

Chapter 5

Immunogenetics of Celiac Disease

Esther C. de Haas, Vinod Kumar, and Cisca Wijmenga

Heritability of Celiac Disease

Celiac disease (CD) can be considered as a model for common complex disorders in which the phenotypes result from a combination of environmental triggers, genetic predisposing factors, and their interactions. Although the gliadin fraction of wheat gluten and similar protein fractions of other grains, which are the primary and necessary environmental triggers for CD, have been well defined since the 1950s, the discovery of genetic risk factors is an ongoing process.

In contrast to Mendelian single-gene disorders, CD does not show a clear pattern of inheritance. The importance of genetics is indicated by familial clustering, shown by epidemiological studies comparing the prevalence of CD in related individuals to the prevalence in unrelated individuals. The reported prevalence in first-degree relatives of CD patients ranges from 2.8 to 22.5 %, with the higher prevalences reported in at-risk relatives undergoing routine intestinal biopsy instead of only serological screening [1]. In contrast, the overall prevalence of CD in North America and Western Europe ranges from 0.5 to 2.9 % [1, 2]. The recurrence risk ratio for siblings (λ_s), the prevalence of CD in siblings divided by the prevalence in the general population, has been reported to be as high as 20–60 [3–5]. A large multicenter study in the USA among at-risk and not-at-risk groups found a prevalence of 4.5 % in first-degree relatives and 2.6 % in second-degree relatives, compared to 0.3 % in not-at-risk children and 0.9 % in not-at-risk adults [6].

In order to distinguish the role of genetic factors in familial clustering from environmental factors, twin studies are performed. In a large population-based twin study in Italy, the estimated case-wise concordance, a measure for disease risk if a co-twin is affected, was 83.3 % (95 % confidence interval 70.3–96.4 %) for

E.C. de Haas, M.D. Ph.D. (✉) • V. Kumar, Ph.D. • C. Wijmenga, Ph.D.
Department of Genetics, University Medical Center Groningen, University of Groningen,
PO Box 30.001, 9700 RB, Groningen, The Netherlands
e-mail: e.de.haas@umcg.nl

monozygotic twin pairs and 16.7 % (3.6–29.8 %) for dizygotic twin pairs [7]. Since the proportion of affected co-twins in dizygotic twin pairs is in line with the reported prevalence of CD in siblings, the role of shared environmental factors, with the exception of exposure to gluten antigen, is thought to be limited [8]. Assuming a population prevalence between 0.1 and 1.1 %, the variance in CD prevalence attributable to genetic variance, the so-called heritability of CD, is estimated to be 57–87 % (95 % confidence intervals 32–100 %) [7].

The Immunogenetics of Celiac Disease

Identification of Susceptibility Genes for Celiac Disease

The above studies on CD prevalence contribute to estimations of the heritability, but do not provide information on which genes or how many genes are actually involved in disease development. Other approaches are needed to identify susceptibility genes: linkage analysis and genetic association analysis (both candidate genes and genome-wide) (see Text Box 5.1 Genetic Linkage and Association analysis).

The first identified genetic risk factor for CD, the human leukocyte antigen DQ (HLA-DQ) genotype, is the strongest known genetic risk factor. Since serological studies in the 1970s discovered the association between HLA and CD, many others have confirmed the strong linkage to HLA, specifically to HLA-DQ2 and HLA-DQ8. However, no other genetic associations were consistently found by linkage studies and candidate gene studies. The recent development of genome-wide association studies (GWAS) has led to the discovery of several additional susceptibility genes for CD. So far, GWAS have identified 57 associated single nucleotide polymorphisms (SNPs) located in 39 non-HLA regions, with most of the positional candidate genes having immunological functions.

Association with HLA Genotype

The major HLA class II, also called Major Histocompatibility Complex II (MHC II), molecules DP, DQ, and DR are cell-surface receptors on antigen-presenting cells involved in the presentation of exogenous peptide antigens to T-helper lymphocytes. The encoding genes are part of the 200 genes encompassing 4 Mb HLA-complex on chromosome 6p21. This region corresponds with the *CELIAC1* locus, a region consistently found to be associated with CD in both linkage and association studies.

The first reports on the association with HLA revealed a link between CD and positive serology for HLA class I B8-antigen and later HLA-DR3-antigen [9, 10]. The encoding alleles of the HLA-B gene and HLA-DR genes are strongly linked in the haplotype A1-B8-DR3-DQ2, which is present in approximately 10 % of

Text Box 5.1 Genetic Linkage and Association Analysis

For a long time, the search for genetic risk factors for complex disorders was based on genetic linkage analysis and association studies with candidate genes. In *genetic linkage analysis*, genetic markers are used to identify chromosomal regions that are shared by affected relatives more often than expected by meiotic segregation. Identified regions may contain up to hundreds of genes. Subsequently, fine mapping is needed in order to pinpoint the causal gene. This can be achieved by association analysis.

