Exploring Genetic Influences on Food Allergies and Auto-Immune Diseases: Understanding Individual Variations

Rashonda D. Anderson

Virginia Polytechnic Institute and State University M.S. Online Master of Agriculture and Life Sciences in Applied Nutrition and Physical Activity

04/23/24

Graduate Committee:

Dr. Deborah J. Good, Committee Chair

Dr. Angela S. Anderson, Co-Committee Chair

Ms. Nikeya Thomas, Committee Member

Dr. Renee Boyer, Committee Member

Keywords: genotype, allergies, auto-immune disease; Single Nucleotide Polymorphism (SNP); protein

Abstract

Food allergies pose a significant health risk, affecting millions of Americans, with symptoms ranging from mild discomfort to life-threatening anaphylaxis. Likewise, autoimmune diseases, where the immune system attacks healthy tissues, encompass a wide range of conditions, each with distinct symptoms and impacts on various organs. Through an analysis of Single Nucleotide Polymorphisms (SNPs), this research explores associations between specific genetic markers and phenotypic outcomes related to allergies and autoimmune diseases. Utilizing data from the genetic testing service 23andMe, this study investigated how genetic makeup may be related to individual responses to food, shedding light on underlying molecular mechanisms. Results demonstrate significant associations between certain SNPs and disease outcomes, highlighting the potential for personalized interventions in managing allergies and autoimmune disorders.

This research contributes to a deeper understanding of personalized nutrition and may pave the way for novel approaches to disease prevention and management.

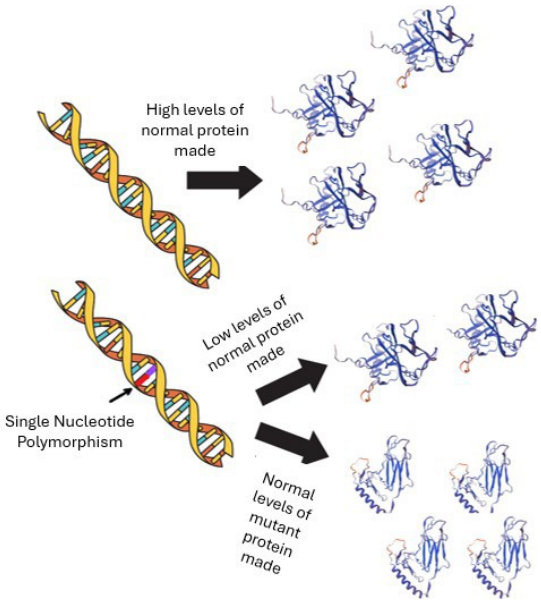
Exploring Genetic Influences on Food Allergies and Auto-Immune Diseases: Understanding Individual Variations

Introduction

The primary objective of this research was to delve into the profound impact of genetic variability on the intricate interplay between an individual's genetic predispositions and their body's response to dietary constituents, particularly in the context of allergies or autoimmune diseases. Food allergies pose a significant threat, with potentially life-threatening consequences, emphasizing the paramount importance of food safety. Based on the Food Allergy Research and Education it is estimated that 33 million Americans have food allergies, including 5.6 million children under age 18. That's one in 13 children, or roughly two in every classroom (FARE 2022). Food allergies can present with a range of symptoms, the most common of which include tingling or itching in the mouth, hives, swelling of the lips, face, tongue, and throat, respiratory issues such as wheezing or difficulty breathing, gastrointestinal symptoms like belly pain, diarrhea, nausea, or vomiting, and sensations of dizziness, lightheadedness, or fainting (Food Allergies | Food Safety and Inspection Service, 2012). However, the most severe allergic reaction triggered by food allergies is anaphylaxis, which manifests as constriction and tightening of the airways, swelling of the throat or the sensation of a lump in the throat, shock with a severe drop in blood pressure, rapid pulse, and potential loss of consciousness (FARE 2022). Recognizing these symptoms is critical for prompt intervention and management of food allergy reactions, especially to prevent the onset of anaphylaxis. Allergenic compounds in foods have long been a concern, given their capacity to trigger immune responses distinct from other food-related reactions such as intolerances, pharmacological responses, or toxicity. Among the myriad of allergens, the "Big Nine" includes soy, milk, eggs, peanuts, shellfish, tree nuts, grains, sesame, and fish (Food Allergies | Food Safety and Inspection Service, 2012).

Autoimmune diseases are when your immune system attacks the healthy cells of your organs and tissues by mistake. This can destroy body tissue and change in organ function, affecting various areas such as blood vessels, connective tissues, glands like the thyroid or pancreas, joints, muscles, red blood cells, and skin (Orbai, 2019). Individuals may experience multiple autoimmune disorders simultaneously, with common examples including Addison disease, Celiac Disease, dermatomyositis, Graves' disease, Hashimoto thyroiditis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, pernicious anemia, reactive arthritis, rheumatoid arthritis, Sjögren syndrome, systemic lupus erythematosus, and type I diabetes. Symptoms vary depending on the affected tissue or organ and may include fatigue, fever, malaise, joint pain, and rash (Orbai, 2019).

DNA is made up of sequences of nucleotides, which are the building blocks. These nucleotides are represented by the letters A (adenine), T (thymine), C (cytosine), and G (guanine). Genes are specific sequences of these nucleotides that encode instructions for building mRNA molecules that then code for proteins. Proteins are large, complex molecules that play many critical roles in the body. They are made up of chains of amino acids, which are coded by the sequences of nucleotides in genes. SNPs (Single Nucleotide Polymorphisms) are variations in a single nucleotide (A, T, C, or G) at a particular position in the DNA sequence within one's genome. They are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block, called a nucleotide (Figure 1). DNA variants can contribute to abnormal protein sequences, protein amounts, or have no effect.



• **Figure 1: DNA variants can contribute to abnormal protein sequences or protein amounts.** Top: Normal DNA sequences for proteins make normal proteins with the correct amount for normal function. Bottom: Single nucleotide variants lead to either lower protein amounts which lower the ability o the protein to perform its function, or mutant proteins that cannot function effectively.

23andMe is a genetic testing service that provides individuals with personalized insights into their ancestry, health traits, and genetic predispositions for certain conditions. Companies like 23andMe have increased a person's access to their genetic information and contributed to advancements in personalized medicine and genomic research. Through a meticulous examination of molecular pathways and mechanisms, this study aims to elucidate how genetic composition intricately modulates individual responses to food choices. By shedding light on the underlying molecular mechanisms, this research endeavors to establish a comprehensive understanding of how genetics influence the body's intricate responses to various types of foods. This exploration holds promise in not only deepening our knowledge of personalized nutrition but also in potentially uncovering novel insights into the prevention and management of allergies and autoimmune diseases. While 23andMe reports typically pull data from the coding strand of the rs (reference SNP), it's important to note that other literature might refer to data obtained from the complementary strand.

Literature Review

Genetic Code

The DNA molecule consists of two strands that wrap around each other to resemble a twisted ladder whose sides, made of sugar and phosphate molecules, are connected by rungs of nitrogen-containing chemicals called bases. Each strand is a linear arrangement of repeating similar units called nucleotides, which are each composed of one sugar, one phosphate, and a nitrogenous base. Four different bases are present in DNA: adenine (A), thymine (T), cytosine (C), and guanine (G). The particular order of the bases arranged along the sugar-phosphate backbone is called the DNA sequence; the sequence specifies the exact genetic instructions required to create a particular organism with its own unique traits (Richards, 1995).

mRNA is pivotal in translating the DNA sequence into a protein, which plays a pivotal role in one's biological manifestations. SNPs denote variations in a singular nucleotide (A, T, C, or G) within the DNA sequence, prevalent within the population. While SNPs constitute alterations in the DNA sequence, their repercussions sometimes extend to protein function and structure, thereby exerting significant influence on a multitude of biological processes and traits. SNPs are tiny changes in our DNA that can affect how our genes work. Some SNPs can influence how genes are turned into proteins, while others can impact how genes are regulated or controlled, and others seem to have no effect at all. For example, SNPs found in certain parts of genes, called exons, can change how a gene is read and used to make proteins. SNPs in other parts, called introns, could affect how the genetic material is cut and spliced together to create the final gene product. Additionally, SNPs located in specific regions before or after the gene sequence can influence how genes are translated into proteins or how certain molecules, called

microRNAs, can regulate gene activity. Overall, these small variations in our DNA can have big effects on our biology, influencing everything from our traits to our susceptibility to diseases (Deng et al., 2017).

Allergies

Research into the "Big 8" list of food allergens is important in supporting allergen prevention, treatment, and cures. "The Big 8" allergens identified since 2004 are milk, eggs, fish, shellfish, tree nuts, peanuts, wheat, and soybean. The 2021 Food Allergy Safety, Treatment, Education and Research (FASTER) Act required the FDA to add sesame as the ninth on its list of food allergens. Sesame will be subject to the same labeling regulations as the other eight major food allergens (Wiklund, 2021). Allergies like milk intolerance provoke immune reactions to food constituents, particularly proteins, leading to the production of immunoglobulin E (IgE). Consequently, the human body releases histamine, cytokines, and other inflammatory molecules, precipitating allergic symptoms ranging from mild inflammation to severe, potentially fatal anaphylactic shock (Krisnawati et al., 2022). Research conducted by Hong et al. in 2009 underscores the interplay of genetic and environmental factors in food allergies. While genetic predisposition plays a significant role, environmental influences and epigenetic modifications also contribute to allergic susceptibility. Gene-environment interactions, encapsulating the interplay between genomic function and environmental stimuli, profoundly impact phenotypic expression. Epigenetic alterations, and modifications in DNA chemical structure without affecting the coding sequence, further underscore the intricate mechanisms underlying food allergies. While genetic predisposition may underpin familial allergies, common environmental conditions likely play a pivotal role in their development (Hong et al., 2009).

Milk allergy and lactose intolerance are distinct conditions with unique causes and symptoms. Milk allergy results from an immune system reaction to specific proteins found in milk, leading to symptoms such as hives, vomiting, diarrhea, and in severe cases, anaphylaxis. On the other hand, lactose intolerance arises from insufficient levels of the enzyme lactase, which hampers the digestion of lactose sugar in milk and dairy products. Symptoms of lactose intolerance typically include bloating, abdominal pain, cramps, diarrhea, and gas, occurring within a few hours of consuming lactose-containing foods. While both conditions involve difficulties with milk consumption, they have different underlying mechanisms and necessitate different approaches to management and treatment (Mayo Clinic, 2018).

