CHAPTER THREE

Profiling Celiac Disease-Related Transcriptional Changes

Ainara Castellanos-Rubio, Jose Ramon Bilbao[1]
University of the Basque Country (UPV/EHU), BioCruces Health Research Institute, Leioa, Basque Country, Spain
[1]Corresponding author: e-mail address: joseramon.bilbao@ehu.eus

Contents

Abstract

Celiac disease (CD) is a chronic, autoimmune disease of the small intestine with a strong but complex genetic component. The disease is triggered by the consumption of dietary gluten through the presentation of immunogenic gliadin peptides to T helper lymphocytes by HLA-DQ2 and DQ8 heterodimers, which are the major contributors to the genetic risk. Recent large-scale genotyping efforts have identified a large number of additional association signals, but the functional role of the underlying genes in the pathogenesis of the disease is still unclear. In the last years, several whole transcriptome analyses have been performed in different tissues, including whole duodenal biopsies, isolated gut epithelial cells or peripheral blood from gluten-consuming CD patients at diagnosis, treated patients on gluten-free diet and nonceliac controls, sometimes after in vitro challenge with gluten-derived gliadin peptides. Although the results from the different experiments are difficult to reconcile, the main findings point to an exacerbation of the immune response, together with the dysregulation of signaling and cell cycle pathways. The effect of associated SNPs on the expression of candidate genes and the role of the noncoding genome are the new territories that the CD research community has only begun to explore.

International Review of Cell and Molecular Biology, Volume 336
ISSN 1937-6448
https://doi.org/10.1016/bs.ircmb.2017.07.003

1. CELIAC DISEASE

Celiac disease (CD; OMIM 212750) or gluten-sensitive enteropathy is a complex chronic, immune-mediated disease that affects roughly 1%–2% of the population and develops in genetically susceptible individuals in response to gluten proteins from wheat, barley, and rye. Clinical symptoms arise from the gluten-induced lesion of the small intestinal mucosa that results in villous atrophy and crypt hyperplasia, and typically include diarrhea, abdominal distension, or vomiting, as well as failure to thrive or psychomotor impairment, as a consequence of nutrient malabsorption. Less frequently neurological events, dental enamel defects, infertility, osteoporosis, joint pain, and elevated liver-enzyme concentrations may occur. The only effective treatment available for CD is life-long removal of gluten from the diet, and in fact, in the great majority of CD individuals, symptoms resolve after several months on a gluten-free diet (GFD), while the small intestinal architecture recovers and circulating gluten-related autoantibodies against tissue transglutaminase TG2 and endomysium clear (Husby et al., 2012).

CD has a strong association with certain human leukocyte antigen (HLA) class II molecules, and virtually all people with CD carry the HLA-DQ2 (encoded by HLA-DQA1*05 and HLA-DQB1*02) or HLA-DQ8 (HLA-DQA1*03-HLA-DQB1*0302) heterodimers (Sollid et al., 1989; Spurkland et al., 1992). There is functional evidence supporting the central role of the HLA-DQ2 and HLA-DQ8 molecules in the pathogenesis of the disease, because CD4+ T lymphocytes isolated from the intestinal mucosa of celiac patients can recognize gluten-derived peptides presented by these molecules (Lundin et al., 1993; Molberg et al., 1997), particularly when those peptides have been previously deamidated by the enzyme TG2. It is estimated that the proportion of the genetic risk attributable to these classical HLA class II variants is around 23% (Gutierrez-Achury et al., 2015). HLA genes are located on the major histocompatibility complex (MHC), a region of chromosome 6p21 harboring many genes involved in the immune response that is characterized by strong linkage disequilibrium (LD) and extended conserved haplotype blocks. A recent fine-mapping study across the entire MHC region has identified additional risk variants on both sides of the HLA-DQ genes, in the more telomeric HLA class I region and in the more centromeric HLA-DPB1 gene, and these new variants would be responsible for an additional 18% in the inheritance of the CD (Gutierrez-Achury et al., 2015). Outside the HLA region, two GWAS

conducted in 2007 and 2010 analyzed approximately 5000 patients and 10,000 controls and identified 26 CD-associated *loci* (Dubois et al., 2010; Hunt et al., 2008; van Heel et al., 2007). One year later, some 12,000 cases and 12,000 controls were genotyped using the Immunochip (a dedicated fine-mapping platform covering 186 loci with evidence of association with 10 immune-related disorders) and another 13 new susceptibility regions were identified, making a total of 39 non-HLA loci associated with the genetic risk of CD (Trynka et al., 2011). All the associated SNPs are common polymorphisms (minor allele frequency in the population above 5%) with modest effects on the risk of developing the disease (OR: 1.2–1.5). The implication of rare variants with clear functional consequences and major effects on risk (like coding mutations) has not been demonstrated except for a nonsynonymous variant in the *NCF2* gene (rs17849502) associated with a small increase in risk (Hunt et al., 2013). Moreover, the vast majority of SNPs are located outside protein-coding sequences, so they are assumed to play a regulatory role in the expression of nearby genes or even genes located elsewhere in the genome. Understanding the biological consequences of associated polymorphisms is a very complicated task, which is still far from being fully resolved (Farh et al., 2015).