In *association analysis* the frequencies of genetic variants are compared between patients and healthy control persons. In *candidate gene association studies*, the analyzed genetic variants are often selected for their location in chromosomal regions linked to the disease or because of their hypothesized function in the disease of interest.

The more recent *GWAS* are hypothesis-free and compare large numbers, hundreds to thousands, of patients and healthy control persons for hundreds of thousands to millions of genetic variants, so-called single nucleotide polymorphisms (SNPs), dispersed over the whole genome. An association analysis relies on the strong linkage over short distances between the analyzed SNPs and potential causal genetic variants; this is the so-called linkage disequilibrium (LD). Most disease-associated SNPs are not the causal variant but a marker or tag for a nearby causal variant. GWAS have contributed greatly to identifying genetic risk factors for complex disorders. The use of hundreds of thousands to millions of SNPs means it is possible to achieve a high resolution compared to linkage studies, while the large numbers of analyzed patients and control persons add to the statistical power to detect genetic variants with small or modest effect sizes. Importantly, the hypothesis-free approach has led to the identification of genes and pathways not previously suspected of being involved in the disease of interest.

Northern Europeans [11]. Subsequent studies pinpointed the association of CD to alleles encoding HLA-DQ2 molecules [12, 13].

The HLA-DQ molecule is a heterodimer consisting of an α chain and a β chain, encoded by HLA-DQA1 and HLA-DQB1. Further characterization of the association between HLA-DQ and CD showed that especially homozygosity for the HLA-DQ2.5 heterodimer (encoded by the alleles DQA1*0501 and DQB1*0201) and, to a lesser degree, heterozygosity for the HLA-DQ2.5 heterodimer combined with the HLA-DQ2.2 heterodimer (encoded by the alleles DQA1*0201 and DQB1*0202) were associated with a strongly increased susceptibility for CD [14, 15] (Fig. 5.1). Assuming a CD prevalence of 1 % in the general population, the absolute risk for CD is estimated at >7 % for this high-risk group [16].

Heterozygosity for the HLA-DQ2.5 heterodimer combined with another HLA-DQ heterodimer, or homozygosity or heterozygosity for the HLA-DQ8

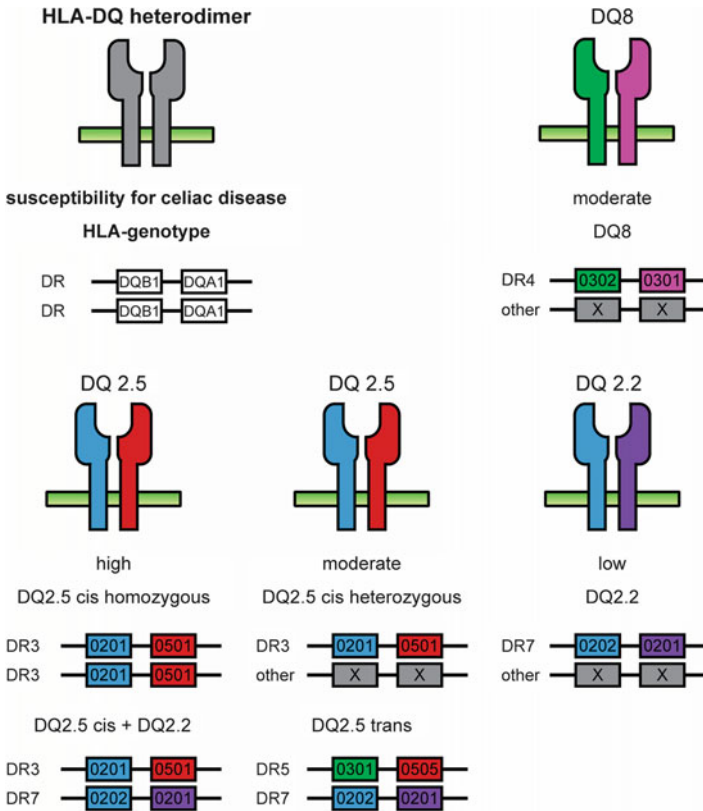


Fig. 5.1 HLA-DQ heterodimers with coding HLA-genotypes and corresponding susceptibility for CD

heterodimer (encoded by the alleles DQA1*0301 and DQB1*0302), confers a more moderately increased risk for CD, with an absolute risk for CD estimated at 0.1–7 % [16]. Functional studies showed that these HLA-DQ2 molecules and, to a lesser degree, HLA-DQ8 molecules have a high affinity for gluten peptides and that gluten-reactive T lymphocytes from the small intestinal mucosa of CD patients preferentially recognize gluten peptides when presented by HLA-DQ2 or HLA-DQ8 [17–20]. Approximately, 95 % of the CD patients carry the HLA-DQ2.5 genotype, and many of the other 5 % carry HLA-DQ8 [13, 21, 22]. There is a significant worldwide correlation between the combination of wheat consumption and frequency of HLA-DQ2 and HLA-DQ8, on the one hand, and the incidence of CD, on the other hand (estimated correlation coefficient $R^2=0.4$) [23]. This observation is in line with a CD model of genetically susceptible individuals in whom dietary gluten triggers intestinal inflammation.

