CHAPTER 1

Genetics of Celiac Disease. HLA and Non-HLA Genes

Leticia Plaza-Izurieta, Nora Fernandez-Jimenez, Jose Ramon Bilbao


Immunogenetics.let@gmail.com, joseramon.bilbao@ehu.es

Doi: http://dx.doi.org/10.3926/oms.249

How to cite this chapter

Abstract

Although the mode of inheritance of celiac disease is still unknown, it has been known for a long time that Genetics participates in the susceptibility to the disease. Studies on the prevalence of CD in affected families, and especially those comparing twin pairs, have been very useful to estimate the proportion in which environmental and genetic factors contribute to the development of this complex disorder. According to these studies, Genetics is a fundamental player both in the triggering and in the latter development of CD.

In general, it is well accepted that the proportion of monozygotic or identical twins concordant for CD is around 75-86%, while in the case of dizygotic twins, this proportion is reduced to 16-20%. This difference between mono- and dizygotic twins has allowed scientists to estimate the genetic component of CD, which is higher than what has been calculated for other immunological complex diseases, such as type 1 diabetes (T1D) (around 30% concordance in monozygotic and 6% in dizygotic twins)\(^1\). Moreover, concordance rates between sib pairs and dizygotic twins are almost the same, indicating that the environmental component has a minimum contribution to the risk of developing CD.

In summary, accumulated evidence suggests that CD has a very strong genetic component and it has been calculated that the heritability of this disease (proportion of the risk of suffering from CD attributable to genetic factors, compared to environmental determinants) is around 87%\(^2\). The largest portion of the genetic risk to develop CD comes from the presence of certain Human Leucocyte Antigen (HLA) alleles. However, even if the role of these HLA molecules is essential in the pathogenesis of the disease, their contribution to the heredity is
modest, and thus, it has been hypothesized on the existence of many small effect, non-HLA susceptibility loci.

**Keywords**

Celiac disease, autoimmune disease, immune-mediated disease, HLA, linkage studies, genome-wide association studies (GWAS), gene expression, pathway analysis.
1. HLA Region and Celiac Disease

1.2. HLA Region

HLA is the name for the Mayor Histocompatibility Complex (MHC) in humans; it is a super locus located on the chromosomal region 6p21 and contains a large number of genes related to the immune response. HLA genes encode antigen presenting proteins that are expressed in most human cells and are essential for the ability of the organism to distinguish between self and foreign molecules.

HLA genes are involved in many inflammatory and autoimmune disorders and also contribute to the susceptibility to develop infectious diseases such as AIDS or malaria. However, due to the high genetic complexity of the region, most of the particular genetic factors and pathogenic mechanisms underlying the susceptibility to each of these disorders remain unknown. In fact, the HLA region presents the highest genic density of the entire genome and a very strong gene expression seems to be favored.

1.2. Contribution to the Genetic Risk and Susceptibility Genes

As previously mentioned, the HLA region is the most important susceptibility locus in CD and explains around 40% of the genetic component of the disease. The first evidence supporting the association between HLA and CD was published in 1973 and was detected using serological methods. Due to the strong linkage disequilibrium present in the area, initial studies identified HLA-A1, HLA-B8 and HLA-DR3 as the etiological variants in the region, but subsequent molecular studies have revealed that the factors directly implicated are the HLA class II genes encoding both HLA-DQ2 and -DQ8 molecules (Figure 1). The strongest association has been found with HLA-DQ2, and 90% of celiac patients present at least one copy of the HLA-DQ2.5 heterodimer (formed by the combination of the products of DQA1*05 and DQB1*02 alleles, that encode the α and β chains of the
heterodimer, respectively). On the other hand, 20-30% of the non-celiac population also presents this HLA-DQ2 variant, making it clear that, even though it is very important, it is not sufficient to develop the disease. Most of the patients who do not carry the HLA-DQ2 genotype are HLA-DQ8 carriers and so have at least one copy of the haplotype containing DQA1*03:01 and DQB1*03:02 alleles\(^5\). A very small portion of the patients are negative for both DQ2 and DQ8, but it has been observed that in these few cases, individuals present at least one of the two alleles encoding the DQ2 molecule (DQA1*05 or DQB1*02)\(^6,7\).