Diagnostic protocols have traditionally relied upon intestinal biopsies as the gold standard for the confirmation of CD and even for patient follow-up, so that this very valuable material has been readily available for investigations aimed to identify altered molecular mechanisms in the disease tissue. This has encouraged the scrutiny of the genomic contribution to the disease from the transcriptional point of view, and a number of whole genome expression analyses have been performed in the last years. Starting from custom microarrays to more recent RNA sequencing approaches, the general aim has been to take an unbiased look at alterations in the expression patterns of single genes, groups of coregulated genes and pathways that are dysregulated in the pathological processes. Together with the identification of altered mRNA levels, the expression of noncoding RNAs like long noncoding RNAs (lncRNAs) and microRNAs (miRNAs) is also possible (Fig. 1).

2. MICROARRAY STUDIES

Microarray-based technology has been used for more than a decade in the study of CD, and it has been employed to answer different questions in distinct experimental scenarios (Table 1). The first expression microarray study in CD (Juuti-Uusitalo et al., 2004) was carried out using a cDNA array

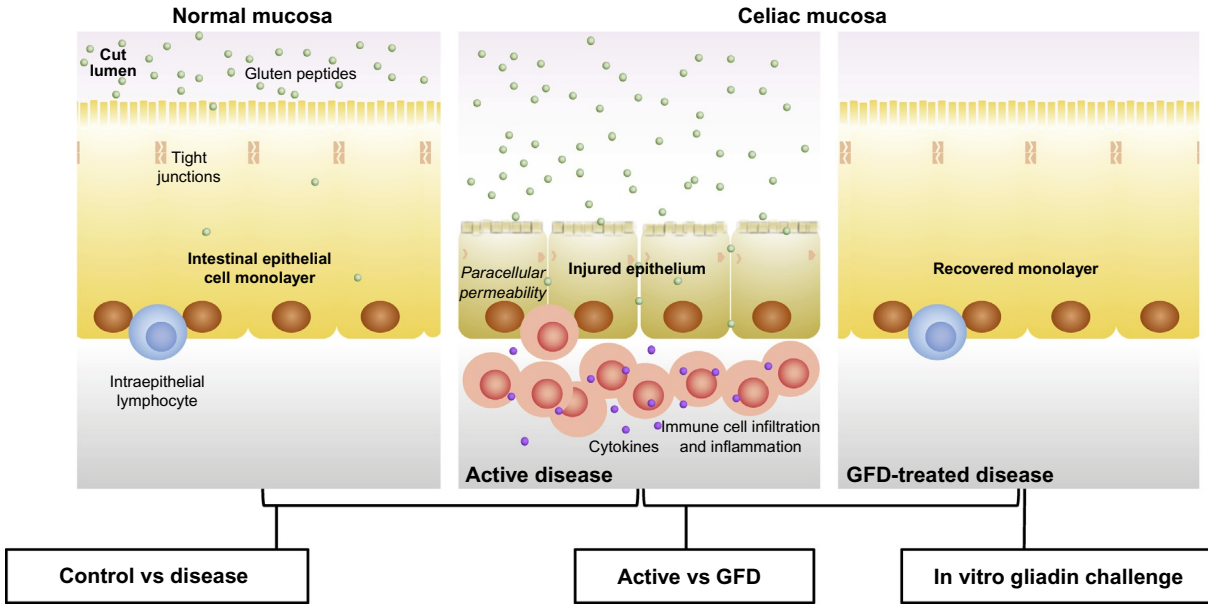

Fig. 1 Celiac disease transcriptome analyses performed so far in the context of disease state and biological sample. The number in brackets indicate the reference where the original study was published.