Although the presence of either HLA-DQ2 or HLA-DQ8 can be considered necessary for the development of CD, neither is sufficient, since only some 3 % of the approximately 40 % of Caucasians who carry either HLA-DQ2 or HLA-DQ8 will

actually develop CD [24]. This suggests that, in addition to wheat gluten consumption and HLA genotype, other environmental and genetic factors must be involved in CD etiology. A recent study on the HLA-complex in CD showed that this region, with its many genes in strong linkage disequilibrium (LD), might contain more susceptibility genes for CD [25].

Based on the assumption of a multiplicative model of genetic risk and an estimated recurrence risk for siblings stratified for HLA genotype (λ_{sHLA}) of 2.3–5.3, the contribution of HLA to the heritability of celiac disease is estimated at between 21 and 44 % [4, 5].

Association with non-HLA Genes

With the completion of the human genome sequence, millions of SNPs have been identified. Using these SNPs as genetic markers, called tag SNPs, GWAS have helped to identify thousands of susceptibility variants for hundreds of complex diseases.

Two GWAS on CD and their follow-up studies revealed 26 non-HLA regions to be associated with CD [26–29]. Denser genotyping with the Immuchip, a custom-made array that covers common variants from 186 GWAS loci associated with 12 immune-mediated diseases, identified 13 additional non-HLA regions [30].

Thus, in addition to HLA, there are now 39 known non-HLA regions associated with CD, of which 36 have been genotyped with a high variant density. These regions contain 57 independently associated SNPs [30] (Fig. 5.2). The association signal can be refined to a single candidate gene for 19 of the regions (see Text Box 5.2. “From Genetic Markers to Candidate Genes and Pathways”). However, only three of the associated SNPs are linked to protein-altering variants located in exonic regions, and eight additional SNPs are localized upstream around the transcription start site (5′ untranslated region) of a specific gene or downstream around the 3′ untranslated region [30]. Although most SNPs are localized in nonprotein coding intergenic and intronic regions, the regions associated with CD are greatly enriched with regions involved in regulating the expression of one or more genes, so-called expression Quantitative Trait Loci (eQTLs) [29].

The candidate genes for CD identified by GWAS provide important clues to the disease pathogenesis, including the pathways that are deregulated in CD. Pathway enrichment analyses using susceptibility genes have shown that most susceptibility genes for celiac disease are involved in immune processes. These pathways concern both the adaptive immune response and the innate immune response. The adaptive immune response includes T-lymphocyte maturation and differentiation (e.g., the *RUNX3*, *ETS1*, *IL2*, *IL21*, *IL12A*, *IL18R1*, and *IL18RAP* genes), T- and B-lymphocyte activation and immune cell signaling (e.g., the *CD28*, *CTLA4*, *ICOS*, *ICOSLG*, *PTPN2*, *SOCS1*, *SH2B3*, *UBASH3A*, and *FASLG* genes), and chemokine-induced cell migration (e.g., the *CCR1-3* and *CCR4* genes). The innate immune response