![Figure 1. Association of the HLA locus with CD. HLA-DQ2 molecule is the major factor conferring risk to CD. Most celiac patients express the heterodimer HLA-DQ2.5, encoded by the alleles HLA-DQA1*05 (\(\alpha\) chain) and HLA-DQB1*02 (\(\beta\) chain), that can be present in cis in the DR3-DQ2 haplotype or in trans, in the heterozygotes DR5-DQ7 and DR7-DQ2.2. The HLA-DQ2.2 dimer, a variant of HLA-DQ2 encoded by the alleles HLA-DQA1*02:01 and HLA-DQB1*02:02, confer a low risk to develop the disease. Most of the patients that are negative for DQ2 express HLA-DQ8, encoded by the DR4-DQ8 haplotype\(^5\).]
HLA-DQ2 and -DQ8 variants are in linkage disequilibrium with DR3 and DR4, respectively. Thus, we often refer to these risk variants as DR3-DQ2 and DR4-DQ8 haplotypes. In several haplotypes, as is the case of DR3-DQ2, the two alleles of the HLA-DQ2.5 heterodimer (DQA1*05:01 and DQB1*02:01) are located in the same chromosome and therefore, encoded in cis. In the heterozygous individuals carrying DR5-DQ7 and DR7-DQ2 haplotypes, the two molecules taking part in the risk heterodimer are encoded in trans, because they are located in different chromosomes. The differences between these two types of HLA-DQ2.5 rely on a single amino acid of the DQα chain (DQA1*05:01 vs. DQA1*05:05) and another residue of the membrane region of the DQβ chain (DQB1*02:01 versus DQB1*02:02), but they seem not to have any functional consequences and are associated with a similar risk effect. However, the risk conferred by another HLA-DQ2 variant, the HLA-DQ2.2 dimer, is very low.

There is also a relationship between the degree of susceptibility to CD and the number of DQ2.5 heterodimers. Homozygous individuals with two DR3-DQ2 haplotypes as well as heterozygous patients presenting DR3-DQ2/DR7-DQ2 express the highest levels of DQ2.5 heterodimers and thus, confer the maximum genetic risk to develop CD. In this sense, it has to be mentioned that patients with refractory CD (those not responding to GFD) present a higher degree of homozygosity for DR3-DQ2 (44-62%) than other celiac patients (20-24%). A similar dose-dependent effect has also been suggested for DQ8 molecules.

Apart from the genes encoding DQ molecules, the HLA region also contains many other genes that participate to the immune response and that could contribute to the susceptibility to CD. Several studies have postulated that polymorphisms in genes such as MICA, MICB or TNF could contribute to the genetic risk to develop this disorder. Nonetheless, most of these works have not paid enough attention to the strong linkage disequilibrium among genes and results are not conclusive. Deep-sequencing and exhaustive mapping of the region will help to determine whether it contains susceptibility factors other than HLA-DQ. Although HLA genes importantly contribute to the
genetic susceptibility, the concordance of the disease in siblings identical for HLA genotype approaches only 30%, so that we can conclude that HLA genes are important but not sufficient to develop CD\textsuperscript{10}.

### 1.3. Role of HLA in the Pathogenesis of CD

The strong association of the HLA class II genes with CD is directly linked to the fundamental role of CD4+ T lymphocytes in the pathogenesis of the disease. In fact, CD4+ T cells that are able to recognize gluten-derived peptides are present in the intestinal mucosa of celiac patients, but not in the case of healthy, non-celiac individuals. When genetically susceptible individuals are exposed to certain gluten-derived epitopes, they are presented by the HLA-DQ2/HLA-DQ8 molecules on the surface of antigen presenting cells (APC), stimulating the proliferation of gluten-specific CD4+ T cells\textsuperscript{14}.

An important landmark in the molecular basis underlying the association between HLA and CD was the discovery that the binding capacity between the HLA-DQ2 and/or -DQ8 and the gliadin peptides increases substantially when the latter have been enzymatically modified by the enzyme tissue transglutaminase type 2, or TG2. The enzyme catalyzes a reaction that provokes the increase of negative charges in the gluten-derived peptides, favoring their binding to certain HLA molecules (DQ2 and DQ8) and thus, triggering the presentation of these gluten peptides to CD4+ T cells.