Table 1 Most Relevant Biological Pathways That Have Been Found to be Altered in Celiac Disease Through Different Transcriptome Analyses (Black Squares)

	References							
Altered Pathway/ Process	**Juuti-Uusitalo et al. (2004)**	**Diosdado et al. (2004)**	**Castellanos-Rubio et al. (2010a)**	**Juuti-Uusitalo et al. (2007)**	**Parmar et al. (2013)**	**Bracken et al. (2008)**	**Quinn et al. (2015)**	**Magni et al. (2014)**
Metabolism	■	■		■		■	■	
Cell cycle, DNA control, and apoptosis	■	■	■	■		■		
Transcription and translation	■	■		■		■		
Cell communication	■	■	■			■		
Signal transduction	■	■	■	■		■	■	
EFG signaling				■				
Jak–Stat signaling			■				■	
NFκB signaling			■					
TGFβ signaling			■				■	
TLR signaling							■	
MAPK signaling			■				■	

Continued

Table 1 Most Relevant Biological Pathways That Have Been Found to be Altered in Celiac Disease Through Different Transcriptome Analyses (Black Squares)—cont'd

	References							
Altered Pathway/ Process	**Juuti-Uusitalo et al. (2004)**	**Diosdado et al. (2004)**	**Castellanos-Rubio et al. (2010a)**	**Juuti-Uusitalo et al. (2007)**	**Parmar et al. (2013)**	**Bracken et al. (2008)**	**Quinn et al. (2015)**	**Magni et al. (2014)**
Protein synthesis and modification	■	■				■		
Immune and inflammatory response related	■	■	■			■	■	■
Cellular and extracellular organization	■		■	■				
Transport	■			■		■		
Antibody production							■	

that interrogated 5188 test sequences (including known genes and expressed sequence tags of unknown function) in four jejunal biopsies from each comparison group of active CD patients, patients on GFD, and nonceliac control individuals. Altogether, 263 genes were found to be differentially expressed in disease, several of them connected to T-cell activation, B-cell maturation, or epithelial cell differentiation, processes that had previously been described as involved in the pathogenesis of the disease. They also looked for changes in the CD predisposition linkage regions known at that time, 5q31–33, 2q31–33, and 15q11–13, but failed to detect any alterations. Constitutively altered genes, whose expression is altered in untreated patients but does not revert to normal after GFD were classified as especially interesting and potentially important in the etiology of CD, but no further analyses were performed on those genes. More recently, studies focused on specific groups of genes have observed that constitutively altered genes usually have more central roles and belong to the core of biological pathways, while those that are altered as a consequence of active disease are located more peripherally (Fernandez-Jimenez et al., 2014).

Two months later, a second microarray study (Diosdado et al., 2004) interrogated 19,200 genes to compare gene expression profiles of duodenal biopsies from 15 CD patients with villous atrophy (Marsh III) and 7 control individuals with normal mucosa (Marsh 0). However, the patients composing the celiac group were heterogeneous because some CD patients were on a gluten-containing diet, while others were on GFD and these subgroups were used for an additional comparison. The study produced two lists of differentially expressed genes: one that reflected intestinal damage regardless of gluten ingestion and another one describing the effect of gluten on the patients. The genes corresponding to intestinal damage in the absence of gluten pointed toward an increased number of T cells and macrophages at the lesion site, a more intense Th1 response, and enhanced cell proliferation. In the case of the effect of gluten, several of the genes were also related to cell cycle and cell division. Based on these results, the authors hypothesized that villous atrophy in CD patients is due to an imbalance between cell proliferation and cell differentiation of the cells that make up the crypt–villi units. On top of this, one of the most differentially expressed genes found was the gene that codes for prolyl endopeptidase (*PREP*), an enzyme that is able to hydrolyze gliadin into small fragments and that had already been proposed as a possible therapeutic agent for CD. However, subsequent studies did not find any sign of genetic association of *PREP* with disease (Diosdado et al., 2005) and have failed to confirm its efficacy as a therapeutic agent (Matysiak-Budnik et al., 2005).

