Genetic Background of Celiac Disease and Its Clinical Implications

Victorien M. Wolters, M.D.; Cisca Wijmenga, Ph.D.

Posted 02/25/2008

Abstract and Introduction

Abstract

Celiac disease (CD) is a complex genetic disorder with multiple contributing genes. Linkage studies have identified several genomic regions that probably contain CD susceptibility genes. The most important genetic factors identified are HLA-DQ2 and HLA-DQ8, which are necessary but not sufficient to predispose to CD. The associations found in non-HLA genomewide linkage and association studies are much weaker. This might be because a large number of non-HLA genes contributes to the pathogenesis of CD. Hence, the contribution of a single predisposing non-HLA gene might be quite modest. Practically all CD patients carry HLA-DQ2 or HLA-DQ8, while the absence of these molecules has a negative predictive value for CD close to 100%. Genetic risk profiles for CD would be helpful in clinical practice for predicting disease susceptibility and progression.

Introduction

Celiac disease (CD) is a common enteropathy with a strong genetic risk. It is characterized by a permanent intolerance for gluten proteins present in dietary wheat, rye, and barley. It affects approximately 1:100-300 individuals,[1-3] although only 1 person in ~8 is aware of being affected because the symptoms may be mild or nonspecific.[4] Environmental, genetic, and immunologic factors are important in the pathogenesis of CD.

Genes play a key role in CD and considerable progress has been made in identifying some of those responsible for CD. The roles of HLA-DQ2 and HLA-DQ8 are well known as almost all patients carry the genes encoding these heterodimers.[5,6] Non-HLA genes also contribute to the development of CD, but these associations are less evident. Genetic linkage analyses have identified susceptibility loci on various chromosomes, such as 2, 5, 6, 9, 15, and 19, revealing the complexity of CD.[7-16]

Gluten is the most important environmental factor. Gluten proteins provoke the disease as the high proline content of gluten is relatively resistant to proteolytic digestion in the intestinal tract.[17,18] The undigested gluten peptides are deamidated by tissue transglutaminase, which results in a better binding capacity to the pocket of HLA-DQ2 or HLA-DQ8 molecules on antigen-presenting cells (see Figure 1). This complex is presented to CD4+ T cells and the ensuing immune response causes inflammation and intestinal tissue damage. A direct response of the epithelium via the innate immune system also plays a role.
Figure 1.

Pathogenesis of celiac disease: Gliadin is absorbed into the lamina propria and presented in conjunction with HLA-DQ2 or HLA-DQ8 cell-surface antigens by antigen-presenting cells, probably dendritic cells, to sensitized T cells expressing the α/β-cell receptor. Tissue transglutaminase deamidates gliadin peptides, generating acidic, negatively charged residues of glutamic acid from neutral glutamines. Because negatively charged residues are preferred in positions 4, 6, and 7 of the antigen-binding groove of HLA-DQ2, deamidated gliadin elicits a stronger T-cell response (with permission of Farrell RJ, Kelly CP in N Engl J Med 2002;346:180-8).

Genetic Epidemiology of CD

Multiple lines of evidence favoring a genetic contribution to the pathogenesis of CD have been suggested by epidemiologic data (Table 1). A familial aggregation is found in 5-15% of CD patients and a striking 83-86% concordance rate was observed among monozygotic twin pairs. CD incidence and prevalence have been found to vary significantly, depending on geographic location and racial or ethnic background; these differences might be reflected by either genetic or environmental susceptibility factors.

Gene Identification in CD

Two complementary approaches are used in the search for genetic susceptibility genes in CD: genetic linkage and genetic association studies. Genetic linkage studies make use of families with affected sibling pairs to identify chromosomal regions shared between the affected siblings above the mean of what is statistically expected. To identify the actual susceptibility locus we use genetic markers (SNPs; single nucleotide polymorphisms). Linkage regions usually encompass 10-100 genes. Once linkage is identified, the next step is a genetic association study to identify the specific disease gene from the candidate gene locus.