Autoimmune Diseases

Research shows that your body's vigilant defense system, known as the immune system, engages in a constant battle to safeguard one’s health. Its primary role is to protect against invading viruses and bacteria, thereby maintaining your well-being. However, there are instances when the immune system errs. Mistakenly identifying your body's healthy cells as foreign invaders, the immune system may launch an attack against them. This aberrant response can lead to the development of autoimmune disorders (NIH 2022).

Gluten, a protein naturally occurring in certain grains like wheat, barley, and rye, is linked to five main health conditions: Celiac Disease, non-celiac gluten sensitivity, wheat allergy, gluten ataxia, and dermatitis herpetiformis (Celiac Disease: Symptoms, Testing, Treatment & Research, n.d.). Each is distinct, but all are related and manageable. A prime example of an autoimmune disease resulting from food allergy is Celiac Disease. This research centers on Celiac Disease, an autoimmune condition triggered by gluten consumption, resulting in small intestine damage. The disease is hereditary, necessitating

the presence of HLA-DQ2 and/or HLA-DQ8 versions of our human leukocyte antigen genes, called the HLA-haplotype, for its development. However, merely possessing these haplotypes isn't sufficient; only about 1% ofindividuals with them develop the disease, despite HLA genes being co-dominantly expressed in everyone (Abadie et al., 2020). Additional factors or environmental influences are thought to affect the immune response causing inflammatory response, prompting the immune system to react to gluten and inflict intestinal damage. Although the precise triggers remain elusive, research suggests that stressors like illness, trauma, puberty, or childbirth may contribute. Beyond Celiac is committed to advancing research to uncover these activation factors, to prevent the onset of the disease (Beyond Celiac, 2019). Symptoms of Celiac Disease typically include gastrointestinal issues such as abdominal pain, diarrhea, and bloating, as well as fatigue and weight loss (Celiac Disease: Symptoms, Testing, Treatment & Research, n.d.).

Non-celiac gluten sensitivity, a milder form of gluten intolerance affecting 0.5–13% of individuals, presents with various symptoms. Bloating, diarrhea, constipation, stomach pain, headaches, fatigue, depression, anxiety, pain, and brain fog are common indicators. While bloating, abdominal discomfort, and digestive issues like diarrhea and constipation may arise after consuming gluten-containing foods, abdominal pain, and discomfort are prevalent symptoms. Additionally, gluten sensitivity may exacerbate migraine episodes, induce fatigue, anxiety, and depression, and contribute to widespread pain and brain fog. These symptoms, often associated with gluten exposure, underscore the importance of recognizing and managing non- celiac gluten sensitivity (21 Common Signs of Gluten Intolerance, 2022).

Wheat allergy is a food allergy eliciting an immune response to proteins found in wheat, including gluten and other compounds. While more prevalent among children,

approximately 66% outgrow wheat allergies by age 12. Symptoms include skin rash, commonly hives, manifesting shortly after wheat consumption and subsiding gradually. Digestive issues such as nausea, vomiting, stomach cramps, and diarrhea are frequent, alongside bloating, pain, and indigestion, driven by the immune system's response to wheat allergens. Nasal symptoms like sneezing, congestion, and a runny nose may indicate a wheat allergy, potentially leading to Baker's Asthma in some cases. Anaphylaxis, a severe allergic reaction causing swelling, hives, nausea, vomiting, and breathing difficulties, can be life-threatening and requires immediate medical attention, typically treated with epinephrine (21 Common Signs of Gluten Intolerance, 2022).

Gluten ataxia, a rare immune-mediated disorder, involves the body's immune system attacking the nervous system in response to gluten ingestion. While it can be associated with Celiac Disease or non-celiac gluten sensitivity, individuals with gluten ataxia may not necessarily exhibit the typical digestive symptoms seen in Celiac Disease. The adoption of a strict gluten-free diet is crucial in managing symptoms and halting the progression of cerebellar damage associated with gluten ataxia. Early diagnosis and intervention with a gluten-free diet can play a pivotal role in mitigating further neurological deterioration and improving the quality of life for affected individuals (*Gluten Ataxia and Celiac Disease | BeyondCeliac.org*, n.d.).

Dermatitis herpetiformis, a chronic skin condition linked to gluten intolerance, is characterized by a gluten-induced autoimmune response that manifests as a rash, typically on the elbows, knees, buttocks, and scalp. Dermatitis herpetiformis can affect individuals of any age but is most seen in those between 30 and 40 years old, particularly those with Celiac Disease or a family history of autoimmune conditions such as anemia, thyroid disease, or Type 1 Diabetes. Symptoms of dermatitis herpetiformis primarily involve skin

issues, such as discolored bumps and itchy blisters forming a rash but may also include oral problems like tooth enamel issues and gastrointestinal symptoms such as sensitivity to gluten and potential manifestations of Celiac Disease (Dermatitis Herpetiformis (DH): Definition, Causes & Treatment, n.d.).

SNPs associated with Food Allergies

With an estimated 33 million individuals sensitive to specific foods, multiple studies have sought out genetic correlations with food allergies. Each of the variants listed below was identified in one or more studies with a significant association with one or more food-based allergies in humans.

rs855791 on TMPRSS6

The rs855791 SNP, located on chromosome 22 at position 37462936 represents a single nucleotide polymorphism within the Transmembrane Protease, Serine 6 (*TMPRSS6*) gene (See Figure 2). For rs855971, if G is the most prevalent allele in the population, it is designated as the wildtype allele, whereas A is considered the mutant allele as it deviates from the wildtype allele. The *TMPRSS6* rs855791 polymorphism has been linked to impaired iron status in children diagnosed with Celiac Disease (CD). Urbaszek et al. conducted a study involving 106 pediatric patients with CD, which aimed to challenge the prevailing belief that persistent iron-deficiency anemia (IDA) is a frequent occurrence within this population. Initial findings revealed IDA in 23.6% of children at the time of diagnosis with CD, yet the majority (88%) experienced normalization of iron levels on a gluten-free diet (GFD) without iron supplementation. Despite observing a tendency toward a higher proportion of the T allele in CD patients with IDA, significant differences in the prevalence of the *TMPRSS6* rs855791 polymorphism were not evident between those with or without anemia. Interestingly, in pediatric CD patients, this polymorphism did not

emerge as a predictive factor for persistent anemia, suggesting potential differences in disease mechanisms between age groups (Urbaszek et al., 2021).

Although the *TMPRSS6* gene function primarily revolves around regulating iron levels within the body, studies have shown that individuals harboring a particular variant of this SNP,

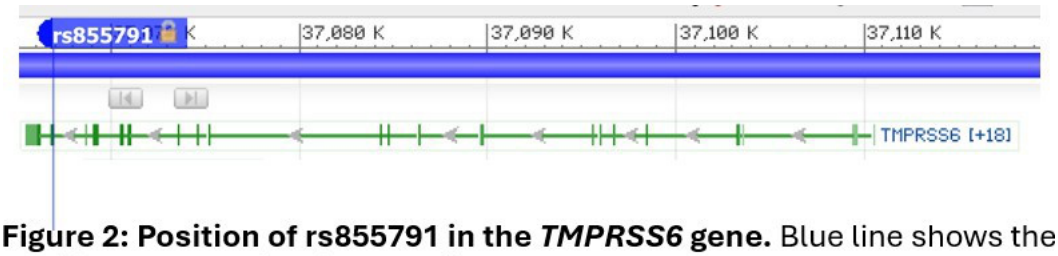


Figure 2: Position of rs855791 in the *TMPRSS6* gene. Blue line shows the position of the variant in the 20th exon of *TMPRSS6*.

characterized by an "A" allele in contrast to a "G" allele, demonstrate a heightened susceptibility to developing milk allergies (Livewello, 2024). While the precise mechanism remains incompletely elucidated, current hypotheses suggest a correlation between the genetic activity of this gene, immune system functionality, and the processing of specific proteins present in milk (Livewello, 2024).

rs7454108 between HLA-DQA1; HLA-DQB1; HLA-DQB2

Rs7454108, located on chromosome 6 at position 32681483, is a significant genetic variant situated between the Human Leukocyte Antigen-DQA1 (*HLA-DQA1*), Human Leukocyte Antigen-DQB1 (*HLA-DQB1*), and Human Leukocyte Antigen-DB2 (*HLA-DQB2*) genes (See Figure 3A and 3B). In the case of rs7454108, if T represents the most common allele in the population, it would be considered the wildtype allele, while C would represent the mutant allele as it differs from the wildtype allele. The HLA genes, located in the major histocompatibility (MHC) region on chromosome 6p21.3, play a role

in multiple autoimmune disorders, like Celiac Disease, Type 1 Diabetes (T1D), Rheumatoid Arthritis, Multiple Sclerosis, Psoriasis, and others. The MHC region is highly polymorphic and some genes in this region are involved in multiple disorders. For example, the *HLA-DQA1* and *-DQB1* genes have alleles that confer risk to both CD and T1D. In most autoimmune diseases, not all patients carry the same risk alleles, and multiple risk alleles are likely to be involved (Monsuur et al., 2008).

Human Leucocyte Antigen (HLA) complex haplotype typing stands as a cornerstone in Celiac Disease diagnosis, complementing serological tests for anti-endomysium and anti- transglutaminase autoantibodies in blood and endoscopic observations of intestinal inflammation. Particularly in the initial stages of diagnosis, HLA typing aids in excluding other potential causes of symptoms, aiding in the differentiation of CD pathogenesis (Sebastian- delaCruz & Castellanos-Rubio, 2023). Central to CD development are specific HLA proteins, notably *HLA-DQ2.5* and *HLA-DQ8*, which are encoded by distinct alleles. The expression of these receptors is contingent upon the presence of certain alleles, a characteristic that can be reliably predicted through the identification of tagging single nucleotide polymorphisms (SNPs). Specifically, the tagging SNP rs2187668 (not investigated in this study) serves as an accurate predictor for *HLA-DQ2.5*, while rs7454108 reliably identifies the presence of *HLA-DQ8* alleles (Sebastian-delaCruz & Castellanos-Rubio, 2023).