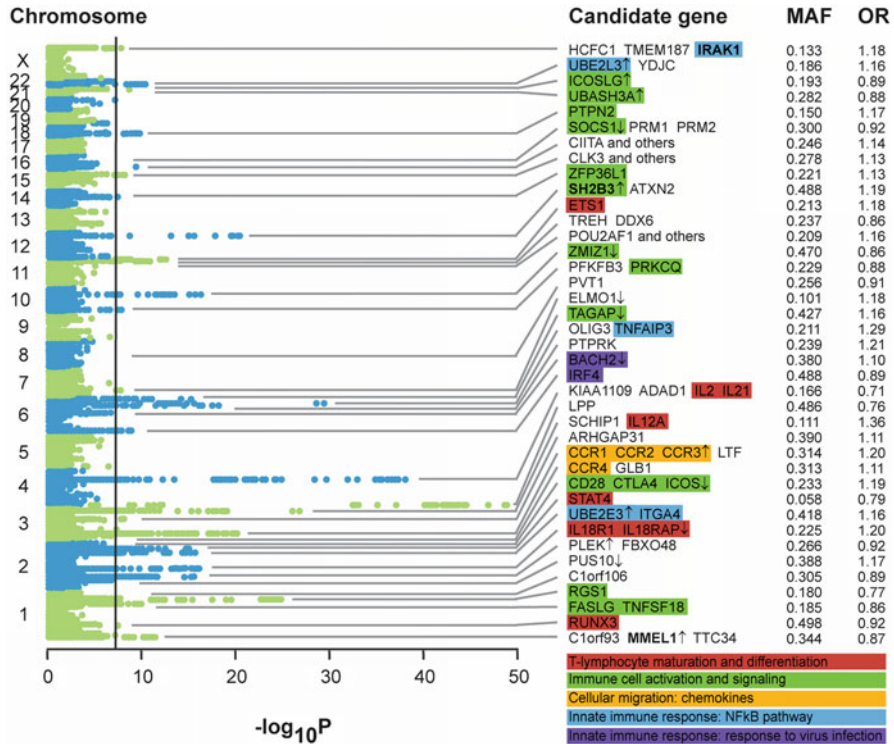


Fig. 5.2 Manhattan plot of the 39 non-HLA regions associated with CD and identified by Immunochip analysis. The vertical line represents the genome-wide significance threshold at $P=5 \times 10^{-8}$. For each associated region the candidate genes are shown with the minor allele frequency (MAF) in the European control population and the odds ratio (OR) for the most significantly associated single nucleotide polymorphism (SNP) of each region. In three regions, the most significantly associated SNP is linked to a protein-altering variant (*IRAK1*, *SH2B3*, and *MMEL1*, in bold). Several SNPs are associated with a change in expression of one or more genes: † for increased expression and ‡ for decreased expression (Kumar V et al., unpublished data). Candidate genes known to be involved in immunological pathways are highlighted

includes the NF κ B-pathway (e.g., the *UBE2E3* and *TNFAIP3* genes) and the response to viral infections (e.g., the *BACH2* and *IRF4* genes) [29, 31] (see Fig. 5.2).

These results do indeed suggest that CD is a T lymphocyte-mediated immune disorder. Furthermore, enrichment of genes involved in natural killer (NK) cell-activation and interferon-gamma production compared to other autoimmune diseases indicates involvement of this pathway in CD [23]. Altogether, celiac susceptibility genes appear to be involved in both the adaptive as well as innate immunity. It has been suggested that the interplay between innate and adaptive immunity on exposure to environmental triggers is a main pathogenic factor in CD [32]. In addition to gluten, various infectious agents have been proposed as triggering environmental factors in genetically predisposed individuals. Moreover, it has been hypothesized that gut microbiota may play a role in CD pathogenesis

Text Box 5.2 From Genetic Markers to Candidate Genes and Pathways

Since the associated SNPs in GWAS will in most cases not be the causal variant but a tag for a nearby causal variant, additional analyses can help to pinpoint the best candidate genes. *Denser genotyping* with the analysis of multiple SNPs located within the same LD block as the tag SNP may help identify SNPs with a stronger association to the disease of interest and narrow down the genetic region to study. *Sequencing* of regions of interest may help to identify less common variants with a potential biological effect. In *eQTL analysis*, the associated SNPs are tested for their correlation with the expression levels of nearby genes (*cis-eQTL*) or genes some distance away (*trans-eQTL*). The result may suggest a functional role for the genes whose expression correlates with the disease-associated SNPs. *Annotation catalogues*, containing information on the functions of genes and genetic variants, and *gene co-expression networks* can further contribute to identifying genes and pathways that may have a function in the disease of interest.

(as discussed in depth in Chap. 7) [33, 34]. Hence, it might be relevant to analyze CD susceptibility genes in the context of interactions between microbes and the host immune response.

Current Challenges in the Search for Genetic Susceptibility Factors

Although GWAS have identified 57 independently associated non-HLA SNPs to CD, the exact causal variant in each region is still unknown. This can be explained by the fact that the analyzed SNPs are in fact tag SNPs, which are in linkage with more than one variant within a so-called LD block. Molecular functional analyses will be necessary to delineate the true causal variants and to understand the mechanism of how these variants affect CD. In addition, most of the associated SNPs are common and associated with a modest increase in CD risk (median odds ratio 1.17, range 1.10–1.70) or a modest decrease in risk (median odds ratio 0.88, range 0.71–0.91) [30]. Since the associated SNPs are tag SNPs, the true effect sizes of the causal variants may be underestimated.

The known non-HLA susceptibility regions, together with HLA, explain approximately 54 % of the heritability of CD [35]. So nearly 50 % of the heritability still needs to be explained. It is not clear whether this hidden heritability is due to thousands of common variants with smaller effect sizes or to individual mutations with strong effect sizes. Future studies performing whole genome sequencing in CD patients may provide answers to this question [36].