Candidate gene association studies search for differences in frequencies of genetic variants in patients compared to control individuals. Such association studies can focus on positional candidate genes from a linkage region, or on functional candidate genes selected from hypothesized disease pathology. More recently, it has become feasible to perform genome-wide association studies -- a hypothesis-free approach which can test thousands of SNPs across the whole genome for association.

Genetic Linkage Studies in CD

CELIAC1 Locus
The CELIAC1 locus on chromosome 6p21 contains HLA class II molecules. It is unequivocal that CD is strongly associated with specific HLA class II genes known as HLA-DQ2 and HLA-DQ8. HLA-DQ molecules are heterodimers consisting of an α and β chain. Particularly the combination of alleles encoding for the α chain DQA1*05 and β chain DQB1*02 of the HLA-DQ2 heterodimer are associated with CD. Most CD patients (~95%) express HLA-DQ2 and the remaining patients are usually HLA-DQ8 positive. The HLA-DQ2 allele is common and is carried by approximately 30% of whites. However, only ~3% of individuals in the general population who carry HLA-DQ2 will develop CD. It is noteworthy that individuals homozygous for the DQ2 molecule comprise approximately 2% of the European population but make up approximately 25% of all CD patients. Thus, HLA-DQ2 or HLA-DQ8 is necessary for disease development but not sufficient as its estimated risk effect is only 36-53%. Thus, non-HLA genes may well contribute more than HLA. The importance of non-HLA genes is supported by the difference in concordance rates seen among HLA identical siblings (~30%). However, linkage peaks observed in non-HLA regions are much lower and not consistent compared to HLA. This might be because many non-HLA genes contribute to the pathogenesis of CD. Hence, the contribution of a single predisposing non-HLA gene might be modest.

CELIAC2 Locus

The support for linkage to the CELIAC2 locus on chromosome 5q31-33 was first identified by Greco et al. This region contains a cytokine gene cluster and it might play a role in immune regulation and inflammation.

CELIAC3 Locus

The CELIAC3 locus on chromosome 2q33 has shown linkage to CD and was replicated by a few but not all the studies performed. The CELIAC3 locus contains the T lymphocyte regulatory genes CD28, CTLA4, and ICOS.

CTLA4 (cytotoxic T lymphocyte-associated antigen 4) is a negative costimulatory molecule of the T-cell response and was already pinpointed as a candidate gene based on this known function before the era of genomewide linkage studies. No evidence for a single mutation in CTLA4 specific to CD has been found but a strong association was suggested at the haplotype level. The genetic variants in the CTLA4 gene in CD were extensively studied in several populations but with opposite results.

CELIAC4 Locus

A genomewide scan by our group identified a region of significant linkage at chromosome 19p13.1. Further association analysis showed association to the myosin IXB gene (MYO9B). Interestingly, MYO9B is a good candidate gene for CD because of its function; it encodes an unconventional myosin molecule that may have a role in actin remodeling of epithelial enterocytes. It is hypothesized that this genetic variant might lead to an impaired intestinal barrier, which might allow the passage of immunogenic gluten peptides. This could be a factor involved in the early mucosal events preceding the inflammatory response in CD. However, in CD populations in the United Kingdom, Spain, Italy, and Scandinavia this association could not be replicated. Furthermore, it is unlikely that the SNP itself is the causing mutation as it is located deep in the intron, which means that it has no function in coding for protein synthesis. Most probably, the SNP is rather a disease marker, an allele “hitchhiking” (in linkage disequilibrium) with the true causative variant.

Chromosome 19p13 was also shown to have significant linkage to inflammatory bowel disease (IBD6) and recently our group showed an association of ulcerative colitis with MYO9B and Morbus Crohn, albeit weaker. This suggests that MYO9B might also promote susceptibility to other intestinal inflammatory diseases, although the precise mechanism of how gene variants in MYO9B can lead to altered gut function is unclear.

A strong association to MYO9B was also reported in patients with a complicated form of CD, known as refractory CD type II (RCDII). In this group, the enteropathy persists despite adherence to a gluten-free diet or it recurs after an initially good response to the diet. RCDII patients are characterized by the presence of aberrant intraepithelial lymphocytes in the small bowel mucosa.

