019.

[100] F. Valitutti, S. Cucchiara, and A. Fasano, "Celiac disease and the microbiome," *Nutrients*, vol. 11, no. 10, 2019.

[101] "Gut microbiome investigation in celiac disease: From methods to its pathogenetic role," *Clin. Chem. Lab. Med.*, vol. 58, no. 3, 2019.

[102] G. K. Makharia, "Current and emerging therapy for celiac disease," 2024.

[103] S. Yoosuf and G. K. Makharia, "Evolving therapy for celiac disease," 2019.

[104] A. Fasano and C. Catassi, "Current approaches to diagnosis and treatment of celiac disease: An evolving spectrum," *Gastroenterology*, vol. 120, no. 3, pp. 636–651, 2001.

[105] A. J. Monsuur and C. Wijmenga, "Understanding the molecular basis of celiac disease: What genetic studies reveal," *Ann. Med.*, vol. 38, no. 8, 2006.

[106] M. F. Kagnoff, "Celiac disease: A gastrointestinal disease with environmental, genetic, and immunologic components," *Gastroenterol. Clin. North Am.*, vol. 21, no. 2, pp. 405–425, 1992.

[107] L. M. Sollid, "Molecular basis of celiac disease," *Annu. Rev. Immunol.*, vol. 18, pp. 53–81, 2000.

[108] M. F. Kagnoff, "Celiac disease: A gastrointestinal disease with environmental, genetic, and immunologic components," *Gastroenterol. Clin. North Am.*, vol. 21, no. 2, pp. 405–425, 1992.

[109] L. M. Sollid, "Molecular basis of celiac disease," *Annu. Rev. Immunol.*, vol. 18, pp. 53–81, 2000.

[110] V. Abadie, L. M. Sollid, L. B. Barreiro, and B. Jabri, "Integration of genetic and immunological insights into a model of celiac disease pathogenesis," *Annu. Rev. Immunol.*, vol. 29, pp. 493–525, 2011.

[111] G. Caio *et al.*, "Celiac disease: A comprehensive current review," *BMC Med.*, vol. 17, no. 142, pp. 1–20, 2019.