Based on a study by Barker et al., rs7454108 is a robust genetic marker for identifying individuals predisposed to Type 1 Diabetes. The high predictive values associated with this SNP highlight its potential clinical relevance in risk assessment and personalized management approaches for individuals at risk of developing Type 1 Diabetes (Barker et al., 2008). The article, Perspectives in Diabetes Cow's Milk and Type

1 Diabetes talks about the link early exposure to cow's milk or lack of breastfeeding to an increased risk of Type 1 Diabetes originated in the 1980s. It suggests that immunization to dietary antigens, which cross-react with pancreatic islet autoantigens, may play a role in this predisposition. For instance, there could be cross-reactivity between bovine insulin and human insulin, as well as between bovine k-casein and tyrosine phosphatase IA-2 (Harrison & Honeyman, 1999). Barker et al. also demonstrated a link between the presence of DQB1*0302 (refers to a specific allele of the human leukocyte antigen (HLA) gene, specifically the DQB1 locus. In the HLA nomenclature system, the asterisk indicates the specific allele designation. HLA genes encode proteins that play a crucial role in the immune system by presenting antigens to immune cells. The DQB1 gene is one of the genes within the HLA complex and encodes the beta chain of the HLA-DQ protein) and the C allele of rs7454108. Building upon this association, they aimed to harness this genetic marker to identify individuals with the highest susceptibility to Type 1 Diabetes. Through the implementation of a two-SNP test, they successfully differentiated various DR/DQ genotypes, particularly focusing on the discriminatory power of rs7454108 in conjunction with rs2040410 (not investigated in this study) for Type 1 Diabetes diagnosis (Barker et al., 2008).

In the study population of Denver, CO, the prevalence of the DQB1*0302 allele was observed to be 21.9% and the rs7454108 C allele demonstrated notable predictive values. Specifically, its positive predictive value stood at 97.5%, indicating its efficacy in identifying individuals with an increased risk for Type 1 Diabetes. Moreover, the rs7454108 C allele exhibited a high negative predictive value of 99.7%, underscoring its ability to accurately exclude individuals from the high-risk category (Barker et al., 2008).

While the rs7454108 represents a promising marker for various autoimmune

disorders, including Celiac Disease (CD) and Type 1 Diabetes (T1D), its location in a non-coding region complicates the understanding of its precise mechanism in promoting food allergies and autoimmune diseases. However, its association with specific HLA proteins,

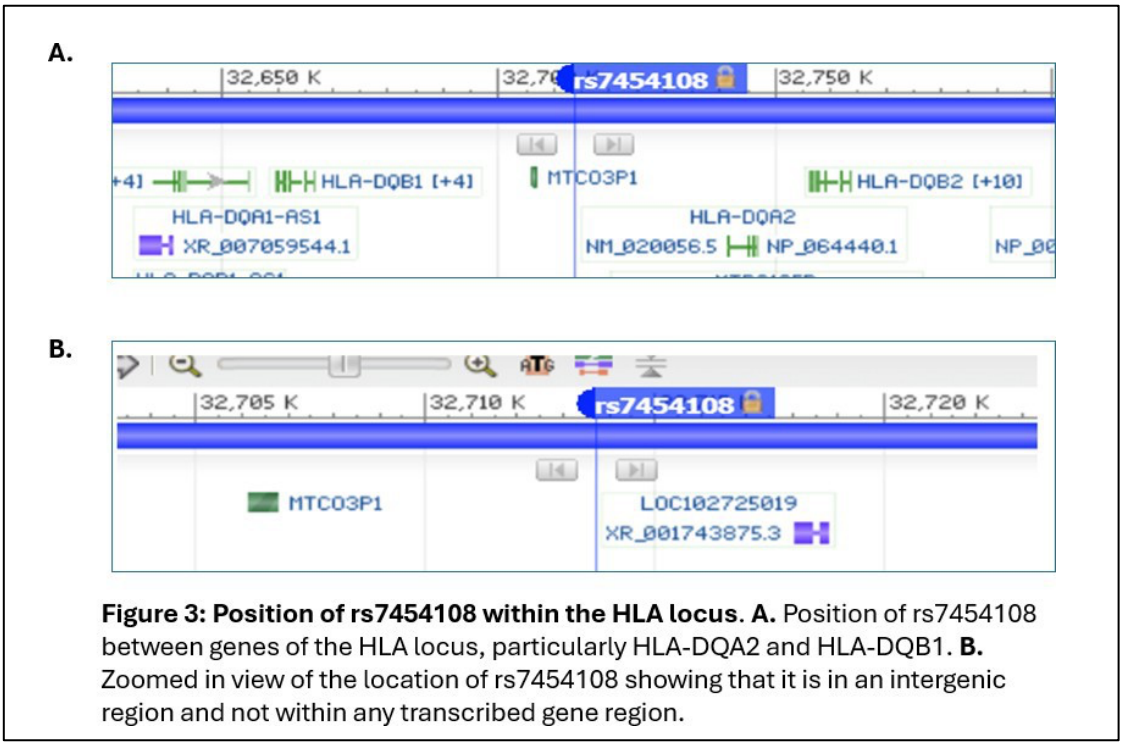


Figure 3: Position of rs7454108 within the HLA locus. A. Position of rs7454108 between genes of the HLA locus, particularly HLA-DQA2 and HLA-DQB1. **B.** Zoomed in view of the location of rs7454108 showing that it is in an intergenic region and not within any transcribed gene region.

such as HLA-DQ8 alleles, underscores its clinical relevance in risk assessment and personalized management strategies.

rs7192 on HLA-DRA

The genetic variant identified by rs7192, is located on chromosome 6 at position 32411646, within the gene Human Leukocyte Antigen-DR Alpha (*HLA-DRA*). For the rs7192 SNP, the wildtype allele is T and the mutant allele is G (See Figure 4). The genome-wide association study (GWAS) on food allergies (FA) highlights a significant locus at 6p21.32 specifically associated with peanut allergy (PA), revealing consistent findings across various analytical approaches. This study pinpoints susceptibility loci within the *HLA-DQ* and *-DR* region, particularly marked by rs7192 and rs9275596. These

genetic variants not only show associations with PA but also correlate with differential DNA methylation (DNAm) levels at multiple CpG sites, with DNAm alterations in the *HLA-DQB1* and *HLA-DRB1* genes potentially mediating the observed SNP-PA associations. With a population-attributable risk (PAR) ranging from 19% to 21%, these findings suggest that the *HLA-DR* and *-DQ* gene regions might represent a pivotal genetic risk factor for PA. The study underscores the necessity for further replication, validation, and functional investigations, which could enhance our understanding of the genetic and epigenetic mechanisms underlying PA risk, ultimately aiding in the development of more effective strategies for prediction, prevention, and treatment of peanut allergies (Hong et al., 2015).

The article, *A Crucial Genetic Factor for Peanut Allergy (HLA)*, is a pivotal 2015 study that included over 2,000 subjects and identified a robust association between two HLA variants and peanut allergies. When a "T" (reference allele) is substituted for an "A" (variant allele) on the rs7192 SNP, in the heterozygous configuration, correlated with 70% higher odds of peanut allergy (Ristic, 2020).

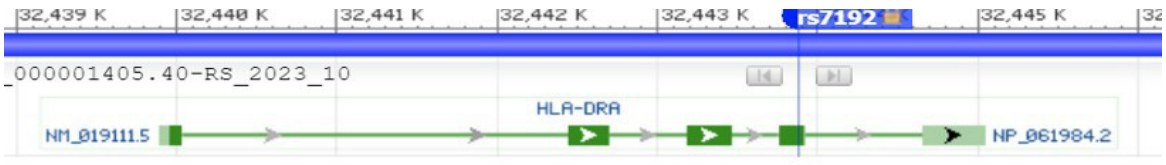


Figure 4: Position of rs7192 on *HLA-DRA* gene. The blue line shows the position of the variant in the 4th exon of *HLA-DRA*.

rs2187668 on HLA-DQA1

The genetic variant with reference number rs2187668, located on chromosome 6 at position 32411646, with the wildtype allele C and the mutant allele T. This represents a

key locus of interest in studies investigating the genetic basis of food allergies, including Celiac Disease (Figure 5). The findings align with the results of the genome-wide association study titled "A genome-wide association study for Celiac Disease identifies risk variants in the region harboring IL2 and IL21," highlighting the critical role of HLA-DQ2/8 in the genetic predisposition to Celiac Disease. The presence of the A allele, tagged by rs2187668, efficiently identifies individuals carrying the HLA-DQ2.5cis haplotype, which is strongly associated with Celiac Disease. This haplotype, wherein both chains of the DQ2 heterodimer are encoded on the same chromosome, was significantly more prevalent in UK celiac patients (89.2%) compared to population controls (25.5%). Further analyses, stratified by rs2187668 genotype in cases with the AG genotype and at rs9275141 in cases with the GG genotype, emphasizing the contribution of HLA-DQ2/8 in antigen presentation in Celiac Disease (van Heel et al., 2007).

In a meta-analysis investigating the association between the *HLA-DQA1* rs2187668 polymorphism and systemic lupus erythematosus (SLE), a total of 10 case-control studies involving 2,574 SLE cases and 3,548 controls were analyzed. The meta-analysis revealed a significant association between the *HLA-DQA1* rs2187668 polymorphism and SLE susceptibility across various populations. Specifically, individuals with the GG genotype and G allele were found to have an increased risk of developing SLE (Li et al. 2018). This suggests a potential role for the *HLA-DQA1* rs2187668 polymorphism in influencing the susceptibility to systemic lupus erythematosus.