Given the importance of HLA molecules in the activation of auto-reactive gluten-specific T cells, it is expected that any modification in their coding sequence will provoke alterations in different steps of this process. In this way, polymorphisms in the sequence encoding the antigen binding sites could affect affinity, favoring or hampering the recognition of the gluten-derived peptides\textsuperscript{15}. On the other hand, several polymorphisms located in regulatory sites can repress or enhance the expression of the HLA molecules, reducing or augmenting the immune response to gluten.
2. Search for Genetic Susceptibility Genes in CD

Given the fact that HLA alone can only explain around 40% of the genetic component of CD, large efforts have been done to localize and identify non-HLA susceptibility genes that could clarify the complex genetics of this disorder. Two have been the major strategies used with this aim: on the one hand, linkage studies in affected families, and on the other hand, association studies based on population screening. More recently, CD has also been studied using Genome Wide Association Studies (GWAS), in which thousands of Single Nucleotide Polymorphisms (SNP) have been genotyped and analyzed. These studies have allowed us to identify several associated loci, but functional studies will be needed to confirm the implication of the proposed candidate genes.

2.1. Linkage Studies

Linkage studies in families have allowed the identification of chromosomal regions which are repeatedly and consistently inherited by the affected members of a family through several generations. Thus, regions potentially relevant to the development of the disease can be selected and fenced in. Genes localized in these regions are considered positional candidates, due to the fact that it is their position in the genome that is conferring them the candidate identity. In the case of CD, apart from the HLA region (or CELIAC1) which obviously is the most consistently replicated signal and the one showing the strongest linkage disequilibrium, three regions containing positional candidates such as a number of interleukins, the SPINK family, CD28, CTLA4, ICOS and MYO9B have been described in the different linkage studies (Figure 2). However, even though consistently replicated in several studies, the certain causes of association with CD have not been identified for these linkage regions.
2.2. Functional Candidate Genes

The candidate gene approach for genetic association studies focuses on associations between genetic variation within pre-specified genes of interest and phenotypes or disease states. This is in contrast to GWA studies, which scan the entire genome for common genetic variation. Candidate genes are most often selected for study based on a priori knowledge of the gene's biological functional impact on the trait or disease in question. This approach has been commonly used in complex disease studies, and also in CD.

Most of the candidate gene studies to date has focused on immune response, since it is generally accepted that CD is a T cell mediated disease, in which gliadin-derived peptides, either in native form or deamidated by transglutaminase, activate lamina propria infiltrating T lymphocytes, leading to both Th1 and Th17 inflammatory responses of the adaptive immune system\textsuperscript{19}. Thus, both the most Th1 response-characteristic cytokine INFγ (encoded by $\text{INFG}$)\textsuperscript{20} and $\text{IL23R}$ the receptor of the best known interleukin in the Th17 cascade\textsuperscript{21-23} have been studied among many others, with not many strong association evidences as conclusion.

During the last decade, however, a growing interest has focused on the possible implication of the innate immune response, based on the fact that gliadin peptides are also able to trigger a non-T-cell-dependent response that could establish the proinflammatory environment necessary for subsequent
T-cell activation and tissue destruction\textsuperscript{24}. Different innate immune genes and gene families have been proposed as putative susceptibility candidates to CD such as the inflammatory mediators \textit{IL1A}, \textit{IL1B}, \textit{IL1RN}, \textit{IL18}, \textit{RANTES} and \textit{MCP1}\textsuperscript{25}, the Killer Immunoglobulin-like receptor (KIR) family\textsuperscript{26}, the Toll-like receptor (TLR) family\textsuperscript{27,28} and the stress molecules \textit{MICA} and \textit{MICB}\textsuperscript{29} but although a general activation of the innate immune system is well known to occur in CD, none of the proposed candidates have shown a strong association with the disease.

Finally, functional players involved in the remodeling of the intestinal epithelia and in the maintenance of the extracellular matrix have also been proposed as putative susceptibility genes, but again, no association has been confirmed for any of them.

\subsection*{2.3. Genome-wide association and follow-up studies in CD}

Hundreds of SNPs have been identified thanks to the Human Genome sequencing projects. Some of those SNPs, called tag SNPs, have been used as genetic markers in GWAS and allow the identification of thousands of susceptibility variants for many complex diseases. The two GWAS performed in CD, together with several follow-up studies, revealed a total of 26 non-HLA associated regions\textsuperscript{30-32}. The most recent large-scale project performed to identify variants associated with CD and other autoimmune diseases is the Immunochip Project, in which a denser genotyping of 186 GWAS loci associated with 12 immune-related diseases identified 13 additional regions associated with CD\textsuperscript{33}.

