From Genotype to Phenotype: Identification of HLA-DQA1 Role in Celiac Disease

Federica Bertolotti
Lake Forest College
Lake Forest, Illinois 60045

AI Disclaimer: I primarily used AI to structure my ideas into paragraphs and improve the general organization. I used AI to make grammatical corrections, such as changing passive voice to active voice and proofreading. Secondly, I used AI to clarify some of the primary articles, especially to simplify the methods section for techniques I did not know. This use of AI to clarify also included some papers in the future experiments section, but I created my experiments drawing from all articles and my lab experience to create truthful and scientifically accurate experiments. Aside from these specific uses, I read and created all the arguments presented in this paper without the help of AI.

I. Abstract:

Celiac disease is an autoimmune disorder characterized by an adverse reaction to the protein gluten, leading to inflammation and damage to the small intestine. Central to this investigation is the association between celiac disease and the *HLA-DQA1* gene, a key player in disease susceptibility. While existing research has established a link between *HLA-DQA1* and celiac disease, my analysis reveals a critical gap in knowledge regarding the specific alleles and their mechanistic pathways underlying their association. Through a comprehensive review of literature, including studies on gene polymorphisms and animal models, this study outlines an experimental approach to elucidate the functional significance of *HLA-DQA1* alleles in celiac disease pathogenesis. By utilizing CRISPR/Cas9 technology to generate knock-in mice expressing disease-associated alleles, I aim to investigate immune responses to gluten and assess tissue damage in the intestinal mucosa of mice. The proposed experiments encompass histological analysis, cytokine production assays, and T-cell activation studies to unravel the mechanistic underpinnings of celiac disease development.

II. The Phenotype

The gene I'm focusing on in this paper is *HDQ-DA1*, particularly due to its association with celiac disease. About a year ago, I underwent blood tests that revealed I have a gluten sensitivity. This sensitivity is shared by other members of my family, prompting me to wonder about the genetic factors underlying the connection between genes, gluten intolerance, and celiac disease. As I delved into research on gluten intolerance and celiac disease, *HDQ-DA1* emerged as a crucial factor. Moreover, I gained insight into the differentiation between celiac disease and gluten sensitivity, often referred to as non-celiac sensitivity.

The decision to center this paper on celiac disease stemmed from two primary reasons. Firstly, there is a significantly greater amount of genetic testing conducted on celiac disease compared to gluten sensitivity, which remains incompletely understood genetically. Secondly,

Italy has a notably high prevalence of celiac disease with rates at 1.6% compared to the global average of 1% ("Italian Screening Study Finds Increasing Prevalence of Celiac Disease, 2021). Given my Italian heritage and the prevalence of the disease within my family, my interest in studying celiac disease grew exponentially. Therefore, I decided to investigate the HDQ-DA1 gene and its alleles.

First, a distinction must be made between celiac disease and gluten sensitivity. Both are a body response to the protein gluten that is found in wheat, rye, and barley. Gluten is made of two different proteins: gliadin and glutenin ("Celiac disease", 2023). Celiac disease is a multifaceted autoimmune disorder (meaning the body's immune system turns against its own tissues.) That is to say, when someone with celiac consumes the gluten protein, an autoimmune reaction is triggered that damages the small intestine's lining and impairs nutrient absorption, which leads to various health complications (Devriese, et al., 2024).

In contrast, gluten intolerance, or non-celiac gluten sensitivity, offers a less intense (although still challenging) condition. While symptoms like abdominal pain, bloating, and diarrhea may occur with gluten intolerance, there's no autoimmune intestinal damage involved ("Gluten Intolerance", 2021). Despite my family's experience with gluten insensitivity, I'm drawn to understanding celiac disease, recognizing its severity as more than just discomfort but a profound health concern requiring deeper exploration and awareness. Therefore, I will focus on celiac disease for my research paper.

In individuals with celiac disease, gluten's gliadin peptides are what trigger the autoimmune response, meaning they recognized as foreign invaders by antigen-presenting cells (APCs; Sciurti, et al., 2019). APCs process the gliadin peptides and present them to T-cells (immune cells), specifically CD4+ T-cells, via major histocompatibility complex class II (MHC

II) molecules (Sciurti, et al., 2019). This interaction activates the CD4+ T-cells, leading to the release of pro-inflammatory cytokines, such as interferon-gamma (IFN-γ) and tumor necrosis factor-alpha (TNF-α).