Figure 5: Position of rs2187668 in the *HLA-DQA1* gene. The blue line shows the position of the variant in the 1st intron of *HLA-DQA1*.

rs41380347 on MCM6

Rs41380347, located on chromosome 2 at position 136608651, is a noteworthy genetic variant situated within the Minichromosome Maintenance Complex Component 6 (*MCM6*) gene which is notably upstream of the lactase enzyme gene *LCT* with a wildtype of A and mutant of G (See Figure 6). Lactose is a disaccharide of glucose and galactose, which accounts for most of the carbohydrates included in milk. Milk is a nutritionally ideal food, but people often suffer from digestive difficulties because the lactose of milk is not digested well but fermented by intestinal microorganisms. This symptom is called lactose intolerance (Chang Seok Oh et al., 2023). Individuals with milk intolerance may exhibit symptoms such as gastrointestinal discomfort or bloating due to genetic variations, which may be influenced by the rs41380347 SNP. In a study conducted by Chang Seok Oh et al., PCR amplification using specific Lac-1 and Lac-2 primer sets targeting various lactase SNPs, including rs41380347, was performed on DNA samples obtained from descendants from the Joseon Dynasty in Korea. Subsequently, DNA sequencing analysis was conducted to determine the genotypes of the SNPs, including rs41380347 (Chang Seok Oh et al., 2023). The consistent presence of the T allele for rs41380347 (LCT-13915) among Joseon Dynasty individuals suggests a high prevalence of lactase non- persistence within this population. This finding aligns with previous studies indicating a correlation between the T allele of rs41380347 and lactase non-persistence in variouspopulations. The genetic

homogeneity observed in the Joseon Dynasty population regarding lactase SNPs, including rs41380347, may reflect dietary practices and evolutionary adaptations specific to this historical period (Chang Seok Oh et al., 2023).

Figure 6: Position of rs41380347 on *MCM6* gene. Blue line shows the position of the variant in the 13th intron of *MCM6*.

rs4988235 on MCM6

SNP rs4988235 is located on chromosome 2 at position 136608646 on the Minichromosome Maintenance Complex Component 6 (*MCM6)* gene (Figure 7). The rs4988235 SNP is situated in the regulatory region upstream of the lactase gene (LCT) and plays a crucial role in controlling lactase enzyme production. Individuals carrying the T allele of rs4988235 exhibit enhanced lactase enzyme activity, allowing for the continued digestion of lactose beyond infancy, whereas individuals with a C allele see a correlation with lactose intolerance. This persistence of lactase activity enables lactase-persistent individuals to tolerate lactose-containing foods throughout adulthood, distinguishing them from lactase non-persistent individuals who experience lactose intolerance symptoms upon lactose consumption (Sivapalan, 2020).

Sivapalan et al. found that individuals with CT or TT genotypes are categorized as lactase persistent (LP), while those with the CC genotype are considered lactase non-persistent (LNP), or genetically lactose intolerant. Moreover, the term "autosomal" denotes that the 'T' variant of rs4988235 is located on an autosome, not on a sex chromosome (i.e.,

X or Y chromosome), thus exhibiting Mendelian inheritance patterns (Sivapalan, 2020).

The genetic basis of lactase persistence, particularly the role of rs4988235, has significant clinical implications. Genetic testing for rs4988235 can provide valuable insights into an individual's predisposition to lactase persistence or non-persistence, thereby informing dietary recommendations and management strategies for lactose intolerance. Additionally, knowledge of the inheritance pattern of rs4988235 aids in familial risk assessment and genetic counseling for individuals with lactose intolerance or lactase persistence (Sivapalan, 2020).

Figure 7: Position of rs4988235 on *MCM6* gene. The blue line shows the position of the variant in the 13th intron of *MCM6*.

rs7216389 on GSDMB

The genetic variant identified by the rs number rs7216389 on chromosome 17 is situated at position 38069949 within the Gasdermin B (*GSDMB*) gene with a wildtype of C and mutant of T (Figure 8A and 8B). Although there is no current research linking rs7216389 to allergies or autoimmune diseases, studies suggest potential associations with asthma and allergic rhinitis. Asthma is a chronic respiratory condition characterized by inflammation and narrowing of the airways, resulting in symptoms such as wheezing, shortness of breath, chest tightness, and coughing. These symptoms often occur in response to triggers such as allergens, pollutants, respiratory infections, exercise, or stress

(Akhouri & House, 2022). Allergic rhinitis, also known as hay fever, is a common allergic condition that affects the nasal passages. It occurs when the immune system overreacts to allergens such as pollen, dust mites, pet dander, or mold spores, leading to inflammation of the nasal lining. Symptoms of allergic rhinitis include sneezing, itching of the nose, congestion, runny nose (rhinorrhea), and postnasal drip. These symptoms can significantly impair quality of life, interfering with sleep, work, and daily activities (Akhouri & House, 2022).

The rs7216389 SNP is within an intron of the *GSDMB* gene. This gene engages in various cellular processes, particularly in inflammation and cell death. A second gene, Cadherin Related Family Member 3 (*CDHR3*) contains the rs6967330 SNP (See Figure 8A and 8B). For *GSDMB*, rs7216389, a C→T risk mutation has been characterized, while for *CDHR3* s6967330 a G→A risk mutation has been characterized, and both have been implicated in sinus disease susceptibility. To evaluate their combined impact, counts of these SNPs were amalgamated to form a genetic risk score based on the number of combined risk alleles. This scoring system categorized individuals into distinct risk groups: 0 risk alleles for the CC + GG genotype combination, 1 risk allele for CT + GG or CC + AG combinations, 2 risk alleles for CC + AA, CT + AG, or TT + GG combinations, 3 risk alleles for CT + AA or TT + AG combinations, and 4 risk alleles for TT + AA genotype combination (Zack et al., 2021). The combined genotype effect of risk alleles (gene dosage effect) at the University of Arizona (UofA) and the University of Pennsylvania (UPenn). In the UofA's population, among controls, individuals with increasing risk allele dosages showed a significant trend, with a decrease in the frequency of 0 alleles and an increase in the frequency of 1, 2, and 3 alleles ($p=0.002$). Similarly, among individuals with chronic rhinosinusitis (CRS), there was a significant trend observed, with a decrease in the

frequency of 0 alleles and an increase in the frequency of 1, 2, and 3 alleles. Notably, individuals with 4 risk alleles were present in both control and CRS groups, albeit at low frequencies. At UPenn, a similar trend was observed, with a significant difference in risk allele dosages between controls and CRS patients (p=0.027). Individuals with CRS exhibited a decreasing trend in the frequency of 0 alleles and an increasing trend in the frequency of 1, 2, and 3 alleles compared to controls (Zack et al., 2021).

Figure 8A: Position of rs7216389 on *GSDMB* gene. The blue line shows the position of the variant in the 1st intron of *GSDMB*.

Figure 8B: Position of rs6967330 on *CDHR3* gene. The blue line shows the position of the variant in the 11th intron of *CDHR3*.

rs12964116 on SERPINB7

Rs12964116, located on chromosome 18 at position 61442619, is a notable genetic variant within the Serpin Family B Member 7 (*SERPINB7*) gene with the wildtype of A and mutant of G (See Figure 9). The Genome-wide association study identifies the *SERPINB* gene cluster as a susceptibility locus for food allergy talked on associated with allergic diseases, specifically food allergies, offering insights into the complex genetic underpinnings of these conditions (Marenholz et al., 2017). The rs12964116 SNP within the *SERPINB7* gene emerges as a significant marker for food allergy susceptibility, displaying consistent association across multiple study sets (Marenholz et al., 2017).

Another SNP, rs1243064, within the *SERPINB* gene cluster exhibits an association with food allergy, suggesting multiple risk haplotypes in this locus. The study's thorough examination of specific genetic markers and their individual links to food allergies marks a significant stride forward in understanding the genetic basis of this condition. This could potentially open avenues for targeted interventions and personalized treatment approaches. (Marenholz et al., 2017).

Figure 9: Position of rs12964116 on *SERPINB7* gene. The blue line shows the position of the variant in the 2nd intron of *SERPINB7*.

rs9272346 on HLA-DQA1

Although rs9272346 located on the *HLA-DQA1* locus has been linked to asthma and allergic diseases (Ranjbar et al., 2022), there are currently no known articles associating this specific SNP with food allergies. The SNP rs9272346, located proximal to *the HLA-DQA1* locus, has been identified as significantly associated with an elevated risk of asthma across both adult and childhood populations Ranjbar et al., 2022). This finding underscores the pivotal role of the *HLA-DQA1* locus, situated within the MHC class II region, which harbors numerous polymorphisms implicated in asthma and allergic diseases. The genetic susceptibility conferredby rs9272346 highlights the complex interplay between immune regulation and asthma pathogenesis, shedding light on potential therapeutic targets for mitigating asthma risk and severity (Ranjbar et al., 2022).

Genome-wide association study (GWAS) further confirms HLA-DQ in the

diagnosis of asthma among adults, a study where Lasky-Su et al. identified 33 variants for further investigation, including rs9272346, situated proximal to *HLA-DQA1* on chromosome 6p21.3 (Figure 10). They demonstrated a significant association with adult-onset asthma (P-value = 2.2E-08), with consistent findings extending to a more diverse cohort of both adults and children (P-value = 1.0E-04). Additionally, several genes previously implicated in asthma through GWAS exhibited nominal associations with the condition within the studied populations, suggesting potential shared genetic mechanisms underlying asthma susceptibility across different demographic groups (Lasky-Su et al., 2012).

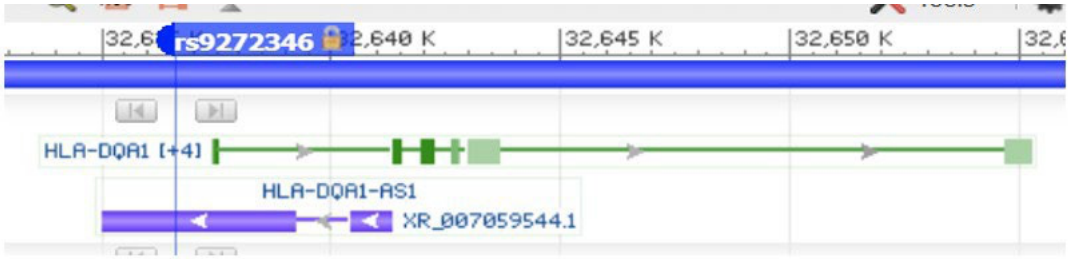
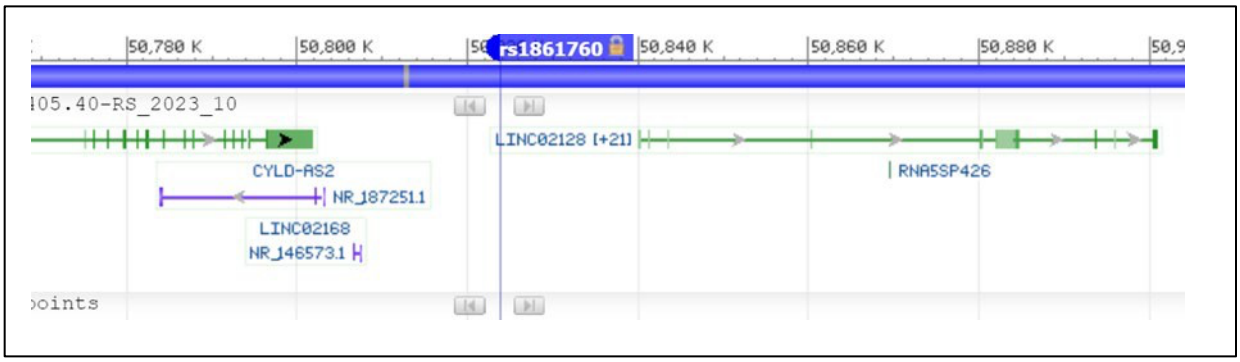