A third microarray study in intestinal biopsies was performed using the standard Affymetrix Human Genome U133 Plus 2.0 Array (Castellanos-Rubio et al., 2010a). The study was designed to analyze the effects of the acute and long-term effects of gliadin exposure on CD-prone duodenal mucosa. To uncover the immediate effects of gliadin, biopsies from nine CD patients on GFD were incubated with gliadin for a 4-h time period and compared with nonstimulated biopsies from the same individuals. For the long-term analysis biopsies from nine active CD patients were compared to those taken from nine patients after more than 2 years on GFD. Using enrichment analyses, differentially expressed genes were analyzed for their involvement in pathogenic biological processes. As in previous studies, cell cycle and apoptosis were found to be altered in both acute and long-term response to gluten. Overall, 6% of the genes affected had cell cycle as one of the biological functions with a tendency toward upregulation of the genes in both analyses. Altered expression of genes from the extracellular matrix and cellular communication pathways was evident also in both acute and chronic effects analyses, suggesting that the relationship of the cells with their environment is affected since very early after the gliadin insult. More recent studies have corroborated the involvement of intestinal barrier-related genes in CD pathogenesis (Kumar et al., 2015). Interestingly, the dysregulation of more complex signaling pathways related to TGF-β, Jak–Stat, and NFκB was preferentially observed in the chronic response experiment, pointing toward their involvement in the perpetuation of the disease process rather than in its onset. Since the main environmental trigger of CD is known, an in vitro recapitulation of the direct effect of the toxic molecule can be performed, making it possible to focus on alterations that are directly related to the pathogenic process. This type of approach helps to identify genes that are altered in different stages of the disease and to draw a timeline of the biological alterations that lead to disease.

The main pathways associated with CD by the first two microarray studies were mainly those related with immune cells (T and B lymphocytes and macrophages) and in the same line, the third microarray also presented increased activity of the antigen processing and presentation pathway. This is not surprising, since CD is an immune-mediated disease in which it is well established that CD4 + T cells recognize deamidated gluten peptides bound to predisposing HLA-DQ molecules on antigen-presenting cells. On the other hand, B cells produce both antibodies against gliadin and autoantibodies against the enzyme TG2, but the way those B cells are activated is still under investigation (du Pre and Sollid, 2015). Moreover, gliadin-derived peptides

have been reported as stimulators of macrophages inducing the secretion of cytokines such as TNF-α, CCL5, and IL12 (Thomas et al., 2006; Tuckova et al., 2002), genes that have been seen to be upregulated in the transcriptome analyses.

The same is true for the alteration of genes related to cell cycle, differentiation, and proliferation pathways. Maintenance of the small intestine functions relies on a proper balance between dying and dividing cells, and the disruption of this control is a hallmark of the damage observed in the celiac gut mucosa. Incubation of biopsies with gliadin provokes an acute alteration in these genes, pointing to dysregulation of these processes as an early event in the development of the disease.

Microarray analyses have also been carried out in human intestinal cell lines and in enterocyte populations separated from duodenal biopsies. In order to find those genes that have similar responses in an epithelial cell differentiation model and in CD biopsies, a combined analysis of 5188 genes was performed (Juuti-Uusitalo et al., 2007) in a three-dimensional TGF-β1-induced T84 epithelial cell differentiation model and in small-bowel mucosal biopsy samples from untreated and treated CD patients as well as healthy controls. The aim of the study was to identify genes induced by gluten in CD samples that were also expressed in the cell model upon differentiation. The experiment identified 30 genes whose mRNA expression was similarly altered in the cell differentiation model and in patient biopsies. Several genes were members of the epidermal growth factor (EGF)-mediated signaling pathway and it was speculated that this could be due to an EGF-like response of epithelial cells to peptic–tryptic digests of gluten and small gluten peptides, as had been recently proposed (Barone et al., 2007). As a result of this work, new candidate genes responsible for disease development were proposed and the epithelial cell differentiation model was accepted as a valid tool for the study of gene expression changes in the crypt–villus axis of the small-intestinal mucosa. Subsequent studies on the involvement of the EGF pathway in CD showed that the increased crypt enterocyte proliferation observed in intestinal biopsy samples from patients with CD could be reduced by EGF receptor and IL15 blocking antibodies when used in combination (Nanayakkara et al., 2013a) but no further studies have been done to explore the therapeutic possibilities of these findings.

Another microarray study was performed in the intestinal cell line Caco2 (Parmar et al., 2013). Undifferentiated Caco2 cells were exposed to pepsin and trypsin alone, pepsin- and trypsin-digested (PT) gliadin and to a

synthetic peptide corresponding to gliadin aminoacids p31–43 for 6 h, in order to screen for downstream gliadin target genes. Although an extensive list of genes was generated, only those genes from PT-gliadin incubated cells were successfully validated by RT-QPCR. However, the same genes were also affected in the pepsin–trypsin only negative controls, so there was no way to conclude that the effects were gliadin specific. The authors hypothesized that gliadin effects only occur in the presence of immune cells. Nevertheless, different studies have detected effects in monolayer formation and in the production of cytokines of Caco2 cells after PT-gliadin incubation (Gujral et al., 2015) without cocultured immune cells. There is also evidence showing that gluten can induce inflammatory responses in the intestinal epithelial layer and provoke the disruption of tight junction proteins thus increasing intestinal permeability. In vitro models modeling the CD epithelial layer, mainly using Caco2 cells, have allowed for significant advances in understanding how epithelial cells respond to different gliadin peptides (Stoven et al., 2013) and using microarray analyses these responses have been quantified at a whole genome gene-expression level. Although cell models are a good way to assess the cellular response to gliadin and microarray studies have been helpful to confirm certain pathological pathways, results from in vitro models have to be handled with caution, using the proper controls and without forgetting that we are lacking the disease-specific genetic background and the presence of the different cell populations composing the disease-affected tissue.