These pro-inflammatory cytokines initiate an immune cascade, resulting in the recruitment of more immune cells, including lymphocytes and macrophages, that go to the intestinal mucosa. This influx of immune cells contributes to inflammation and tissue damage in the small intestine, particularly in the villi, which are finger-like projections responsible for nutrient absorption (Besser, et al., 2023).

Furthermore, the activation of tissue transglutaminase (TG2), an enzyme in the intestinal mucosa, plays a crucial role in the pathogenesis of celiac disease (Besser, et al., 2023). TG2 modifies gliadin peptides, increasing their immunogenicity and enhancing their binding to MHC II molecules on APCs. This process further amplifies the immune response and perpetuates tissue damage in the gut lining.

Ultimately, the chronic inflammation and tissue damage caused by this immune response leads to villous atrophy, crypt hyperplasia, and flattening of the intestinal mucosa. Villous atrophy is the damage to the finger-like projections (villi) lining the small intestine, and is a hallmark of celiac disease, leading to villi erosion or loss (Schiepatti et al., 2022). Crypt hyperplasia is the increased proliferation and enlargement of the intestinal crypts, which are invaginations between the villi, often seen in response to mucosal injury or inflammation

(Villanacci et al., 2020). Lastly, the flattening of the intestinal mucosa is as it sounds, a reduction of the characteristic folds (rugae) and protrusions (villi) in the lining of the intestine (Villanacci et al., 2020). All these characteristics combined produce the resulting and typically indicative of mucosal damage that impairs nutrient absorption and can manifest clinically as malabsorption, gastrointestinal symptoms, and systemic complications.

III. The Genotype

Further studies of celiac disease identified *HDQ-DA1* playing a crucial role in the development of the disease (Devriese, et al., 2024). It is involved in the regulation of the immune response to gluten, with variations in the gene contributing to an increased risk of developing the condition. Research into *HDQ-DA1* revealed insights into the mechanisms underlying celiac disease development (Devriese, et al., 2024). Certain variants of *HDQ-DA1* have been linked to heightened immune reactivity to gluten peptides, leading to the characteristic inflammatory response seen in individuals with celiac disease (Fernández-Mestre, et al., 2023). Understanding the role of *HDQ-DA1* in this process is essential for developing targeted therapies and interventions for celiac disease management.

*This author wrote this paper for Neuroscience 352: From Genotype to Phenotype taught by Dr. Karen Kirk.

The gene *HDQ-DA1*, also known as *HLA-DQ1*, is a key player in the pathogenesis of celiac disease. *HLA-DQ1* is found in chromosome 6p21.32 and belongs to the human leukocyte antigen (HLA) gene complex, which encodes major histocompatibility complex (MHC) class II molecules (Fernández-Mestre, et al., 2023). More specifically, the gene comprises 5 exons: exon 1 codes for the leader peptide, exons 2 and 3 code for the two extracellular domains, while exon 4 codes for the transmembrane domain and the cytoplasmic tail (Fernández-Mestre, et al., 2023). Polymorphisms within both the alpha and beta chains of the DQ molecule dictate peptide binding specificities, potentially yielding four distinct molecules.

In celiac disease, gluten-derived peptides, particularly gliadin fragments as mentioned before, are presented by *HLA-DQ1* molecules to CD4+ T-cells in the intestinal mucosa. This interaction between gliadin peptides and *HLA-DQ1* molecules triggers an immune response, leading to the production of pro-inflammatory cytokines and the recruitment of immune cells to the intestinal epithelium (Fernández-Mestre, et al., 2023). *HLA-DQ1* can do so because the protein made has a high affinity for certain gluten epitopes, such as those found in gliadin (Albert, et al., 2024).