Hence, there is a total of 39 non-HLA regions associated with CD, containing 57 independent association signals. Nineteen of those regions pinpoint to a single candidate gene, but only 3 associated SNPs are linked to protein-altering variants located in exonic regions, although some potentially causative genes have been proposed due to the existence of signals near the 5' or 3' regulatory regions(Figure 3).

Even though most SNPs localize to nonprotein coding intergenic and intronic regions, CD associated variants seem to be located in expression
quantitative trait loci or eQTLs, genomic loci that regulate expression levels of mRNAs or proteins. When eQTLs map to a genomic location close to the regulated gene they are referred to as cis-eQTLs; in contrast, when the eQTL maps far from the gene (even on different chromosome), it is referred to as trans-eQTL. After a meta-analysis of a genome-wide eQTL dataset of 1,469 human whole blood samples, supposed to reflect primary leukocyte gene expression, 38 genome-wide CD associated non-HLA loci were assessed for cis expression-genotype correlation. Twenty significant eQTLs were identified, more than expected by chance, indicating that CD associated regions are greatly enriched for eQTLs. These data may indicate that some risk variants could have an influence in CD susceptibility by altering gene expression, however, there are many evidences indicating that cis-eQTLs differ between different tissues and can even have completely opposite effects.

Thereby, it is important to perform functional analysis of the proposed candidate genes in the disease tissue. The eight association peaks from the first CD GWAS were replicated in a Spanish population in 2011, identifying four genes (IL12A, LPP, SCHIP1 and SH2B3) whose expression in the intestinal mucosa varied according to disease status and the genotype of the associated variant. These results suggest that these genes may be constitutively altered in celiac patients, probably before the onset of observable symptoms of the disease, and therefore could have a primary role in its pathogenesis.
A second work took a step forward and identified two genes (PTPRK and THEMIS), located in the same associated region, which were co-expressed both in active disease and in response to \textit{in vitro} stimulation by gliadin of intestinal biopsies of celiac patients with inactive disease who have adhered to the gluten free diet for at least two years\textsuperscript{35}. Therefore, it seems that associated variants in this region affect the expression of different genes, but not constitutively from the time of birth of the future celiac patient, but only in the presence of a toxic stimulus that triggers an immune response. The implications of this finding are of great importance because they highlight the existence of common regulatory mechanisms for different genes in the DNA
sequence that only have an effect in the presence of a disease-provoking immunogenic stimulus.

In order to elucidate the substantial fraction of heritability that remains unexplained in most complex diseases, a novel hypothesis has recently been postulated. It has been called the “rare-variant synthetic genome-wide-association hypothesis” and it is based on the assumption that unobserved rare causal variants lead to association detected at common tag variants. However, a recent work in which sequencing and genotyping for coding exons of 25 GWAS risk genes were performed in 41,911 UK residents of white European origin (24,892 subjects with six autoimmune disease phenotypes and 17,019 controls) has revealed that rare coding-region variants at known loci have a negligible role in common autoimmune disease susceptibility, including CD.

A different approach was taken to fine map the LPP locus in the search for possible functional variants. This strategy revealed 6 SNPs that overlap regulatory sites, with rs4686484 having a possible effect on LPP gene expression in patients.

Almost all associated regions contain genes with an immunological function, many of which act in the same biological pathways. T-cell development in the thymus, a pathway previously not explored in CD pathogenesis, is one of those pathways. A study carried out by Amundsen et al. aimed to explore the regulatory potential of the CD-associated SNPs by eQTL analysis in thymic tissue. They found 43 nominally significant (p<0.05) eQTLs within 24 CD-associated chromosomal regions, corresponding to 27 expression-altering SNPs and 40 probes that represent 39 unique genes. When compared across different tissues, they found that 14 eQTLs could represent potentially novel thymus-specific eQTLs. This implies that CD risk polymorphisms could affect gene regulation in the thymus.