To further test whether dysregulation of the epithelial cell differentiation in the small intestine is important for the development of the celiac lesion, another independent study examined whole genome expression in epithelial cells isolated from intestinal biopsies of five CD patients and five controls (Bracken et al., 2008). The analysis was performed in Atlas Glass Human 3.8 I microarrays and produced a list of 102 genes that were grouped according to functional categories. Among others, they found gene expression changes in the proliferation, cell death, and differentiation pathways, as had all previous experiments, although they also identified several genes previously unrelated to CD. The study of a pure population of epithelial cells is able to eliminate the contribution of genes that are expressed in other cells of the gut mucosa and thus detect cell-type specific, differentially expressed genes. The separation of the different cell populations that make up the affected tissue will help identify discordant alterations among cell types and will expose new candidates that are unnoticeable when analyzing whole intestinal biopsies.

3. RNA-SEQ STUDIES

In the last years, RNA sequencing has become a powerful alternative to microarrays with lower background signals and higher sensitivity. Besides, RNA-seq is not limited to the detection of previously described transcripts and allows the identification of novel transcribed regions that could be important in disease pathogenesis (Wang et al., 2009).

So far, only one RNA-seq study in CD has been published (Quinn et al., 2015) in which the transcriptome of CD4+ T lymphocytes derived from peripheral blood mononuclear cells of patients and controls was studied in basal conditions and in response to anti-CD3/CD28 antibody and phorbol 12-myristate 13-acetate (PMA) stimulations, detected 87, 1069, and 2007 disease-related differentially expressed genes in each group, respectively. Subsequent pathway enrichment analyses were not able to find any altered pathway in the unstimulated cell group while the most enriched pathway in the antibody-stimulated group was the "Cytokine–cytokine receptor interaction" pathway. Interestingly, *CCR1*, *IL18RAP*, *IL21*, and *IL2RA*, (Dubois et al., 2010; Trynka et al., 2011) four genes that are genetically associated with the disease are members of this pathway and showed differential expression in the study. The PMA-stimulated samples also showed enrichment of immune-related genes, with a high degree of overlap in their differential expressed genes and thus, pathways.

Altered genes from the anti-CD3/CD28 stimulation were grouped into modules based on pairwise correlation between their expression levels. The authors considered the two correlation modules most enriched for differentially expressed genes to be potentially related to CD pathogenesis. Subsequent analyses showed an overrepresentation of immune-related pathways and the TGF-β receptor pathway, and both modules contained genes that had been associated with genetic risk. In particular, *BACH2* (Dubois et al., 2010) was downregulated under all conditions tested, so that a network of *BACH2*-regulated genes that are altered in CD was identified. These findings support previous evidence pointing toward an important role of *BACH2* in the regulation of T cell differentiation and autoimmune disease prevention, due to its function as a transcription factor which binds super enhancers that control cell identity (Roychoudhuri et al., 2013; Vahedi et al., 2015).

The study was also able to find CD-related differential exon usage in 647 exons from 542 genes in the anti-CD3/CD28-stimulated samples, out of

which only 4% were differentially expressed. *YDJC*, *SH2B3*, *ITGA4*, *UBE2E3*, four genes associated with disease susceptibility, were also among the differential exon usage genes. Thus, CD gene expression profiling with RNA-seq provided an extra piece of information on top of the differential expression of genes, opening a new field of study based on the ability of genes to use different exons preferentially in the CD scenario. Finally, although T cells are pivotal in the pathogenesis of the disease, the differences between circulating and tissue resident CD4+ T cells must be taken into account, and results must be interpreted with caution.