Furthermore, genetic variations in the *HLA-DQ1* gene, particularly certain haplotypes such as *HLA-DQ2* and *HLA-DQ8*, are strongly associated with susceptibility to celiac disease (Albert, et al., 2024). Individuals carrying these *HLA-DQ1* haplotypes have an increased risk of developing celiac disease, as their *HLA-DQ1* molecules are more efficient in presenting gluten peptides to CD4+ T-cells, thus facilitating the initiation of the autoimmune response.

This paper aims to contribute to the ongoing literature review of the *HLA-DQ1* gene by determining the influence of genetics on celiac disease. It is already known that the *HLA-DQ1* gene increases the susceptibility to celiac disease; however, there is very little research done on the influence of specific alleles on the disease, or even on gluten sensitivity. The gene is studied for its high levels of polymorphisms, and although studies looked at the different molecule products, the effects of each allele on each molecule are still understudied in the context of celiac disease, which most papers focusing on the immune response, rather than the subsequent immune disease. For this reason, I am narrowing the focus of this paper and the experiment proposed to the diversity of gene alleles and their effect on celiac disease.

The investigation into the presence of different alleles of *HLA-DQA1* in association with celiac disease started in the late 80s. Gyllensten and Erlich (1988) demonstrated an innovative single-step method for amplifying a single-copy gene and producing an excess of single-stranded DNA for sequencing, aiding in the study of allelic diversity at the *HLA-DQA* locus. Their work was followed by researchers like Briata et al., (1989) who presented evidence suggesting polymorphic cis-acting elements within the *HLA-DQB* gene control splicing and polyadenylation, introducing complexities in gene regulation. These findings underscore the intricate interplay between *HLA-DQA1* alleles and disease susceptibility, not just for celiac disease but also other autoimmune reactions.

In the later years, increased research on the gene began using animal models such as mice. Rajagopalan et al. (2008) aimed to elucidate the impact of *HLA-DQ* polymorphisms on immune response using transgenic mice when presented with specific antigens from bacteria, a topic relevant to autoimmune diseases like celiac disease. Their hypothesis posited that variations in *HLA-DQ* would lead to differential susceptibility to the different superantigens of bacteria, possibly influencing immune activation and downstream responses. To investigate this, the researchers generated transgenic mice expressing different *HLA-DQ* alleles and exposed them to the superantigens of interest. They observed distinct immune responses among the transgenic mice, with certain *HLA-DQ* alleles conferring increased susceptibility to bacterial superantigens, while others exhibited more robust immune regulation. These findings are of great importance to *HLA-DQ* polymorphisms in modulating immune responses to external stimuli, such as bacterial superantigens, and provide insights into how *HLA-DQ* variants are highly associated with the immune response.

To continue my exploration of the gene, I investigated whether researchers are able to use mice as animal models to create transgenic mice and how they do that. Evidence from multiple studies, such as Dewan et al. (2021) confirmed that *HLA-DQA1* can be successfully inserted in mice. Moreover, the human *HLA-DQA1* gene has a mouse homolog called histocompatibility 2, class II antigen A, alpha, noted as H2-Aa (NCBI Gene: 14960, Mouse Genome Informatics, 2024; NCBI, 2024).

Dewan et al. (2021) utilized a gene-targeting approach known as knock-in technology, where they introduced the human *HLA-DQA105:01* allele into the mouse genome at a specific locus. This involved designing a targeting vector containing the human *HLA-DQA105:01* gene sequence flanked by homology arms that match the genomic sequence surrounding the insertion site in the mouse genome. The targeting vector was then introduced into mouse embryonic stem cells, allowing for homologous recombination between the targeting vector and the endogenous mouse DNA. Through a series of selection steps, cells with the desired gene insertion were isolated and used to generate transgenic mice. The resulting mice carried the human *HLA-DQA105:01* gene in their genome, allowing for the study of its function and contribution to celiac disease susceptibility in vivo. Unfortunately, there hasn't been much research done on using this innovative approach to investigate the role of specific *HLA-DQ* alleles in disease pathogenesis and testing potential therapeutic interventions in a relevant animal model. Their work on transgenic mice confirmed the possibility of exploring different human alleles of interest in a mouse model. Hence, my interest in looking at this in a multitude of ways is to provide future research on the precise mechanisms underlying the association between *HLA-DQ* alleles and celiac disease susceptibility.