Figure 10: Position of rs9272346 upstream *HLA-DQA1* gene. The blue line shows the position of the variant within the *HLA-DQA1-AS1* antisense transcript.

rs1861760 Between LOC105371251 and LINC02128

Rs1861760, positioned on chromosome 16 at position 50857693, represents a significant genetic variant located between the genes LOC105371251 and LINC02128, with a widtype of A and a mutant of T (Figure 11). The genetic variant rs1861760, situated proximal to Cylindromatosis (*CYLD)* and Nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) genes on chromosome 16q12, holds significant implications for immune- related diseases. Notably, its location coincides with a binding site for FOXJ1, a transcription factor linked to allergic rhinitis, suggesting potential involvement in allergic responses (Sarnowski et al., 2016). Furthermore, variants within a 130-kb region

encompassing rs1861760 have been associated with inflammatory bowel diseases, particularly Crohn's disease, as well as leprosy, highlighting its multifaceted role in modulating immune function and susceptibility to diverse pathological conditions (Sarnowski et al., 2016).



rs1059513 on STAT6

Rs1059513 is located on chromosome 12 at position 57489709, within the Signal Transducer and Activator of Transcription 6 (*STAT6*), with a wildtype of T and mutant of C (Figure 12). Compelling evidence suggests that the A allele association with rs1059513 A correlates with allergic conditions, including asthma, atopic dermatitis, heightened IgE levels, and sensitization to common food and inhalant allergens (Ginkel et al., 2018). In a groundbreaking study, both A alleles of rs324015 and rs1059513 were found to correlate with food allergy and peanut allergy, IgE sensitization to peanut and cow's milk, and more severe allergic reactions. These findings suggest a potential role for *STAT6* genetic polymorphisms in the pathophysiology of food allergy, indicating their involvement independent of the specific allergenic food. STAT6 is a protein that plays a crucial role in transmitting signals from various cytokines, particularly interleukin-4 (IL- 4) and interleukin-13 (IL-13), in immune responses. This underscores the complexity of genetic factors in allergic diseases and emphasizes the importance of further research to elucidate the mechanisms underlying these associations and their clinical implications (Ginkel et al., 2018).

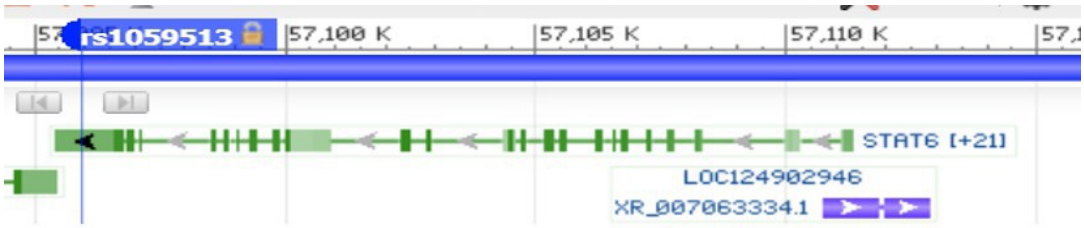


Figure 12: Position of rs1059513 on Non-Coding Region *STAT6* gene. The blue line shows the position of the variant in the 23rd exon of the *Non-Coding Region STAT6 gene.*

Hypothesis Statement and Aims

Through an exploration of genetic factors and specific SNP variants identified via 23andMe data and survey responses, this study aims to enhance our understanding of the interplay between genetics and immune-related conditions, specifically related to food allergies. By analyzing genetic data alongside survey responses, this research seeks to correlate food allergies and autoimmune diseases with various gene SNPs, shedding light on potential genetic predispositions. To test the hypothesis that one or more of the SNPs identified would be linked to food allergy in our dataset, a comprehensive study involving 85 participants was undertaken. Each of the 12 different SNPs from each participant was analyzed for the reference or variant alleles. Additionally, participant responses to three survey questions addressing personal medical histories, familial health conditions, and individual tolerance levels related to prevalent food allergies were used to correlate phenotype with genotype.

Methods

This study was conducted as a component of a broader investigation aimed at exploring the correlation between phenotypic outcomes and Single Nucleotide Polymorphisms (SNPs).

Specifically, the focus of this study was directed toward understanding the phenotypic outcomes associated with allergies and autoimmune diseases. The data for this investigation were gathered through the utilization of both 23andMe genetic testing and survey instruments.

Participants

The study sample comprised a previously gathered dataset from 85 undergraduate students between the age 18-22 enrolled in a large lecture class at a Research 1 public state university, who actively participating during the spring semesters of 2019 and 2020. 23andMe (23andMe, San Francisco, CA) genetic information about SNPs was obtained from each participant, alongside responses to three distinct survey questions. A total of 11 SNPs were collected from the participants, and subsequent analysis involved the examination of 33 unique outcomes resulting from the interaction between the genetic data and responses to the three survey questions. The questions pulled for data were:

- Do you have any of the following diseases (check all that apply: Type 1 Diabetes, Type 2 Diabetes, obesity, Celiac Disease, High Blood Pressure, Rheumatoid arthritis, and None)
- Does anyone in your family have any of the following diseases (check all that apply: Type 1 Diabetes, Type 2 Diabetes, obesity, Celiac Disease, High Blood Pressure, Rheumatoid arthritis, and None)
- I draw questions from two distinct participant groups, each pertaining to either

food allergies or autoimmune diseases. Subsequently, I merge these questions to generate comprehensive consumption data results. Participant C: Are you able to consume dairy products? Participant D: Or are you able to consume gluten without irritation?

All data were meticulously gathered in Excel spreadsheets, with each participant's information carefully organized. The total number of participants for each genotype category—whether homozygous for the reference allele, heterozygous, or homozygous for the variant allele—was then computed. To streamline the analysis, questions regarding milk and Celiac Disease were grouped under the category of "consumption," eliciting simple yes or no responses. These genotype-phenotype relationships were then collectively subjected to statistical analysis.

Data Analysis

The Fisher's test, commonly referred to as Fisher's exact test (https://www.danielsoper.com/statcalc/calculator.aspx?id=58), served as a statistical tool for assessing the presence of nonrandom associations between the two categorical variables. In this study, it was employed to categorize genotype and phenotype, enabling the calculation of the likelihood of a potential correlation to food allergies or autoimmune diseases based on both survey responses and nucleotide sequences. Significance was determined if the *P* value was equal or less than 0.05.

Results

The results of this study demonstrate the analysis of the associations between specific SNPs and phenotypic outcomes related to allergies and autoimmune diseases. Eleven variants were analyzed for genotype-phenotype association.

The SNP rs855791, located within the *TMPRSS6* gene, did not demonstrate any significant statistical correlations with either milk, gluten consumption, or disease presence, as evidenced by the data presented in Tables 1-3.

Table 1. Family Disease frequency and rs855791 on gene *TMPRSS6*.

Nucleotide Sequence	Yes	No
GG reference	11	18
AG	9	26
AA variant	3	4

*P=0.49 for an association of family disease and RS855791.

Table 2. Personal Disease frequency and rs855791 on gene *TMPRSS6*.

Nucleotide Sequence	Yes	No
GG reference	2	27
AG	3	40
AA variant	0	11

*P=1.0 for an association of personal disease and RS855791.

Table 3. Consumption frequency and rs855791 on gene *TMPRSS6*.

Nucleotide Sequence	Yes	No
GG reference	10	1
AG	36	7
AA variant	27	1

*P=0.24 for an association of consumption and RS855791.

Likewise, the SNP rs7454108, positioned between the genes *HLA-DQA1*, *HLA-DQB1*, and *HLA-DQB2*, did not exhibit any statistical associations with milk, gluten consumption, or disease presence. This lack of correlation was evident from the data presented in Tables 4-6.

Table 4. Family Disease frequency and rs7454108 between *HLA-DQA1*; *HLA-DQB1*; *HLA-DQB2*

Nucleotide Sequence	Yes	No
TT reference	28	37
CT	5	8
CC variant	3	1

*P=0.51 for an association of family disease and RS7454108.

Table 5. Personal Disease frequency and rs7454108 between *HLA-DQA1*; *HLA- DQB1*; *HLA-DQB2*.

Nucleotide Sequence	Yes	No
TT reference	3	64
CT	1	12
CC variant	1	3

*P=0.14 for an association of personal disease and RS7454108.

Table 6. Consumption frequency and rs7454108 between *HLA-DQA1*; *HLA-DQB1*; *HLA-DQB2*.

Nucleotide Sequence	Yes	No
TT reference	59	8
CT	11	2
CC variant	4	0

*P=0.80 for an association of consumption and RS7454108.

Similarly, the SNP rs7192, located within the *HLA-DRA* gene, did not show any statistical associations with milk or gluten consumption. This observation is supported by the data presented in Tables 7-9.

Table 7. Family Disease frequency and rs7192 on *HLA-DRA*.

Nucleotide Sequence	Yes	No
TT reference	5	13
GT	13	15
GG variant	19	18

*P=0.25 for an association of family disease and RS719.

Table 8. Personal Disease frequency and rs7192 on *HLA-DRA*.

Nucleotide Sequence	Yes	No
TT reference	1	17
GT	3	26
GG variant	1	36

*P=0.51 for an association of personal disease and RS7192.

Table 9. Consumption frequency and rs7192 on *HLA-DRA*.

Nucleotide Sequence	Yes	No
TT reference	15	3
GT	26	2
GG variant	32	5

*P=0.63 for an association of consumption and RS7192.