Given the diversity of cell types and specialization of functions within the immune system, Xinli Hu et al. studied genetic and cellular traits of CD4+ effector memory T (CD4+ TEM) cells, which are particularly important in the onset of CD. They purified CD4+ T cells form a cohort of healthy
individuals and assayed genome-wide SNPs, abundance of CD4+ TEM cells in blood, proliferation upon T cell receptor stimulation, and 215 gene transcripts both in resting and stimulated states. They found that expression levels of 46 genes were regulated by nearby SNPs, including disease-associated SNPs. Many of these eQTLs had not been previously observed in studies of more heterogeneous peripheral blood cells, however they were not able to demonstrate that disease alleles confer risk by modulating these traits in this particular cell type.

The last work published in this field tried to scrutinize the functional implication of 45 candidate genes that were not studied in previous works\(^{40}\) (Figure 3). The expression of those genes was analyzed in the disease tissue of celiac patients at diagnosis and after treatment, and compared to non-celiac controls. Moreover, the SNP genotype effect in gene expression was also investigated and coexpression analyses were performed. Several genes showed differential expression among disease groups, most of them related to immune response. Multiple *trans*- but only 4 *cis*- eQTLs were found, and surprisingly the genotype effect seems to be stimulus dependent as it differs among groups. Coexpression levels vary from higher to lower levels in active patients at diagnosis, treated patients and non-celiac controls respectively. A subset of 18 genes tightly correlated in both groups of patients but not in controls was identified. Interestingly, this subset of genes was influenced by the genotype of 3 SNPs. These results strongly suggest that the effects of disease-associated SNPs go far beyond the oversimplistic idea of transcriptional control at a nearby locus.

In conclusion, recent studies stress the need of developing functional studies and the importance of avoiding arbitrary selection of susceptibility candidate genes. Additionally, they reveal the huge work that remains to be done in order to identify the elements underlying the complex regulatory system of the genome, while opening the door to future studies, in which the scientific community will need to exhaustively analyze both different classes of variation (such as structural variants of the genome or epigenetic features) and the vast non-coding genome, in order to shed light on the complex
Genetics of common disorders and to be able to understand the effect of the disease-associated variants found by the numerous GWA studies.

3. Novel Approaches to Unravel the Genetics of CD

A unique Copy Number Variation (CNV) study has been performed in CD. In this work, *TLR2*, *TLR4*, and the β-defensin cluster (*DEFB4*, *DEFB103* and *DEFB104*) were analyzed by gene-specific, real-time PCR in 376 CD patients and 376 controls. TLR genes did not show CNV, and all samples presented with two copies. β-defensin clusters varied between 2 and 9 copies per genome, and when grouped into bins, high copy numbers (>4) were underrepresented among patients, suggesting that increased copy numbers could protect from CD, possibly by impeding bacterial infiltration more efficiently and preserving gut epithelial integrity.

On the other hand, genome-wide expression analysis have consistently been used to draw maps of the most common functional alterations in different complex diseases. Sometimes, these approaches have also been used to identify associated variants that could explain different pathological situations. In the case of CD, Castellanos-Rubio et al. designed in 2008 a strategy that combined gene expression profiling of intestinal biopsy specimens, linkage region information, and different bioinformatics tools for the selection of potentially regulatory single-nucleotide polymorphisms. Among other results, they found evidence of association with several SNPs and identified *SERPINE2* in 2q33, and *PBX3* or *PPP6C* in 9q34 as potential role players in the development of the disease.

Following the results from the ENCODE project, it is now known that a substantial fraction of genetic variants contributing to complex traits in humans are involved in gene regulation. Most phenotype-associated variants discovered in GWA studies are far away from protein coding regions, and even appear in gene deserts. This distribution is similar to that shown by most of the cis regulatory modules such as promoters and enhancers, and it is expected that many variants associated with complex traits may affect gene
expression. Furthermore, virtually any noncoding sequence in the human genome could potentially be a regulatory element and even act far away from its genomic location and globally alter whole pathways and signaling routes. Thus, oversimplistic and arbitrary selection of nearby and single candidate genes should be avoided and Systems Biology approaches should be implemented to find the relations and common regulatory mechanisms conjugating the genes that interact to generate the celiac phenotype.

In this sense, in a genome-wide expression microarray carried out several years ago, some signaling routes were found to be altered in CD, such as the Jak-Stat, NFκB, MAPK or TGFB pathways. Some of the genes participating in these routes have been studied to determine whether they contain CD-associated variants. One of these genes is STAT1, whose expression is altered in the disease. However, no associated SNPs have been found.