IV. Future Directions

This essay outlines key experiments that I am interested in conducting to advance our understanding of celiac disease using transgenic mouse models and cutting-edge genetic techniques.

First, I aim to investigate the functional role of *HLA-DQ1* alleles in celiac disease pathogenesis using transgenic mouse models. To accomplish this, I will generate *HLA-DQ1* knock-in mice expressing disease-associated *HLA-DQ1* alleles via CRISPR/Cas9 technology. The CRISPR/Cas9 system allows scientists to introduce a mutation and delete (or alter) the genome at a precise sequence of DNA. I would be interested in using this system, in the case of the *HLA-DQ1* gene, by looking at the H2-Aa mouse homolog, and modifying the sequence to create the *HLA-DQA105:01* allele, which is highly associated with disease progression. Once I know the specific DNA sequence, I can design a single guide RNA (sgRNA) that complementary binds to this sequence. When introduced *in vitro,* it will promote the binding of the Cas9 enzyme to the sgRNA and then to the target DNA by recognizing a PAM (5′ – NGG – 3′) site near the cutting region. This will produce the cut in the DNA that allows me to introduce my wanted allele or site change in case it only requires single nucleotide changes. I hypothesize that *HLA-DQ1* alleles contribute to celiac disease susceptibility by modulating gluten-specific immune responses in the intestinal mucosa.

Following successful integration, the knock-in mice will be bred, and I will test the presence of the *HLA-DQ1* alleles through screening using PCR genotyping. This test could be accomplished by doing PCR on the exact sequence of the allele I am interested in at position chromosome 6. Experimental and control groups will be established, with each group consisting of at least 10 mice for a total of 20 mice in the study. For my study, it is essential to ensure that the immune system of the knock-in mice has sufficiently matured to be able to carry out follow-

up experiments and appropriate testing of immune response to gluten antigens. For this reason and based on previous literature treating transgenic mice, the mice will be tested four weeks after birth (Dewan et al., 2021). Because breeding will be involved, I would expect to see both homozygous and heterozygous mice in my studies. I will categorize homozygous for the human-associated allele as OA (homozygous celiac disease-associated allele) and heterozygous as EA. There will be three groups:

Group 1: OA mice with *HLA-DQA105:01* allele

Group 2: EA mice with *HLA-DQA105:01* allele

Group 3: Control mice (no CRISPR-altered, normal mice homolog)

To prepare mice, both experimental and control groups will be fed a gluten-free diet for the first four weeks of life. Subsequently, once the immune system is mature, all groups will be introduced to a gluten-containing diet which they will be kept on for another four weeks or until time allows. After this time, gluten-specific immune responses will be assessed.

To continue the study and the immune response, I am interested in conducting T-cell activation. This procedure involves the isolation of T-cells from a lymphoid organ of the mice, such as the spleen. Most labs use the density gradient centrifugation to successfully extract the T cells. This is followed by plating the isolated cells in an appropriate culture dish medium that can be tested and stimulated using gluten peptides that are common T-cell activators such as CD3 (T lymphocytes) and CD20 (B lymphocytes) that help identify patients with celiac disease (Mubarak et al., 2015). The cells would then need to be incubated at 37°C with 5% CO2 for an appropriate duration, typically between 24 to 72 hours, as other literature describes for best incubation and experimental requirements (Qiao et al., 2021).

One very important aspect of the study is having correct control groups. Apart from my group 3 T-cell activation assays, I would have two other control groups: Unstimulated T-cells and Irrelevant T-cell stimulation. The first one entails having a group of T-cell cultures that undergoes all the same conditions as the experimental groups but does not get stimulated with the antigens. These groups would serve as a baseline for T-cell background activation and test background cytokine production without any stimuli. The latter group, the irrelevant group, involves another T-cell culture but in this case, it would get an irrelevant stimulus that is not specific to celiac disease and should not be activated. This control helps me to distinguish specific responses to the experimental stimuli from non-specific effects that could affect result interpretation.