However, when looking at the SNP rs2187668, situated on the *HLA-DQA1* gene, a statistically significant association with personal diseases ($P < 0.05$) was found, as indicated in Table 11, especially among individuals with the reference homozygous CC genotype. However, family disease and food consumption did not demonstrate any statistical associations, as illustrated in Table 10 and Table 12.

Table 10. Family Disease frequency and rs2187668 on *HLA-DQA1*.

Nucleotide Sequence	Yes	No
CC reference	28	17
CT	7	14
TT variant	1	1

*P=0.29 for an association of family disease and RS2187668.

Table 11. Personal Disease frequency and rs2187668 on *HLA-DQA1*.

Nucleotide Sequence	Yes	No
CC reference	2	59
CT	2	19
TT variant	1	1

*P=0.04 for an association of personal disease and RS2187668.

Table 12. Consumption frequency and rs2187668 on *HLA-DQA1*.

Nucleotide Sequence	Yes	No
CC reference	54	7
CT	19	2
TT variant	1	1

*P=0.31 for an association of consumption and RS2187668.

The SNP rs41380347, located within the *MCM6* gene, similarly did not exhibit statistical associations with milk, gluten consumption, or disease presence. This finding is supported by the data presented in Tables 13-15.

Table 13. Family Disease frequency and rs41380347 on *MCM6*.

Nucleotide Sequence	Yes	No
AA reference	35	47
AG	1	0
GG variant	0	0

*P=0.43 for an association of family disease and RS41380347.

Table 14. Personal Disease frequency and rs41380347 on *MCM6*.

Nucleotide Sequence	Yes	No
AA reference	5	79
AG	0	1
GG variant	0	0

*P=1.0 for an association of personal disease and RS41380347.

Table 15. Consumption frequency and rs41380347 on *MCM6*.

Nucleotide Sequence	Yes	No
GG reference	6	10
AG	20	22
AA variant	10	15

*P=0.73 for an association of consumption and RS41380347.

Also, on the *MCM6* gene is the rs4988235 SNP, which did exhibit a statistically significant association with personal diseases ($P < 0.05$), particularly among individuals with the heterozygous AG genotype as highlighted in Table 17. However, no statistical associations were observed between family disease and food consumption, as depicted in Table 16 and Table 18.

Table 16. Family Disease frequency and rs4988235 on *MCM6*.

Nucleotide Sequence	Yes	No
GG reference	12	1
AG	2	41
AA variant	2	23

*P=0.43 for an association of family disease and RS4988235

Table 17. Personal Disease frequency and rs4988235 on *MCM6*.

Nucleotide Sequence	Yes	No
GG reference	13	3
AG	37	6
AA variant	24	1

*P=0.0 for an association of personal disease and RS4988235.

Table 18. Consumption frequency and rs4988235 on *MCM6*.

Nucleotide Sequence	Yes	No
GG reference	13	3
AG	37	6
AA variant	24	1

*P=0.33 for an association of consumption and RS4988235.

The SNP rs7216389, located within the *GSDMB* gene, likewise did not demonstrate statistical associations with milk, gluten consumption, or disease presence. This conclusion is reinforced by the data provided in Tables 19-21**.**

Table 19. Family Disease frequency and rs7216389 on *GSDMB*.

Nucleotide Sequence	Yes	No
CC reference	7	15
CT	18	19
TT variant	11	13

*P=0.46 for an association of family disease and RS7216389

Table 20. Personal Disease frequency and rs7216389 on *GSDMB*.

Nucleotide Sequence	Yes	No
CC reference	1	21
CT	4	34
TT variant	0	24

*P=0.26 for an association of personal disease and RS7216389.

Table 21. Consumption frequency and rs7216389 on *GSDMB*.

Nucleotide Sequence	Yes	No
CC reference	21	1
CT	33	5
TT variant	20	4

*P=0.43 for an association of consumption and RS7216389.

The SNP rs12964116, situated on the *SERPINB7* gene, also did not show statistical associations with milk, gluten consumption, or disease presence. This conclusion is supported by the data presented in Tables 22-24.

Table 22. Family Disease frequency and rs12964116 on *SERPINB7*.

Nucleotide Sequence	Yes	No
AA reference	32	44
AG	2	4
GG variant	1	0

*P=0.65 for an association of family disease and RS12664116

Table 23. Personal Disease frequency and rs12964116 on *SERPINB7*.

Nucleotide Sequence	Yes	No
AA reference	5	71
AG	0	6
GG variant	0	1

*P=1.0 for an association of personal disease and RS12664116.

Table 24. Consumption frequency and rs12964116 on *SERPINB7*.

Nucleotide Sequence	Yes	No
AA reference	66	10
AG	6	0
GG variant	1	0

*P=1.0 for an association of consumption and RS12664116.

Another SNP to show a significant correlation was the rs1861760 SNP, located between the genes *LOC105371251* and *LINC02128*, which showed a statistically significant association with consumption frequency ($P < 0.05$), especially among individuals with the mutant homozygous CC genotype, as indicated in Table 27. However, no statistical associations were observed between family disease and personal disease, as illustrated in Table 25 and Table 26.

Table 25. Family Disease frequency and rs1861760 between *LOC105371251 and LINC02128*.

Nucleotide Sequence	Yes	No
AA reference	0	0
AC	3	2
CC variant	32	45

*P=0.64 for an association of family disease and RS1861760

Table 26. Personal Disease frequency and rs1861760 between *LOC105371251* and LINC02128.

Nucleotide Sequence	Yes	No
AA reference	0	0
AC	0	5
CC variant	5	72

*P=1.0 for an association of personal disease and RS1861760

Table 27. Consumption frequency and rs1861760 between *LOC105371251* and LINC02128.

Nucleotide Sequence	Yes	No
AA reference	0	0
AC	4	1
CC variant	10	68

*P=0.0 for an association of consumption and RS1861760.

Also significant was the rs9272346 SNP, located on the *HLA-DQA* gene, which exhibited a statistically significant association with consumption frequency ($P < 0.05$), particularly among individuals with the heterozygous AG genotype as demonstrated in Table 30. However, no statistical associations were observed between family disease and personal disease, as depicted in Tables 28 and 29.

Table 28. Family Disease frequency and rs9272346 on *HLA-DQA1*.

Nucleotide Sequence	Yes	No
GG reference	8	20
AG	20	20
AA variant	5	7

*P=0.21 for an association of family disease and RS9272346

Table 29. Personal Disease frequency and rs9272346 on *HLA-DQA1*.

Nucleotide Sequence	Yes	No
GG reference	8	24
AG	20	40
AA variant	5	11

*P=0.71 for an association of personal disease and RS9272346.

Table 30. Consumption frequency and rs9272346 on *HLA-DQA*.

Nucleotide Sequence	Yes	No
GG reference	0	2
AG	34	6
AA variant	11	1

*P=0.03 for an association of consumption and RS9272346.

Likewise, the rs1059513 SNP, located on the *STAT6* gene, showed a statistically significant association with family disease ($P < 0.05$), especially among individuals with the variant homozygous CC genotype as demonstrated in Table 31. However, no statistical associations were observed between family disease and consumption frequency, as depicted in Table 32 and Table 33.

Table 31. Family Disease frequency and rs1059513 on *STAT6*.

Nucleotide Sequence	Yes	No
TT reference	0	0
CT	17	7
CC variant	11	41

*P=0.0 for an association of family disease and RS1059513

Table 32. Personal Disease frequency and rs1059513 on *STAT6*.

Nucleotide Sequence	Yes	No
TT reference	0	0
CT	0	4
CC variant	5	64

*P=0.58 for an association of personal disease and RS1059513

Table 33. Consumption frequency and rs1059513 on *STAT6*.

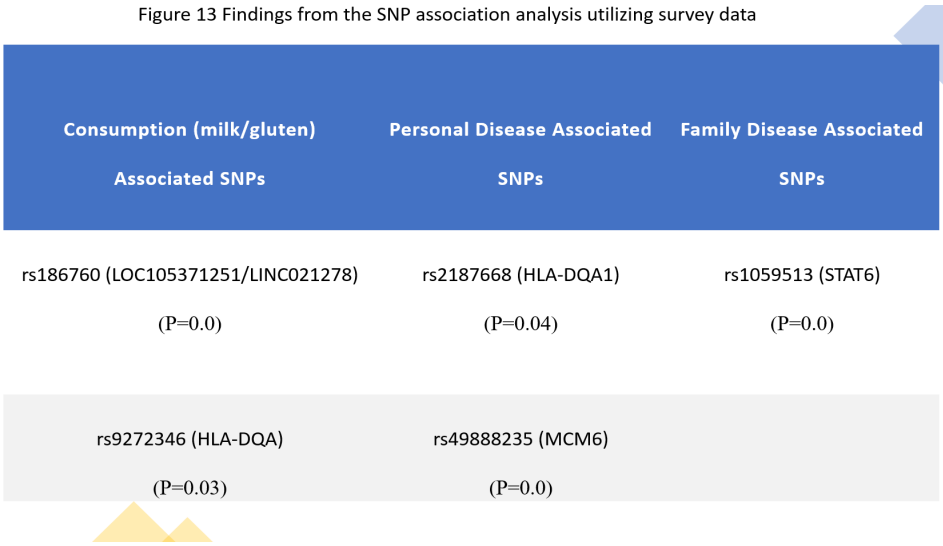
Nucleotide Sequence	Yes	No
TT reference	0	0
CT	12	2
CC variant	61	8

*P=0.67 for an association of consumption and RS1059513.

Discussion

The findings of this study provide valuable insights into the associations between specific SNPs and phenotypic outcomes related to allergies and autoimmune diseases. Despite an extensive analysis, several genotypes and phenotypes did not show statistical associations. However, statistical significance was seen with the rs2187668 on the *HLA-DQA1* gene and the rs49888235 on the *MCM6* gene and personal disease; as well as the rs1861760 between *LOC105371251* and *LINC021278* genes and the rs9272346 *HLA-DQA* gene and consumption frequency; and lastly between the rs1059513 on the *STAT6* gene and family disease, see Figure 13.