Genetic Association Studies

CTLA4
Although increasing data pinpoint \textit{CTLA4} as a candidate gene, results of association studies did not support certain polymorphisms in \textit{CTLA4} being the major susceptibility locus for CD\textsuperscript{[14,28,29,32,34,42-44]} A recent study analyzing all common SNPs in \textit{CTLA4} suggested an association on the haplotype level rather than on the single SNP variant level\textsuperscript{[34]} In Figure 2, the suggested genetic loci of CD are shown.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.pdf}
\caption{Genetic loci in CD.}
\end{figure}

\textbf{Other Candidate Genes}

Other studies did not confirm association for promising genes like tTG, FAS, MMP-1 and 3, TCR\textsubscript{\beta}, IL12\textsubscript{\beta}, CD28, CD80, CD86, KIR, LILR, STAT 1, PGPEP1, IRF1, DPPIV, TGM2, NOS2, or IFN\textsubscript{\gamma}\textsuperscript{[35]}

\textbf{Other New (Genetic) Developments in CD}

\textbf{Factor V Leiden}

Recently, a genetic association between factor V Leiden (FVL) and CD was discovered\textsuperscript{[45]} A family was described in which the two diseases segregated in all cases as no sibling was affected by only one of the diseases, suggesting that the genetic mutation responsible for the development of CD in this family occurs in a gene very close to FVL on chromosome 1q. Further assessment of this association is needed to unravel the pathogenesis of CD.

\textbf{Possible Causal Factors}

\textbf{Breastfeeding, Amount of Gluten, and Age of Gluten Introduction}

Sweden experienced a threefold increase in incidence of CD in children younger than 2-yr old in the mid-1980s\textsuperscript{[27]} This was partly explained by changing recommendations on gluten introduction in infants (in 1982) from 4 to 6 months of age; as a result more infants were introduced to gluten without ongoing breastfeeding. At the same time, the content of gluten in baby food was increased. In 1996, the Swedish authorities recommended a gradual introduction of gluten
from the age of 4 months while breastfeeding, and the incidence of CD soon dropped dramatically. The current European recommendation states that breastfeeding during gluten introduction might be beneficial to children at high risk for CD, although it is unclear if breastfeeding prevents CD or simply delays the onset of disease.\cite{46}

**Infectious Episodes**

Interestingly, in the Swedish epidemic of CD, children born during the summer had a greater risk for CD, which might be because gluten was introduced during the winter when infections are more common.\cite{47} Two explanations for this greater CD risk were suggested: infections might change gut permeability leading to the passage of immunogenic gluten peptides through the epithelial barrier. The other possibility implies that sequence similarities exist between proteins produced during adenovirus infections and proteins of gluten. In 1987, Kagnoff et al. suggested a role for human intestinal adenovirus.\cite{48} Recently, a prospective study showed that multiple rotavirus infections predicted a higher risk of CD.\cite{49}

**Medication**

The onset of CD has been reported during a course of treatment of hepatitis C with interferon (IFN).\cite{50-52} This medication might increase epithelial permeability and proinflammatory cytokine production.\cite{53,54} In hepatitis C patients, the activation of silent CD during IFN treatment should be suspected although symptoms subsided in almost all patients after IFN was withdrawn and without a gluten-free diet.\cite{55} However, the histological abnormalities were still seen several months after discontinuing IFN, which means a timely diagnosis of CD after IFN.

**Animal Model**

So far there is no good animal model for CD although some experience is available with gluten-sensitive enteropathy (GSE) in Irish setter dogs.\cite{56}

**Genes and Syndromes**

Several genetic syndromes are associated with CD, e.g., Down syndrome (DS), Turner syndrome (TS), and Williams syndrome (WS), and it is tempting to speculate that possible chromosomal derangements might influence a disturbed immune response in these syndromes.