Once my groups are set and the incubation period is over, I will finally assess the activation of the T-cells in the three different groups. These tests include cytokine production, which in the case of celiac disease, gluten peptides activate gluten-specific T-cells, leading to the production of pro-inflammatory cytokines such as interferon-gamma (IFN-γ), tumor necrosis factor-alpha (TNF-α), interleukin-2 (IL-2), and interleukin-17 (IL-17). Therefore, I would collect the aforementioned intestinal tissue samples and T-cell cultures from all groups and analyze them using enzyme-linked immunosorbent assay (ELISA) and reverse transcription-quantitative polymerase chain reaction (RT-qPCR). For group 1, I would expect to see increased levels of pro-inflammatory cytokines such as IFN-γ, TNF-α, IL-2, and IL-17 in the intestinal mucosa compared to control mice. In group 2, since these are heterozygous for the allele, I would expect to see some elevated levels of cytokines (although not as high as group 1 since they only have one copy of the human allele) and thus would modify the HLA complex.

The RT-qPCR I would expect to reveal the upregulation of genes encoding for these proinflammatory cytokines. This result can be obtained because the RT-qPCR amplifies and quantifies specific RNA that is reverse-transcribed into DNA sequences in a sample, such as *HLA- DQA1*. This reversion is followed by designing complementary primers, performing PCR amplification with fluorescent detection, and analyzing the fluorescence signals to determine the amount of target RNA present (and thus the expression of the gene.) Through this analysis, I could focus on the mechanism of action of these alleles by studying the expression of the gene and the activation of any signaling pathways that are commonly associated with immune activation and inflammation, such as NF-kB, MAPK, and STAT.

Finally, to complete a thorough investigation of the gene alleles on both genotype and phenotype, I would conduct a histological analysis. This assay allows us to investigate tissue damage, which is essential for seeing if the allele produced intestinal damage after having a gluten-containing diet. For this assay, I would need to collect samples of the small intestine of the mice (properly handling the tissue with paraffin wax) and then section the samples.

The staining process follows using specific antibodies, in the case of immunohistochemistry, such as those markers for inflammation, CD3 and CD20, as well as tissue damage (Parham, 2015). These tissue sections would be examined under a microscope. I would expect that group 1 present the most damage to the lining of the intestine while groups 2 and have less. Group 3 should not show any damage at all, since this is the control mice. This is to say, I expect to see after the histological analysis greater intestinal tissue damage and inflammation in knock-in mice following gluten exposure.

To ensure rigorous data collection and analysis, all experiments will be conducted following approved institutional animal care and use protocols. These experiments will provide valuable insights into the functional role of *HLA-DQ1* alleles in celiac disease pathogenesis and the mechanisms underlying the autoimmune disorder. In my analysis in the lab, I have a newly identified allele of the gene, whose role and association with celiac disease I want to study.

I would be very excited to carry out this long and ambitious analysis of different alleles on disease progression and mechanistic insights into *HLA-DQ1.* However, even though the knock-in mice model has many advantages to studying disease, this model cannot fully capture the complexity of human celiac disease. As discussed before, genetics are an important component, but the disease is greatly influenced by environmental and other immunological factors. Some of the differences in immune system regulation, gut microbiota composition, and environmental exposures (social bonds or isolation) between mice and humans may limit the translational relevance of the literature and my hypothetical potential findings.

V. Conclusion

In conclusion, this study delves into the intricate interplay between genetics, gluten intolerance, and celiac disease, with a particular focus on the role of the *HLA-DQA1* in disease susceptibility. Through a comprehensive review of existing literature and proposed experimental analyses, this research is meant to shed light on the underlying mechanisms driving celiac disease pathogenesis. Understanding specific variants of the *HLA-DQA1* and their role not only in celiac disease but overall in the HLA complex would help develop better therapeutic techniques for patients that involve direct targeting of the biology and not just a diet change. Further research should explore demographics by conducting sequencing analysis of different populations, and thus, create a study of different prevalence of alleles across the world. As of now, most of the research done on celiac disease is done on specific populations, mostly white. Thus, the literature focuses on the alleles present in world regions. The importance of genetics and to provide specialized therapeutics that can be applied worldwide. Ultimately, a deeper understanding of their mechanistic pathways has the potential to inform the development of broader scientific understanding of autoimmune disorders.