Figure 13 Findings from the SNP association analysis utilizing survey data



The intricate relationship between genetic predispositions and disease outcomes is highlighted by the observed association between rs2187668 and lack personal diseases in our study. The significance of rs2187668 in the context of food allergies, including Celiac Disease, underscores its importance as a potential biomarker for disease risk. This genetic variant, situated on chromosome 6 at position 32411646, represents a focal point for further investigation into the genetic underpinnings of food-related immune responses. In a broader context, our findings align with previous research linking rs2187668 to autoimmune

diseases. A meta-analysis exploring the association between *HLA-DQA1* rs2187668 polymorphism and systemic lupus erythematosus (SLE) revealed a significant correlation across diverse populations. Specifically, individuals carrying the GG genotype and G allele exhibited an elevated risk of developing SLE (Li et al., 2018). This suggests a potential role for rs2187668 in influencing susceptibility to autoimmune disorders. The observed association between rs2187668 and personal diseases in our study underscores the complex interplay between genetic predispositions and disease outcomes. Further research is suggested to understand the molecular mechanisms underlying this association and to explore potential therapeutic targets for individuals at heightened risk of developing food allergies and autoimmune diseases. Our study highlights the significance of the *HLA-DQA1* rs2187668 polymorphism in the context of personal disease susceptibility, particularly in the realm of food allergies and autoimmune disorders.

The SNP rs4988235, located on chromosome 2 at position 136608646 on the Minichromosome Maintenance Complex Component 6 (*MCM6*) gene, plays a pivotal role in regulating lactase enzyme production. Positioned in the regulatory region upstream of the lactase gene (LCT), rs4988235 influences lactase enzyme activity, thereby impacting an individual's ability to digest lactose beyond infancy. The analysis of personal disease frequency concerning the rs4988235 SNP on the MCM6 gene revealed notable associations. Specifically, individuals with the AG genotype exhibited a higher frequency of personal disease compared to those with the GG reference genotype. Furthermore, individuals homozygous for the AA variant genotype showed the highest frequency of personal disease among the three genotypes. These findings indicate a statistically significant association between personal disease and rs4988235, with a *P*- value of 0.0. Individuals carrying the T allele of rs4988235 demonstrate heightened lactase enzyme

activity, allowing for continued lactose digestion into adulthood. This persistence of lactase activity characterizes lactase-persistent (LP) individuals, distinguishing them from lactase non-persistent (LNP) individuals who experience lactose intolerance symptoms upon lactose consumption. Lactase-persistent (LP) individuals are people who continue to produce lactase enzyme into adulthood. This persistence allows them to digest lactose, the sugar found in dairy products, without experiencing symptoms of lactose intolerance. Lactase non-persistent (LNP) individuals are individuals have reduced lactase enzyme activity as they age, leading to symptoms of lactose intolerance, such as bloating, gas, and diarrhea, when they consume lactose-containing foods. Sivapalan et al. (2020) classified individuals with CT or TT genotypes as LP, while those with the CC genotype are considered LNP or genetically lactose intolerant. Furthermore, the autosomal nature of the 'T' variant of rs4988235 indicates Mendelian inheritance patterns, independent of sex chromosomes. The clinical significance of understanding the genetic basis of lactase persistence, particularly the role of rs4988235, is profound. Genetic testing for rs4988235 offers valuable insights into an individual's predisposition to lactase persistence or non- persistence, guiding dietary recommendations, and management strategies for lactose intolerance. Additionally, knowledge of the inheritance pattern of rs4988235 facilitates familial risk assessment and genetic counseling for individuals with lactose intolerance or lactase persistence. Despite exhibiting a statistically significant association with personal diseases, particularly among individuals with the mutant heterozygous AG genotype and AA homozygous genotype, no statistical associations were observed between family disease and food consumption in our study. This underscores the complex interplay between genetic factors and disease susceptibility, emphasizing the need for further research.

The investigation into the genetic variant rs1861760, between LOC105371251 and

LINC02128 genes, reveals a noteworthy gap in research concerning its association with food allergies and autoimmune diseases. Despite limited prior exploration, our findings demonstrate a statistically significant association with the frequency of milk or gluten consumption ($P < 0.05$). The examination of consumption frequency in relation to the rs1861760 SNP, located between LOC105371251 and LINC02128, revealed significant associations. Specifically, individuals carrying the CC variant genotype exhibited a significantly higher frequency of not being able t consume either dairy or gluten. No individuals with the AA reference genotype were observed in the analysis. These findings underscore a statistically significant association between consumption frequency and rs1861760, with a p-value of 0.0. This suggests a potential role for the rs1861760 SNP in influencing consumption behaviors, highlighting the need for further investigation into its mechanistic implications. This observation suggests a potential role for rs1861760 in modulating dietary responses and immune function. Interestingly, while no statistical associations were observed between family disease and personal disease in our study, existing literature sheds light on potential mechanisms underlying the influence of rs1861760 on allergic and autoimmune responses. Studies have highlighted its association with FOXJ1 and Th2-type responses in allergic rhinitis suggests a potential role for FOXJ1 in modulating the immune response underlying allergic reactions, possibly by influencing the expression of genes involved in Th2-type cytokine production or other aspects of allergic inflammation. Furthermore, variants within the genomic region encompassing rs1861760 have been linked to inflammatory bowel diseases, including Crohn's disease, and even leprosy, underscoring its multifaceted role in immune modulation and susceptibility to diverse pathological conditions (Sarnowski et al., 2016). These insights broaden our understanding of rs1861760's impact beyond its association with food

allergies and autoimmune diseases. They highlight its potential as a pivotal genetic marker influencing immune function and susceptibility to various pathological conditions. Despite the need for further, our study contributes valuable evidence towards unraveling the complex interplay between genetic variants, dietary factors, and immune responses in the context of allergic and autoimmune diseases.

The SNP rs9272346, located on the *HLA-DQA* gene, exhibited a statistically significant association with consumption frequency ($P < 0.05$), particularly among individuals with the mutant heterozygous AG genotype and AA variant (giving a possible protective effect) in our study. However, no statistical associations were found between family disease and personal disease, indicating a nuanced relationship between this genetic variant and health outcomes. Notably, limited research has been conducted to explore the potential associations of rs9272346 with food allergies and autoimmune diseases, leaving a notable gap in our understanding of its broader implications. Although existing literature links rs9272346 to asthma and allergic diseases, particularly absent are studies associating this SNP specifically with food allergies or autoimmune conditions (Lasky-Su et al., 2012). This gap underscores the need for further investigation into the role of rs9272346 in immune-mediated disorders beyond asthma. Despite its established significance in asthma susceptibility, the potential involvement of rs9272346 in other immune-related conditions remains unexplored.

In addition to rs9272346 on the HLA-DQA gene, another SNP of interest in our study is rs1059513 located on the STAT6 gene. Rs1059513 demonstrated a statistically significant association with an increase in family disease, highlighting its potential relevance to health outcomes within familial contexts. While rs1059513 has been extensively studied in the context of asthma and allergic diseases (Ginkel et al., 2018), its

specific associations with food allergies and autoimmune conditions remain largely unexplored. The limited research on rs1059513 underscores the need for further investigation into its broader implications in immune-mediated disorders beyond asthma. Despite its established significance in asthma susceptibility, the potential involvement of rs1059513 in other immune-related conditions presents an intriguing avenue for future research efforts.

In conclusion, while our study did not find significant associations for the SNPs rs41380347, rs7216389, rs12964116, rs855791, rs7454108, and rs7192 with food allergies and autoimmune diseases, further research is warranted to explore their potential roles in disease susceptibility. Future studies with larger sample sizes, diverse populations, and comprehensive phenotypic characterization are needed to elucidate the complex genetic architecture of these conditions and identify novel genetic markers for risk assessment and therapeutic targeting.

Limitations

The present study is not immune to certain limitations that warrant consideration. Firstly, reliance on self-reporting for data collection may introduce inherent biases, including recall bias and social desirability bias, potentially impacting the accuracy and reliability of the gathered information. Additionally, volunteers may have had trouble in recalling or accurately reporting the numbers provided to them for the survey, leading to potential inconsistencies or errors in data entry. Moreover, while efforts were made to control for environmental factors that could influence the outcomes under investigation, such as variations in geographical location or exposure to allergens, it is acknowledged that the complexity and multitude of environmental variables make it challenging to comprehensively account for their effects. These limitations underscore the need for a

cautious interpretation of the findings and highlight avenues for future research to employ more rigorous methodologies, such as objective measures of data collection and comprehensive environmental assessments, to mitigate these potential sources of bias and enhance the robustness of conclusions drawn. Addition, the phenotype questions encompassed conditions beyond autoimmune diseases, such as obesity and hypertension. Additionally, it's essential to recognize that aversion to milk and gluten consumption does not necessarily indicate the presence of milk allergy or Celiac Disease. These factors may influence the interpretation of our findings and should be taken into consideration when evaluating the results. Next instead of using generic answers, researchers could customize responses to better capture the nuances of autoimmune diseases. This could involve providing a range of options that reflect the diverse manifestations and severity of these conditions.

Dissemination

The findings from this study shed light on the complex relationship between genetic markers and phenotypic outcomes associated with food allergies and autoimmune diseases. With millions of Americans impacted by these conditions, understanding the underlying genetic factors is crucial for personalized interventions and improved management strategies. By analyzing Single Nucleotide Polymorphisms (SNPs) using data from the genetic testing service 23andMe, this research uncovers how individual genetic makeup may influence responses to food and susceptibility to allergies and autoimmune disorders. Significant associations were identified between specific SNPs and disease outcomes, offering promising avenues for tailored interventions. This research not only deepens our understanding of personalized nutrition but also holds potential for the development of novel approaches to disease prevention and management, ultimately contributing to

improved health outcomes for individuals affected by these conditions. In future research, I intend to develop and implement customized phenotype questions and answers tailored specifically to the scope of autoimmune diseases and related conditions. This approach will ensure a more focused and comprehensive understanding of the factors under investigation. Additionally, I plan to select a specific genetic variant (rs) associated with autoimmune diseases for in-depth analysis, allowing for a deeper exploration of its implications. Furthermore, I aim to delve into a particular topic within autoimmune diseases, such as milk allergy, to elucidate its mechanisms, risk factors, and potential interventions, thereby contributing to the advancement of knowledge in this field.