NFκB1 has also been studied but, although it is constitutively active in the intestinal mucosa of CD patients, it does not seem to contain any genetic alteration that could explain its overexpression. It has been suggested that the pathogenic effects assigned to this transcription factor (TF) could be caused by a regulatory defect and that variants or alterations in genes upstream NFκB could trigger the enhanced transcriptional activity observed in CD. It has been speculated that two of the genes identified in a follow-up study after the GWAS (REL and TNFAIP3) could underlie the deregulation of this biological route. A regulatory SNP in the UBD gene that is involved in the activation of NFκB has been associated to CD in Spanish population. This gene is overexpressed in active disease and the allelic distribution of the associated polymorphism presents a significant correlation with expression levels.

In this context, a recent study that tried to normalize the altered expression of the NFκB pathway in vitro using a MALT1 paracaspase inhibitor discovered a strong coexpression among genes of the route in healthy gut mucosa, while intestinal biopsies from active CD patients presented a completely deregulated pathway (Figure 4). This disruption of coexpression persisted in treated, inactive patients, especially after acute gliadin
stimulation *in vitro*, and could be reverted to a regulated pattern similar to the one seen in controls through MALT1 inhibition. These results strongly suggest that unknown regulatory mechanisms behind the tight coexpression of the NFκB pathway observed in non-inflamed gut mucosa could be the ones affected by putative genetic or epigenetic alterations rather than single genes taking part in the activating cascade.

![Gene expression matrixes for different disease statuses](image)

*Figure 4. Gene pair coexpression matrixes for the different disease statuses. Each small square represents the p value for the correlation of the expression level in a specific gene pair. White, light gray, dark gray and black indicate Pearson's correlation p values >0.05, <0.05, <0.01 and <0.001, respectively.*

Other recent pathway analysis includes the study of genes whose expression was previously shown to be altered in celiac disease and that shared “angiogenesis” GO terms. A regulatory polymorphism mapping to *TNFSF13* was shown to be associated with CD, and several antiangiogenic genes such as *TGM2* and *PML* were found to be upregulated, while some proangiogenic genes were notably downregulated. Another study has confirmed the involvement of tight junction genes related to permeability, polarity, and cell proliferation in the epithelial destruction observed in CD. Coexpression patterns of several genes of the tight junction pathway support the idea of a common regulatory mechanism that seems to be altered in active CD. In general, GFD normalization confirms the reversibility of the process, except for the constitutive downregulation of *PPP2R3A*, suggestive of a genetic implication.
4. Preliminary steps on the epigenetics of CD

For the moment, only a few attempts have been performed to unravel the epigenetic landscape of CD. However, taking into account that genetic and epigenetic variation, together with environmental factors that shape expression and methylation patterns are known to underlie the vast complexity of common disorders, it is probable that epigenetic studies will increase in the following years.

As far as we know, the first miRNA expression analysis in CD was performed in 2011 by Capuano et al. In this study, they tested the expression of a large set of miRNA molecules and found out that nearly the 20% was differentially expressed when celiac patients were compared to control individuals. Moreover, they discovered that high miR-449a levels targeted and reduced NOTCH1 signaling and suggested that NOTCH pathway could be constitutively altered in the celiac small intestine due to the overexpression of this miRNA, and therefore, could drive the increased proliferation and the decreased differentiation of intestinal cells towards the secretory goblet cell lineage.

On the other hand, DNA methylation has also been studied in the context of CD. It is known that methylation of cytosines, usually at CpG dinucleotides, is involved in epigenetic regulation of gene expression. Promoter methylation is typically associated with repression, whereas genic methylation correlates with transcriptional activity. It has been recently found that 96% of CpGs exhibit differential methylation in at least one cell type or tissue assayed and that levels of DNA methylation correlate with chromatin accessibility. Additionally, chronic inflammation have been linked to several epigenetic alterations. Thus, methylation level was measured in several NFκB-related genes in celiac active and inactive mucosa and compared to control, non-inflamed tissue. Surprisingly, partially reversible, subtle but still significant methylation differences were found in active celiac biopsies and disease samples showed significant correlations among the methylation levels.
of different genes (co-methylations). These relationships seemed to somehow disrupt the coexpression patterns observed in health among those same genes.