4. miRNA EXPRESSION

miRNAs are a class of short noncoding RNA molecules that play an essential role in regulating mRNA levels at the posttranscriptional stage (Bartel, 2004). The connections between miRNAs and human disease have shown that certain miRNA expression patterns are disease specific and could be useful as prognostic markers. For example, a comprehensive miRNA profiling study demonstrated that distinct miRNA expression patterns were specific to various types of cancers and could reflect the developmental lineage and differentiation state of tumors (Lu et al., 2005). Further research has demonstrated that miRNAs play key roles in the regulation of the immune response and in autoimmune disorders (Chen et al., 2016). These studies have also provided insights about miRNA-mediated inflammatory responses and highlighted miRNAs as potential drug targets for the regulation of the immune system. To date, one miRNA study has been performed in the context of CD using miRNA arrays (Magni et al., 2014) to compare the miRNA profiles of intestinal biopsies of adult patients and controls. Four miRNAs (miR-192-5p, miR-31-5p, miR-338-3p, and miR-197) were found to be significantly decreased in the duodenum of patients particularly in those with more severe histological lesions. In silico analysis of the targets of these miRNAs pointed to genes involved in the adaptive and innate immune responses, *CXCL2*, *NOD2*, and *FOXP3* among others. Some of the targets were upregulated in the disease tissues and were inversely correlated with their putative regulatory miRNAs.

Moreover, significant expression changes of miR-192-5p and its targets *CXCL2* and *NOD2* were triggered in GFD-treated CD patients upon incubation with PT-gliadin, supporting the important role of the innate immune response in the development of this disease. Further analysis of the miRNA array showed that several other genes that are altered in CD and could be

involved in the immune response and intestinal permeability appeared to be regulated through a disease-related reduction in the levels of certain miRNAs. Interestingly, a subsequent study in which the expression of several miRNAs was studied in pediatric CD did not confirm the relationship between miR-192-5p and its putative targets *CXCL2* and *NOD2* but was able to replicate the downregulation of miR-31-5p and an upregulation of its target, *FOXP3* (Buoli Comani et al., 2015). These studies suggest that the pattern of expression of some miRNAs and their predicted gene targets could be somewhat different between adult and pediatric celiac patients and emphasizes the importance of the homogeneity of the samples when analyzing expression patterns.

5. DISEASE ASSOCIATION AND FUNCTION: EXPRESSION QUANTITATIVE TRAIT LOCI

As described earlier, there is an important number of genetic variants that have been associated to CD through GWAS and the follow-up Immunochip project. However, the identification of the causal variant within each association peak is a complicated task due to the strong LD between close markers within a *locus* and because most associations cannot be attributed to a single gene or are located in noncoding regions.

Expression quantitative trait *locus* (eQTL) mapping is the approach used to determine the phenotypic effects of associated genetic variants. Several studies have attempted to find out whether the different alleles of those SNPs influence expression levels of nearby genes, what is known as *cis*-eQTL. It has been described that more than half of the associated SNPs correlates with the expression of at least one adjacent gene (Dubois et al., 2010). However, in most of the cases, mRNA levels were measured in peripheral blood leukocyte populations and the results should be viewed with caution as gene expression and regulation vary greatly among different tissues and cells. Another additional level of complexity of this type of analysis is the fact that in many cases, the effect of a SNP on the expression of a gene only occurs in the presence of specific stimulus, such as the disease itself, intestinal inflammation, or gluten peptides. For example, an analysis of the expression of two GWAS candidate genes (*PTPRK* and *THEMIS*) demonstrated their coregulation in biopsies from active CD patients and from individuals on GFD after in vitro stimulation with gliadin, but not in nonceliac controls, suggesting that *cis*-eQTLs could be stimulus dependent (Bondar et al., 2014). Furthermore, a more recent study of 44 SNPs and 45 candidate genes in intestinal biopsies found

only four *cis*-eQTLs and it was observed that most often, SNP genotypes correlated with the expression levels of one or more genes in other chromosomal regions or *trans*-eQTLs (Plaza-Izurieta et al., 2015).

The effect of the SNPs associated with disease goes far beyond the simplistic idea of a transcriptional control over an adjacent *locus*. To come to understand the complex mechanisms bridging genotype and phenotype, studies must take into account the different cell types present in the celiac lesion, together with the influence of different stimuli. Future eQTL-mapping studies will focus on broadening the range of available tissues and cell types, in order to determine the key tissues and cell types involved in complex traits. Large meta-analyses will be able to pinpoint the causal variants within the trait-associated loci and determine their downstream effects in greater detail (Westra et al., 2013). Also, one must bear in mind that genes act in a coordinated manner within complex biological pathways, whose homeostasis is key to health, and may be disturbed by alterations on different genes in different patients.