From the ENCODE project, it is now apparent that most of the human genome is transcribed to not only protein-coding transcripts but also to large numbers of noncoding RNA molecules of different size [37]. These noncoding RNAs include short noncoding RNAs such as microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs), as well as a new class of long noncoding RNAs (lncRNAs), which are larger than 200 nucleotides [38]. Interestingly, several CD-associated SNPs map to lncRNA regions [39], and it has been shown that disease-associated SNPs can alter lncRNA expression [40]. Hence, the identification of genetic variants associated with CD and that map to noncoding transcripts could help us not only to explain the hidden heritability of CD but also to better understand the disease mechanism.

Shared Immunogenetics with Immune-Related Diseases

CD is associated with several other autoimmune diseases and immune-mediated diseases. Patients with type 1 diabetes mellitus and autoimmune thyroiditis (Graves' disease and Hashimoto's disease) belong to the high-risk populations for CD, with estimated prevalences of 3–6 % in type 1 diabetes patients [1] and 3–8 % in autoimmune thyroiditis patients [41]. In addition, the prevalence of immune-related diseases, including type 1 diabetes and autoimmune thyroiditis, is increased in CD patients compared to the general population [42–45]. In different cohorts of CD patients, the overall prevalence of autoimmune diseases, excluding dermatitis herpetiformis, was approximately 20 %, compared to approximately 11 % in control groups [44–46]. In a retrospective study of CD patients, their cumulative risk for autoimmune disease increased from 8 % at 15 years of age to 33 % at 50 years of age. Type 1 diabetes comprised 29 % of the reported cases and autoimmune thyroiditis 26 %, while other diseases such as psoriasis (8 %), inflammatory bowel disease (7 %), and rheumatoid arthritis (3 %) were reported in fewer patients [45].

Risk factors for the development of another autoimmune disease were a positive family history for autoimmune disease (hazard ratio 2.4, 95 % confidence interval 1.7–3.3) and a diagnosis of CD before 36 years of age (hazard ratio 2.7, 95 % confidence interval 1.8–3.9). In contrast, a positive family history for CD was not associated with an increased risk for other autoimmune diseases in CD patients [45]. This suggests that some susceptibility genes for CD may be shared with other immune-related diseases.

Many immune-related diseases have been linked to the HLA region. The CD HLA risk haplotypes containing HLA-DQ2.5 and HLA-DQ8 also belong to the most susceptible HLA haplotypes for type 1 diabetes [47]. It is estimated that approximately one-third of the type 1 diabetes patients who are homozygous for HLA-DQ2.5 have CD-associated transglutaminase autoantibodies, compared to 1 % of the patients without HLA-DQ2 or HLA-DQ8. It is therefore likely that CD and type 1 diabetes share more risk factors [48].