References

Albert, E. A., Smirnova, D. I., Zobkova, G. Y., Smirnova, A. V., & Volchkov, P. Y. (2024). Identification of the novel HLA-DQA1*05:107 allele. *HLA*, *103*(4), e15500.

Besser, H. A., & Khosla, C. (2023). Celiac disease: mechanisms and emerging therapeutics. *Trends in Pharmacological Sciences*, *44*(12), 949–962. https://doi.org/10.1016/j.tips.2023.09.006

Celiac disease—Symptoms and causes. (2023). Mayo Clinic.

Dewan, A. E., Koentgen, F., Johannesen, M. K., du Pre, M. F., & Sollid,

L. M. (2021). Generation of an HLA-DQ2.5 Knock-In Mouse. *ImmunoHorizons*, *5*(1), 25–32.

Entry - * 146880; HLA-DQA1 - OMIM. (2021).

Espino, L., & Núñez, C. (2021). The HLA complex and coeliac disease. *International Review of Cell and Molecular Biology*, *358*, 47–83.

Gluten Intolerance: Symptoms, Test, Non-Celiac Gluten Sensitivity. (2021). Cleveland Clinic.

H2-Aa histocompatibility 2, class II antigen A, alpha [Mus musculus (house mouse)] – Gene – NCBI. (2024).

H2-Aa MGI Mouse Gene Detail – MGI:95895 – Histocompatibility 2, class II antigen A, alpha. Mouse Genome Informatics. (2024).

Italian screening study finds an increasing prevalence of celiac disease. (2021). Beyond Celiac.

Mubarak, A., Wolters, V. M., Houwen, R. H., & ten Kate, F. J. (2015). Immunohistochemical CD3 staining detects additional patients with celiac disease. *World Journal of Gastroenterology*, 21(24), 7553–7557. https://doi.org/10.3748/wjg.v21.i24.7553

Parham D. M. (2015). Immunohistochemical Markers of Soft Tissue Tumors: Pathologic Diagnosis, Genetic Contributions, and Therapeutic Options. *Analytical Chemistry Insights*, *10*(Suppl 1), 1–10. https://doi.org/10.4137/ACI.S32730

Qiao, S. W., Dahal-Koirala, S., Eggesbø, L. M., Lundin, K. E. A., & Sollid, L. M. (2021). Frequency of Gluten-Reactive T Cells in Active Celiac Lesions Estimated by Direct Cell Cloning. *Frontiers in Immunology*, *12*, 646163. https://doi.org/10.3389/fimmu.2021.646163

Rajagopalan, G., Polich, G., Sen, M. M., Singh, M., Epstein, B. E., Lytle, A. K., Rouse, M. S., Patel, R., & David, C. S. (2008). Evaluating the role of HLA-DQ polymorphisms on the immune response to bacterial superantigens using transgenic mice. *Tissue Antigens*, *71*(2), 135–145. https://doi.org/10.1111/j.1399-0039.2007.00986.x

Schiepatti, A., Sanders, D. S., & Biagi, F. (2022). Chapter 4—Seronegative villous atrophy. In A. Schieptti & D. Sanders (Eds.), *Coeliac Disease and Gluten-Related Disorders* (pp. 69–87). Academic Press. https://doi.org/10.1016/B978-0-12-821571-5.00001-5

Sciurti, M., Fornaroli, F., Gaiani, F., Bonaguri, C., Leandro, G., Di Mario, F., & De' Angelis, G. L. (2018). Genetic susceptibility and celiac disease: what role do HLA haplotypes

play? *Acta Biomedica: Atenei Parmensis*, *89*(9-S), 17–21. https://doi.org/10.23750/abm.v89i9-S.7953

Villanacci, V., Vanoli, A., Leoncini, G., Arpa, G., Salviato, T., Bonetti, L. R., Baronchelli, C., Saragoni, L., & Parente, P. (2020). Celiac disease: histology-differential diagnosis complications. A practical approach. *Pathologica*, *112*(3), 186–196. https://doi.org/10.32074/1591-951X-157