5.1 Studies in Blood

To date, the majority of eQTL studies in CD have been performed using blood samples. A study in primary cell RNA from peripheral blood mononuclear cells (PBMCs) from CD patients found 1273 SNPs that influenced the expression of 372 genes. When these results were compared to a published HapMap immortalized B cell line dataset, 135 different SNPs regulating 51 different genes were common to both samples, indicative of major conserved functions. However, most *cis*-eQTLs were only detected in one of the two cell populations, confirming that the functional consequences of genetic variation is cell-type specific, which is important for studies aiming to functionally characterize risk variants for common diseases., as studying immune tissues for immune-mediated diseases. In this particular study, *cis*-eQTLs related to "immunity and defense" were more abundant in CD patient derived PBMCs than in the HapMap B cell line. Interestingly, when focusing on SNPs that had been previously associated with CD risk, variants from the *IL18RAP* and *CCR3* were shown to be disease-specific eQTLs, because there was no significant genotype-expression correlation in the HapMap B cells (Heap et al., 2009). However, it could well be that these SNPs are not functional eQTLs in B cells. Another study (Hunt et al., 2008) correlated gene expression in whole blood RNA samples with the first GWAS SNPs from non-HLA celiac-associated regions (van Heel et al., 2007). They found that the minor allele for SNP rs917997, which is associated to

disease risk correlates with lower mRNA expression of *IL18RAP*, a finding that is somewhat controversial because this gene is actually overexpressed in intestinal biopsies of CD patients (Castellanos-Rubio et al., 2016; Plaza-Izurieta et al., 2015). Although *IL18RAP* has been described as a blood cell eQTL in various studies in CD and in T1D (Dubois et al., 2010; Heap et al., 2009; Myhr et al., 2013), the associated SNP does not seem to control the expression of *IL18RAP* gene in the intestinal tissue.

In a study that assessed 38 genome-wide significant and suggestive CD-associated non-HLA *loci* for *cis* expression-genotype correlations identified eQTLs in 20 of them (Dubois et al., 2010). Among the genes regulated by disease-associated SNPs, there was enrichment for the viral RNA detection pathway genes, and *TLR8* gene expression in whole blood seems to be regulated by an associated SNP. This finding implicated viral infection as a putative environmental trigger in CD, an attractive hypothesis that has been widely studied but no conclusive results have been obtained so far.

The last study performed in blood analyzed the relationship of SNPs associated to different autoimmune diseases with gene expression (Ricano-Ponce et al., 2016). In the case of CD, the authors detected 10 *cis*-eQTLs corresponding to protein-coding genes. The study revealed two novel candidate genes in 1q32.1 and 15q24.1, corresponding to genes *GPR25* and *ULK3*, respectively. *ULK3* encodes a kinase involved in autophagy (Young et al., 2009), a pathway that had not been previously implicated in CD. Autophagy pathway genes were observed to be enriched among the differentially expressed genes in intestinal biopsies of patients. Interestingly, lipopolysaccharide-stimulated PBMC production of IL6 was correlated with the *ULK3* SNP genotypes, with the risk allele producing less IL6. The study proposes that *ULK3*-dependent autophagy could be involved in regulating inflammation and highlights the importance of studying nongluten antigens, as the members of microbiome, in the context of disease pathogenesis.

In summary, although gene expression studies in blood have proposed a handful of genes that seem to be regulated by CD-associated polymorphisms, functional and confirmation studies in the pertinent tissues and cells will be necessary to clearly establish the allele—gene expression relationships.

5.2 Studies in Intestinal Biopsies

The correlation between genotype and expression of GWAS candidate genes in CD gut mucosa was first tested in a Spanish cohort. In this study, the 11 top SNPs from 8 GWAS regions were genotyped and the expression