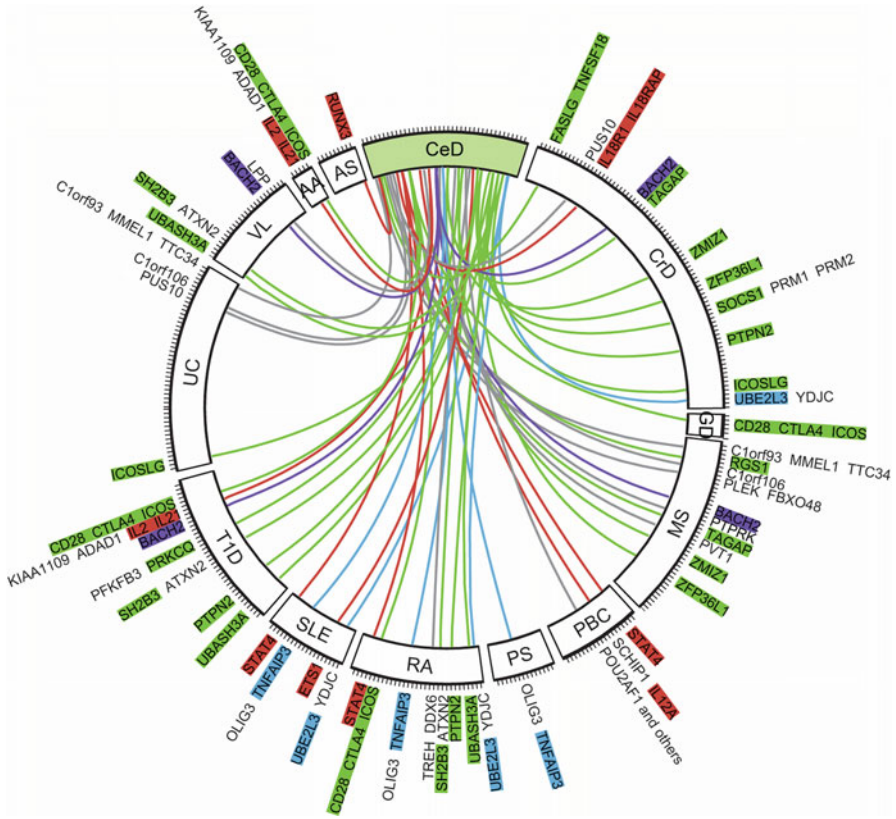


Fig. 5.3 Non-HLA susceptibility regions shared between celiac disease (CeD) and the following immune-related diseases (using a genome-wide significance threshold at $P = 5 \times 10^{-8}$): *CrD* Crohn’s disease, *GD* Graves’ disease, *MS* multiple sclerosis, *PBC* primary biliary cirrhosis, *PS* psoriasis, *RA* rheumatoid arthritis, *SLE* systemic lupus erythematosus, *T1D* type 1 diabetes mellitus, *UC* ulcerative colitis, *VL* vitiligo, *AA* alopecia areata, *AS* ankylosing spondylitis. Adapted from [59]

A meta-analysis on genetic susceptibility regions discovered by GWAS for immune-related diseases (including CD and type 1 diabetes) showed that 44 % of the regions were associated with more than one immune-related disease. This confirms a widespread sharing of non-HLA susceptibility regions between immune-related diseases [49]. Of the non-HLA susceptibility regions for celiac disease, 30/39 (80 %) have been associated with other immune-related diseases (Fig. 5.3).

Analysis of pathways shared by candidate genes directly associated with two or more immune-related diseases showed that there are three major immunological pathways involved: T-lymphocyte differentiation, immune cell signaling, and the innate immune response [50]. CD susceptibility genes appear to be linked to pathways that are strongly involved in T lymphocyte-mediated autoimmune diseases,

such as type 1 diabetes and autoimmune thyroiditis, but not in inflammatory bowel disease, for example [23].

In addition, some of the identified candidate genes appear to be specifically associated with one or a few immune-related diseases. Disease-specific genes could provide insight into specific aspects of the disease pathogenesis. For example, LPP (LIM domain-containing preferred translocation partner in lipoma), which is shared with vitiligo, is involved in cell adhesion and may have a structural role in the intestine [50].

The combined analysis of immune-related diseases in the future may add power to studies to discover more common genetic variants with smaller effect sizes and may contribute to insights into their shared etiological pathways.

Towards Clinical Applications

The ultimate aim of discovering causal genes and pathways involved in CD is to improve the accuracy of diagnosis and to contribute to risk stratification for determining the follow-up and treatment needed.

Thus far, HLA-DQ genotype is the strongest genetic factor linked to CD. HLA-genotyping with a tag SNP method, using six HLA-tagging SNPs, predicts HLA-DQ risk type with high accuracy and is a cost-effective method suited for large-scale use [51, 52].

HLA-DQ2 and HLA-DQ8 combined have a sensitivity of median 96 % [53]. Individuals who have neither HLA-DQ2 nor HLA-DQ8 are unlikely to have CD. HLA-DQ genotyping could therefore be used to exclude CD or make it unlikely in patients with an uncertain diagnosis. In addition, genotyping could be used as a first-line test in the screening of asymptomatic individuals with an increased risk for CD, for example, patients with type 1 diabetes or autoimmune thyroiditis, and the first-degree relatives of CD patients [53, 54]. For example, in siblings of children with CD, who have an overall risk for CD of 10 %, HLA-DQ genotyping was used to stratify ~40 % of the siblings into a group with a small residual risk of <1 % and another ~30 % into a group with a residual risk of 1–10 % [55].

In contrast, the specificity of HLA-DQ genotyping is rather low, with a combined median specificity for HLA-DQ2 and HLA-DQ8 of 54 % [53]. The presence of HLA-DQ2 or HLA-DQ8, in combination with the presence of CD-specific antibodies, could strengthen the diagnosis in patients in whom CD is clinically strongly suspected but in whom no intestinal biopsy will be performed [53]. However, even at a relatively high a priori chance for CD, the proportion of individuals with false-positive results is rather high. Thus, because of the low positive predictive value of HLA-DQ genotyping, a combination with additional risk factors may lead to better prediction of CD risk. A two-step approach could be applied: first excluding individuals without HLA-DQ2 and HLA-DQ8, and second classifying the remaining individuals into different risk groups based on their non-HLA genetic risk factors [16].

In a genetic risk model based on HLA-DQ genotype and ten non-HLA susceptibility SNPs, the presence of 13 or more risk alleles in an individual implied an odds ratio of 6.2 (95 % confidence interval 4.1–9.3) for CD, compared to individuals with 5 or fewer risk alleles [56]. An intermediate HLA-genotype risk combined with 13 non-HLA risk alleles led to an increased risk for CD (odds ratio 6.1) comparable to a high HLA-genotype risk combined with no non-HLA risk alleles (odds ratio 6.2) [56].