References

21 Common Signs of Gluten Intolerance. (2022, June 3). Healthline. https://www.healthline.com/nutrition/signs-you-are-gluten-intolerant#gluten-sensitivity- symptoms

Abadie, V., Kim, S.M., Lejeune, T. et al. IL-15, gluten and HLA-DQ8 drive tissue destruction in coeliac disease. Nature 578, 600–604 (2020). https://doi-org.ezproxy.lib.vt.edu/10.1038/s41586-020-2003-8

Barker, J. M., Triolo, T. M., Aly, T. A., Baschal, E. E., Babu, S. R., Kretowski, A., Rewers, M. J., & Eisenbarth, G. S. (2008). Two Single Nucleotide Polymorphisms Identify the Highest-Risk Diabetes HLA Genotype: Potential for Rapid Screening. Diabetes, 57(11), 3152–3155. https://doi.org/10.2337/db08-0605

Beyond Celiac. (2019). Beyond Celiac. https://www.beyondceliac.org/

Chang Seok Oh, Myeung Ju Kim, Yi-Suk Kim, Sori Min, Kyong Taek Oh, Soong Deok Lee, & Dong Hoon Shin. (2023). Revealing Joseon period People's single nucleotide polymorphism associated with lactase gene by ancient DNA analysis of human remains from archaeological sites in Korea. Anatomy & Cell Biology, 56(1), 54–60. https://doi.org/10.5115/acb.22.178

Celiac Disease: Symptoms, Testing, Treatment & Research. (n.d.). Beyond Celiac. Retrieved April 13, 2024, from https://www.beyondceliac.org/celiac-disease/?gad_source=1&gclid=Cj0KCQjw2uiwBhCXARIsACMvIU1iXeiX8UHPKcU6T TBer6nUB8U93lwp4udH-rpNfzHnhiknHy2zJwcaAs0GEALw_wcB

Deng N., Zhou H., Fan H., Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget. 2017; 8: 110635-110649. Retrieved from https://www.oncotarget.com/article/22372/text/

Dermatitis Herpetiformis (DH): Definition, Causes & Treatment. (n.d.). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/21460-dermatitis-herpetiformis

FARE. (2022). Facts and Statistics. Food Allergy Research & Education. https://www.foodallergy.org/resources/facts-and-statistics

Food Allergies | Food Safety and Inspection Service. (2012). Usda.gov. https://www.fsis.usda.gov/food-safety/safe-food-handling-and-preparation/food-safety- basics/food-allergies

Food Allergens - The Big 8 is Now The Big 9. (n.d.). Eurofins Scientific. Retrieved April 8, 2024, from https://www.eurofinsus.com/food-testing/resources/food-allergens-the-big-8- is- now-the-big-9/

Ginkel, C. D., Pettersson, M. E., Dubois, A. E. J., & Koppelman, G. H. (2018). Association

of STAT6 gene variants with food allergy diagnosed by double-blind placebo-controlled food challenges. Allergy, 73(6), 1337–1341. https://doi.org/10.1111/all.13432

Gluten Ataxia and Celiac Disease | BeyondCeliac.org. (n.d.). Beyond Celiac. Retrieved April 15, 2024, from https://www.beyondceliac.org/celiac-disease/related-conditions/gluten- ataxia/#:~:text=Gut%2DBrain%20Connection%3A%20What%20is

Harrison, L. C., & Honeyman, M. C. (1999). Cow's milk and Type 1 Diabetes: the real debate is about mucosal immune function. Diabetes, 48(8), 1501–1507. https://doi.org/10.2337/diabetes.48.8.1501

Hong, X., Hao, K., Ladd-Acosta, C., Hansen, K. D., Tsai, H.-J., Liu, X., Xu, X., Thornton, T. A., Caruso, D., Keet, C. A., Sun, Y., Wang, G., Luo, W., Kumar, R., Fuleihan, R., Singh, A. M., Kim, J. S., Story, R. E., Gupta, R. S., & Gao, P. (2015). Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nature Communications, 6(1). https://doi.org/10.1038/ncomms7304

Hong, X., Tsai, H.-J., & Wang, X. (2009). Genetics of food allergy. Current Opinion in Pediatrics, 21(6), 770–776. https://doi.org/10.1097/mop.0b013e32833252dc

Krisnawati, D. I., Alimansur, M., Atmojo, D. S., Rahmawati, E. Q., Rahayu, D., Susilowati, E., & Kuo, T.-R. (2022). Food Allergies: Immunosensors and Management. Applied Sciences, 12(5), 2393. https://doi.org/10.3390/app12052393

Lasky-Su, J., Himes, B. E., Raby, B. A., Klanderman, B. J., Sylvia, J. S., Lange, C., Melen, E., Martinez, F. D., Israel, E., Gauderman, J., Gilliland, F., Sleiman, P., Hakonarson, H., Celedón,
J. C., Soto-Quiros, M., Avila, L., Lima, J. J., Irvin, C. G., Peters, S. P., & Boushey, H. (2012). HLA-DQ strikes again: Genome-wide association study further confirmsHLA-DQin the diagnosis of asthma among adults. Clinical & Experimental Allergy, 42(12), 1724–1733. https://doi.org/10.1111/cea.12000

Li, S., Wu, W., & Jin, Z. (2018). Association between the HLA-DQA1 rs2187668 polymorphism and systemic lupus erythematosus: A meta-analysis. Medicine, 97(45), e13043. https://doi.org/10.1097/MD.0000000000013043

L. M., Karunas, A. S., James, A., Probst-Hensch, N. M., von Mutius, E., & Pin, I. (2016). Identification of a new locus at 16q12 associated with time to asthma onset. Journal of Allergy and Clinical Immunology, 138(4), 1071–1080. https://doi.org/10.1016/j.jaci.2016.03.018

Marenholz, I., Grosche, S., Kalb, B., Rüschendorf, F., Blümchen, K., Schlags, R.,

Harandi, N., Price, M., Hansen, G., Seidenberg, J., Röblitz, H., Yürek, S., Tschirner, S., Hong, X., Wang, X., Homuth, G., Schmidt, C. O., Nöthen, M. M., Hübner, N., & Niggemann, B. (2017). Genome-wide association study identifies the SERPINB gene cluster as a susceptibility locus for food allergy. Nature Communications, 8(1). https://doi.org/10.1038/s41467-017-01220-0

Gupta, R. S., Kim, J. S., Springston, E. E., Pongracic, J. A., Wang, X., & Holl, J. (2009). Development of the Chicago Food Allergy Research Surveys: assessing knowledge, attitudes, and beliefs of parents, physicians, and the general public. BMC Health Services Research, 9(1). https://doi.org/10.1186/1472-6963-9-142

Mayo Clinic. (2021, December 31). Food allergy - Symptoms and causes. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/food-allergy/symptoms-causes/syc-20355095

Mayo Clinic. (2018). Milk Allergy - Symptoms and Causes. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/milk-allergy/symptoms-causes/syc- 20375101

Monsuur, A. J., de Bakker, P. I. W., Zhernakova, A., Pinto, D., Verduijn, W., Romanos, J., Auricchio, R., Lopez, A., van Heel, D. A., Crusius, J. B. A., & Wijmenga, C. (2008). Effective Detection of Human Leukocyte Antigen Risk Alleles in Celiac Disease Using Tag Single Nucleotide Polymorphisms. PLoS ONE, 3(5), e2270. https://doi.org/10.1371/journal.pone.0002270

Ranjbar, M., Whetstone, C. E., Omer, H., Power, L., Cusack, R. P., & Gauvreau, G. M. (2022). The Genetic Factors of the Airway Epithelium Associated with the Pathology of Asthma. Genes, 13(10), 1870. https://doi.org/10.3390/genes13101870

Ristic , A. (2020, August 8). A Crucial Genetic Factor for Peanut Allergy (HLA). Selfdecode. https://selfdecode.com/app/article/hla-peanut-allergy/#your-hla-results-for-peanut-allergy

Richards, E. P. (1995). *Introduction*. Introduction: Primer on Molecular Genetics. https://biotech.law.lsu.edu/blaw/hgp/doe1/prim1.html

Sarnowski, C., Sugier, P.-E., Granell, R., Jarvis, D., Dizier, M.-H., Ege, M., Imboden, M., Laprise, C., Khusnutdinova, E. K., Freidin, M. B., Cookson, W. O. C., Moffatt, M., Lathrop, M., Siroux, V., Ogorodova

Sebastian-delaCruz, M., & Castellanos-Rubio, A. (2023, January 1). Chapter 16 - A fast, cheap, and easy protocol for Celiac Disease HLA haplotype typing using buccal swabs (A. Castellanos- Rubio & L. Galluzzi, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/pii/S0091679X22001625

Sivapalan, H. (2020, September 20). Trait: Lactose Tolerance [Review of Trait: Lactose Tolerance]. Fitness Genes. https://www.fitnessgenes.com/blog/lactose-tolerance

Urbaszek, K., Drabińska, N., Szaflarska-Popławska, A., & Jarocka-Cyrta, E. (2021). TMPRSS6 rs855791 Polymorphism Status in Children with Celiac Disease and Anemia. Nutrients, 13(8), 2782. https://doi.org/10.3390/nu13082782

van Heel, D. A., Franke, L., Hunt, K. A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M. C., Barnardo, M. C. N. M., Bethel, G., Holmes, G. K. T., Feighery, C., Jewell, D., Kelleher, D., Kumar, P., Travis, S., Walters, J. R., Sanders, D. S., Howdle, P., Swift, J., & Playford, R. J. (2007). A genome-wide association study for Celiac Disease identifies risk variants in the region harboring IL2 and IL21. Nature Genetics, 39(7), 827–829. https://doi.org/10.1038/ng2058

Wiklund, J. (2021, December 29). *The Big 9 Food allergens*. The Food Institute. https://foodinstitute.com/focus/the-big-9-food-allergens/

Orbai, A.-M. (2019). What Are Common Symptoms of Autoimmune Disease? Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/wellness-and-prevention/what- are- common-symptoms-of-autoimmune-disease

Zack, D. E., Stern, D. A., Willis, A. L., Kim, A. S., Mansfield, C. J., Reed, D. R., Brooks, S. G., Adappa, N. D., Palmer, J. N., Cohen, N. A., Chiu, A. G., Song, B. H., Le, C. H., & Chang, E. H. (2021). The GSDMB rs7216389 SNP is associated with chronic rhinosinusitis in a multi-institutional cohort. *International Forum of Allergy & Rhinology*, *11*(12), 1647– 1653. https://doi.org/10.1002/alr.22824