The ENCODE project has also been able to find CpGs with allele-specific methylation consistent with genomic imprinting, and determined that these loci exhibit aberrant methylation in cancer cell lines. Very recently, Hutchinson and collaborators hypothesized that the phenomenon of allele-specific methylation may underlie the phenotypic effects of multiple variants identified by genome-wide association studies, so that they evaluated this in an initial screen at up to 380,678 sites within the genome\textsuperscript{54}. They showed that many of the cis-regulated allele-specific methylation variants are also eQTLs in peripheral blood mononuclear cells and monocytes and/or in high linkage-disequilibrium with variants associated to complex disease. Finally, they found out that, among others, the CD-associated SNP rs2762051, was associated to one of such methylation variants, opening the door to a novel way to relate the epigenetic, non-coding variation to the GWAS-derived results.
References


Genetics of Celiac Disease. HLA and Non-HLA Genes

http://dx.doi.org/10.1038/nri885  
PMid:12209133

http://dx.doi.org/10.1038/sj.gene.6364061  
PMid:15014431

http://dx.doi.org/10.1111/j.1399-0039.1993.tb01998.x  
PMid:8362409

http://dx.doi.org/10.1084/jem.178.1.187  
PMid:8315377

http://dx.doi.org/10.1053/j.gastro.2009.09.008  
PMid:19766641

http://dx.doi.org/10.1038/nature07524  
PMid:19037317 PMCid:PMC3784325

http://dx.doi.org/10.1086/301754  
PMid:9497251 PMCid:PMC1376948

http://dx.doi.org/10.1111/j.1399-0039.2004.00189.x  
PMid:14989710

http://dx.doi.org/10.1016/S0016-5085(03)01205-8
http://dx.doi.org/10.1080/08916930802350789
PMid:19127457

http://dx.doi.org/10.1016/j.jaut.2004.05.004
PMid:15324937

http://dx.doi.org/10.1111/j.1572-0241.2007.01660.x
PMid:18047540

http://dx.doi.org/10.1186/1471-2350-10-8
PMid:19175939 PMCID:PMC2642807

http://dx.doi.org/10.1371/journal.pone.0031244
PMid:22350581 PMCID:PMC3281077

http://dx.doi.org/10.1016/S0140-6736(03)13803-2

http://dx.doi.org/10.1186/1471-2350-6-29
PMid:16078996 PMCID:PMC1190178

http://dx.doi.org/10.1038/sj.gene.6364367
PMid:17215859
http://dx.doi.org/10.1097/MPG.0b013e31816de885 
PMid:18728522

http://dx.doi.org/10.1111/j.1399-0039.2007.00945.x 
PMid:17927684

http://dx.doi.org/10.1007/s00251-004-0724-8 
PMid:15490153

http://dx.doi.org/10.1038/ng.102 
PMid:17558408 PMCid:PMC2274985

http://dx.doi.org/10.1038/ng.102 
PMid:18311140 PMCid:PMC2673512

http://dx.doi.org/10.1038/ng.543 
PMid:20190752 PMCid:PMC2847618

http://dx.doi.org/10.1038/ng.998 
PMid:22057235 PMCid:PMC3242065

http://dx.doi.org/10.1136/jmg.2011.089714 
PMid:21490378


http://dx.doi.org/10.1038/nature11247
PMid:22955616 PMCid:PMC3439153

http://dx.doi.org/10.1074/jbc.R112.352427
PMid:22952232 PMCid:PMC3438926

http://dx.doi.org/10.1073/pnas.0903103106
PMid:19474294 PMCid:PMC2687147

http://dx.doi.org/10.109/08916930903225229
PMid:19814655

http://dx.do.org/10.1038/sj.ejhg.5201667
PMid:16773129

http://dx.oi.org/10.1136/gut.2008.169052
PMid:19240061

http://dx.doi.org/10.1016/j.humimm.2009.09.359
PMid:19808075

http://dx.oi.org/10.1093/hmg/ddt520
PMid:24163129 PMCid:PMC3919015

http://dx.doi.org/10.3109/08916934.2011.637531
PMid:22136669
http://dx.doi.org/10.1097/mpg.0000000000000338

http://dx.doi.org/10.1371/journal.pone.0029094
PMid:22194996 PMCid:PMC3240641

http://dx.doi.org/10.1371/journal.pone.0098464
PMid:24911414 PMCid:PMC4049588