The combination of HLA-DQ genotype and the 57 known non-HLA susceptibility SNPs in a genetic risk score leads to a further increase in accuracy of the genetic risk score, with 11 % of the individuals being reclassified to a more accurate risk group. This combination shows a moderate discriminative accuracy with an area under the receiver operating characteristic (ROC) curve of 0.854, corresponding with a chance of 85 % to classify a random CD patient correctly as having a higher risk for the disease than a random individual without CD (J. Romanos et al. unpublished data). However, there is still a large overlap in genetic risk scores between CD patients and healthy individuals, as shown by the percentage of patients and healthy controls classified at intermediate risk: 51 % of the patients and 40 % of the healthy controls. The high-risk category, consisting of the top 25 % of genetic risk scores, has a sensitivity of 43 % and a specificity of 93 %. The positive predictive value for the high-risk category is estimated to be 6 % at a CD prevalence of 1 % and 43 % for a high-risk population with an a priori risk of 10 %. In both situations, the negative predictive value is expected to be higher than 99 %. With a prospective cohort study, the positive and negative predictive values of a genetic prediction model can be estimated more accurately. For example, the PreventCD Study encompasses a European multicenter study among high-risk CD families, in which approximately 1000 newborns who tested positive for HLA-DQ2, and/or HLA-DQ8 will be genotyped in more detail [57].

New insights into genetic risk factors, including their interaction with environmental factors, may contribute to further refining of the prediction models. Genetic data will probably need to be combined with other biomarkers in order to identify subgroups that can usefully guide follow-up and treatment [58].

An important aspect of genetic studies remains the discovery of causal genetic variants and new pathways, including the pathways shared with other immune-related diseases. These may eventually contribute to the identification of new therapeutic targets.

Acknowledgments We thank Jackie Senior for editing the manuscript.

Funding

Work in the Wijmenga group is supported by grants from the Celiac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government (grant BSIK03009 to CW)) and the Dutch Digestive Disease Foundation (MLDS WO11-30 to CW).

References

1. Dubé C, Rostom A, Sy R, Cranney A, Saloojee N, Garritty C, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology*. 2005;128:S57–67.
2. Myléus A, Ivarsson A, Webb C, Danielsson L, Hernell O, Högborg L, et al. Celiac disease revealed in 3 % of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr*. 2009;49:170–6.
3. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet*. 1987;40:1–14.
4. Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet*. 1997;61:307–17.
5. Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, et al. Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet*. 1999;36:687–90.
6. Fasano A, Berti I, Gerarduzzi T, Not T, Colletti RB, Drago S, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med*. 2003;163:286–92.
7. Nisticò L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut*. 2006;55:803–8.
8. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, et al. The first large population based twin study of coeliac disease. *Gut*. 2002;50:624–8.
9. Falchuk ZM, Rogentine GN, Strober W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest*. 1972;51:1602–5.
10. Keuning JJ, Peña AS, van Leeuwen A, van Hooff JP, van Rood JJ. HLA-DW3 associated with coeliac disease. *Lancet*. 1976;1(7958):506–8.
11. Price P, Witt C, Allcock R, Sayer D, Garlepp M, Kok CC, et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol Rev*. 1999;167:257–74.
12. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, et al. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. *Clin Immunol Immunopathol*. 1983;28:395–404.
13. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ α/β heterodimer. *J Exp Med*. 1989;169:345–50.
14. Van Belzen MJ, Koeleman BP, Crusius JB, Meijer JW, Bardoel AF, Pearson PL, et al. Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients. *Genes Immun*. 2004;5:215–20.
15. Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, et al. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc Natl Acad Sci USA*. 2003;100:12390–5.
16. Romanos J, Wijmenga C. Predicting susceptibility to celiac disease by genetic risk profiling. *Ann Gastroenterol Hepatol*. 2010;1:11–8.
17. Lundin KE, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, et al. Gliadin-specific, HLA-DQ(α 1*0501, β 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J Exp Med*. 1993;178:187–96.
18. Lundin KE, Scott H, Fausa O, Thorsby E, Sollid LM. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. *Hum Immunol*. 1994;41:285–91.
19. Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci USA*. 2004;101:4175–9.