**Clinical Relevance**

To date, only HLA-DQ2 or HLA-DQ8 typing is clinically relevant, as all the other promising genetic loci could not be replicated consistently. The main role of HLA typing lies in its high negative predictive value to exclude CD (close to 100%). CD can be virtually excluded in nonbiopsy-proven white individuals on a gluten-free diet who are non-HLA-DQ2 and non-HLA-DQ8. HLA typing can be useful to help exclude the possibility of the future development of CD in patients at high risk and can provide additional information if the clinical picture is unclear. First-degree relatives of CD patients should be HLA typed and if CD cannot be excluded (i.e., HLA-DQ2 or HLA-DQ8 positive) serologic tests might be performed in asymptomatic patients with a frequency of approximately every 5-10 yr when patients still have growing potential (<20 yr). In asymptomatic HLA-DQ2/DQ8 positive first-degree relatives >20 yr old one single screening at the age of ~50 yr might be indicated as complications of CD can develop. When a first-degree relative is symptomatic, a low threshold for biopsy is indicated.

**Future Prospects**

We assume an etiology model for CD comprising a major gene (HLA) and several low risk genes with a function in the intestinal barrier and immune system. It is therefore important to identify susceptibility genes as well as protective genes in CD.

In the coming years, identifying other target genes and understanding the pathways they influence will lead to a better understanding of CD pathogenesis. Ultimately, we might be able to define genetic risk profiles for more precise diagnoses and for predicting disease progression, and they may lead to novel therapies.

**Clinically Relevant Conclusions**
Genetic susceptibility is a prerequisite for developing CD, with HLA class II as the most important genetic factor identified so far.

The absence of HLA-DQ2/DQ8 can exclude the possibility or future development of CD with a certainty close to 100%.

In asymptomatic HLA-DQ2/DQ8 positive first-degree relatives <20 yr old serologic screening is indicated approximately every 5-10 yr. In asymptomatic HLA-DQ2/DQ8 positive first-degree relatives >20 yr old one single screening at the age of approximately 50 yr is indicated, as complications of CD can develop.

The presence of HLA-DQ2/DQ8 is clinically irrelevant since 30-40% of the general population is positive for one or both of these factors and only a fraction (0.5-1%) of these individuals has CD.

Considerable advances in the genetics of CD will identify more genetic determinants for CD development and disease progression and establish clinically relevant genetic risk profiles. CD might serve as a model for other autoimmune diseases.

---

Table 1. Evidence for Genetic Susceptibility to Celiac Disease

<table>
<thead>
<tr>
<th>Evidence for Genetic Susceptibility</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnic differences in disease incidence/prevalence</td>
<td>(21)</td>
</tr>
<tr>
<td>Familial aggregation</td>
<td>5-15% first-degree relatives of affected individuals are also affected</td>
</tr>
<tr>
<td></td>
<td>30% HLA identical sibs</td>
</tr>
<tr>
<td>Twin studies concordance rates</td>
<td>Monozygotic twins 83-86%</td>
</tr>
<tr>
<td></td>
<td>Dizygotic twins 11%</td>
</tr>
<tr>
<td>Identification of susceptibility loci by genome screening</td>
<td>Chromosomes 2, 5, 6, 9, 15, 19</td>
</tr>
<tr>
<td>Genetic association studies of functional candidate genes</td>
<td>CTLA4</td>
</tr>
<tr>
<td>Association found with genetic syndromes</td>
<td>Down syndrome (57,58)</td>
</tr>
<tr>
<td></td>
<td>Turner syndrome (59-62)</td>
</tr>
<tr>
<td></td>
<td>Williams syndrome (63)</td>
</tr>
</tbody>
</table>

---

References


Acknowledgements

We thank Chris Mulder, Jackie Senior, and Alexandra Zhernakova for commenting on the article.

Reprint Address

Cisca Wijmenga, Ph.D., Department of Genetics, University Medical Centre Groningen, P.O. Box 30 001, 9700 RB Groningen, The Netherlands.

Victorien M. Wolters, M.D.,1 Cisca Wijmenga, Ph.D.2,3

1Department of Pediatric Gastroenterology, University Medical Centre Utrecht, Utrecht, The Netherlands
2Department of Biomedical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands
3Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands

Disclosure: Victorien Wolters received funding from The Wilhelmina Research Fund and Cisca Wijmenga received funding from the Netherlands Organization of Scientific Research, the Dutch Digestive Diseases Foundation, and the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009).