20. Henderson KN, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T, et al. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity*. 2007;27:23–34.
21. Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E. HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum Immunol*. 1992;35:188–92.
22. Louka AS, Sollid LM. HLA in coeliac disease: unravelling the complex genetics of a complex disorder. *Tissue Antigens*. 2003;61:105–17.
23. Abadie V, Sollid LM, Barreiro LB, Jabri B. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annu Rev Immunol*. 2011;29:493–525.
24. Liu E, Rewers M, Eisenbarth GS. Genetic testing: who should do the testing and what is the role of genetic testing in the setting of celiac disease? *Gastroenterology*. 2005;128:S33–7.
25. Ahn R, Ding YC, Murray J, Fasano A, Green PH, Neuhausen SL, et al. Association analysis of the extended MHC region in celiac disease implicates multiple independent susceptibility loci. *PLoS One*. 2012;7:e36926.
26. Van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet*. 2007;39:827–9.
27. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet*. 2008;40:395–402.
28. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, Turner G, et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF- κ B signalling. *Gut*. 2009;58:1078–83.
29. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010;42:295–302.
30. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet*. 2011;43:1193–201.
31. Trynka G, Wijmenga C, van Heel DA. A genetic perspective on coeliac disease. *Trends Mol Med*. 2010;16:537–50.
32. Fasano A, Shea-Donohue T. Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nat Clin Pract Gastroenterol Hepatol*. 2005;2:416–22.
33. Forsberg G, Fahlgren A, Hörstedt P, Hammarström S, Hernell O, Hammarström ML. Presence of bacteria and innate immunity of intestinal epithelium in childhood celiac disease. *Am J Gastroenterol*. 2004;99:894–904.
34. Collado MC, Donat E, Ribes-Koninckx C, Calabuig M, Sanz Y. Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease. *J Clin Pathol*. 2009;62:264–9.
35. Kumar V, Wijmenga C, Withoff S. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Semin Immunopathol*. 2012;34:567–80.
36. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13:135–45.
37. The ENCODE Project Consortium. Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799–816.
38. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–89.
39. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–27.
40. Kumar V, Westra H, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet*. 2013;9:e1003201.

41. Hadithi M, de Boer H, Meijer JW, Willekens F, Kerckhaert JA, Heijmans R, et al. Coeliac disease in Dutch patients with Hashimoto's thyroiditis and vice versa. *World J Gastroenterol*. 2007;13:1715–22.
42. Collin P, Reunala T, Pukkala E, Laippala P, Keyriläinen O, Pasternack A. Coeliac disease-associated disorders and survival. *Gut*. 1994;35:1215–8.
43. Ventura A, Magazzù G, Greco L. Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. *Gastroenterology*. 1999;117:297–303.
44. Sategna Guidetti C, Solerio E, Scaglione N, Aimo G, Mengozzi G. Duration of gluten exposure in adult coeliac disease does not correlate with the risk for autoimmune disorders. *Gut*. 2001;49:502–5.
45. Cosnes J, Cellier C, Viola S, Colombel JF, Michaud L, Sarles J, et al. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *Clin Gastroenterol Hepatol*. 2008;6:753–8.
46. Viljamaa M, Kaukinen K, Huhtala H, Kyrönpalo S, Rasmussen M, Collin P. Coeliac disease, autoimmune diseases and gluten exposure. *Scand J Gastroenterol*. 2005;40:437–43.
47. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes*. 2008;57:1084–92.
48. Bao F, Yu L, Babu S, Wang T, Hoffenberg EJ, Rewers M, et al. One third of HLA DQ2 homozygous patients with type 1 diabetes express celiac disease-associated transglutaminase autoantibodies. *J Autoimmun*. 1999;13:143–8.
49. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet*. 2011;7:e1002254.
50. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet*. 2009;10:43–55.
51. Monsuur AJ, de Bakker PI, Zhernakova A, Pinto D, Verduijn W, Romanos J, et al. Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PLoS One*. 2008;3:e2270.
52. Koskinen L, Romanos J, Kaukinen K, Mustalahti K, Korponay-Szabó I, Barisani D, et al. Cost-effective HLA typing with tagging SNPs predicts celiac disease risk haplotypes in the Finnish, Hungarian, and Italian populations. *Immunogenetics*. 2009;61:247–56.
53. Husby S, Koletzko S, Korponay-Szabó IR, Mearin ML, Phillips A, Shamir R, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr*. 2012;54:136–60.
54. Hill ID, Dirks MH, Liptak GS, Colletti RB, Fasano A, Guandalini S, et al. Guideline for the diagnosis and treatment of celiac disease in children: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. *J Pediatr Gastroenterol Nutr*. 2005;40:1–19.
55. Bourgey M, Calcagno G, Tinto N, Gennarelli D, Margaritte-Jeannin P, Greco L, et al. HLA related genetic risk for coeliac disease. *Gut*. 2007;56:1054–9.
56. Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhernakova A, Fu J, et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology*. 2009;137:834–40.
57. Hogen Esch CE, Rosén A, Auricchio R, Romanos J, Chmielewska A, Putter H, et al. The Prevent CD Study design: towards new strategies for the prevention of coeliac disease. *Eur J Gastroenterol Hepatol*. 2010;22:1424–30.
58. Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med*. 2011;365:1612–23.
59. Hindorff LA, MacArthur J (European Bioinformatics Institute), Morales J (European Bioinformatics Institute), Junkins HA, Hall PN, Klemm AK, and Manolio TA. A catalog of published genome-wide association studies. Available at: <http://www.genome.gov/gwastudies>. Accessed 01/15, 2013.