of the candidate genes *RGS1*, *IL18R1/IL18RAP*, *CCR3*, *IL12A/SCHIP1*, *LPP*, *IL2/IL21-KIAA1109*, *TAGAP*, and *SH2B3* was quantified in the intestinal mucosa of active and GFD-treated CD patients and nonceliac controls. Disease association of several SNPs was replicated in this cohort and differential expression of some of the candidate genes was also observed in several, but the correlation between SNP genotype and mRNA levels was only detected in *IL12A*, *LPP*, *SCHIP1*, and *SH2B3* (Plaza-Izurieta et al., 2011). *LPP* is highly expressed in the small intestine, has been implicated in cell adhesion, and could have a structural role in the maintenance of epithelial integrity (Trynka et al., 2010). However, the two studies that have found altered expression in biopsies from CD patients are contradictory in the direction of the change (Almeida et al., 2014; Plaza-Izurieta et al., 2011). Intriguingly, in the study that found overexpression of *LPP* in CD biopsies, the risk allele had an opposite effect to what was expected, since biopsies from active patients homozygous for the risk allele had the lowest *LPP* mRNA levels (Plaza-Izurieta et al., 2011). In turn, the study that found lower amounts of *LPP* in CD described six SNPs in the *LPP* locus with potential regulatory functions but did not analyze the correlation between the genotype of those SNPs and the expression of the gene. Another study that did not detect disease-related differences in expression levels found that the subcellular distribution of LPP protein is altered in CD, with increased LPP in focal adhesions and reduced LPP in the nucleus (Nanayakkara et al., 2013b). Thus, although the implication of this gene in the disease seems clear, further studies are needed to confirm an eQTL in this *locus* and to understand the specific role of LPP in the development of the disease. *SCHIP1* has been shown to be underexpressed in both active CD and GFD-treated patient tissue samples. As with *LPP*, the effect of the risk allele on *SCHIP1* gene expression goes in the opposite direction than predicted, since the risk allele is associated with a higher expression of *SCHIP1*. This gene has been implicated in in other immune-related diseases (Jin et al., 2014) but it is close to *IL12A* within the same GWAS-associated region (3q25-3q26). *IL12A* expression is known to be increased in CD (Castellanos-Rubio et al., 2009) and the Plaza-Izurieta study showed that a single copy of rs9811792*C SNP allele is capable of increasing the production of *IL12A* mRNA. The two top SNPs in these genes (rs17810546, rs9811792) are located in a 70-kb LD block that shows strong association with CD. It has been suggested that IL12A and SCHIP represent independent association signals (Hunt et al., 2008) but the functionality of the variants is still not clear. *SH2B3* is constitutively upregulated in patient mucosa, independent of disease status, suggesting a defect that precedes disease development and could be

influenced by a genetic variant (Hunt et al., 2008; Plaza-Izurieta et al., 2015). Moreover, investigation of the effect of the *SH2B3* genotype in response to lipopolysaccharide and muramyl dipeptide revealed that carriers of the *SH2B3* risk allele showed stronger activation of the NOD2 recognition pathway, suggesting that *SH2B3* could be involved in protection against bacterial infections (Zhernakova et al., 2010). The presence of the risk allele also correlates with a higher expression in the mucosa of active patients, making *SH2B3* a good candidate for functional follow-up studies.

Apart from the GWAS studies, pathway analyses can also identify disease-associated regulatory variants. Previous microarray experiments have shown an overall upregulation of the ubiquitin–proteasome system pathway in intestinal mucosa from active CD patients (Castellanos-Rubio et al., 2010a). This pathway plays a central role in the selective degradation of intracellular proteins, including those involved in antigen presentation, NFκB-mediated inflammatory processes, cell cycle regulation, and cytokine gene expression (Elliott et al., 2003). The most pronounced overexpression was observed for the ubiquitin D (*UBD*) gene with a fold change over 14 in active CD compared to GFD-treated patients. Due to its participation in the activation of the NFκB pathway and apoptosis, two important pathways in CD pathogenesis, this gene was further investigated in the context of the disease. Overexpression of *UBD* was confirmed in an independent set of samples. After evaluating the presence of putative regulatory SNPs in the region the best candidate was genotyped and found to be associated with the disease. Additionally, there was a significant correlation between the SNP genotype and the amount of *UBD* mRNA, with the highest levels present among those individuals homozygous for the risk allele (Castellanos-Rubio et al., 2010b). Subsequent studies have observed highly upregulated expression of UBD induced by proinflammatory cytokines IFN-γ and TNF-α and by the synergistic interaction NFκB and STAT3 in several cell types and tissues. *UBD* is also increased in many cancer types including colorectal and gastric cancer and regulates the transcription of the tumor suppressor p53 (Aichem and Groettrup, 2016; Choi et al., 2014). Although, the relationship between *UBD* and inflammation is now clear, no further studies have been performed around the possible implication of this SNP in the levels of *UBD* mRNA.

5.3 CD-Associated Variants and the Thymus

Almost all regions associated to CD by GWAS contain genes with a relevant immune function, and many of them participate in the same biological

pathways. One of the pathways in which CD-associated genes are implicated is "thymic T-cell development" (Dubois et al., 2010) so the thymus can be considered a relevant organ to address the regulatory role of CD genetic risk variants. A study (Amundsen et al., 2014) that analyzed thymic tissues from 42 subjects to determine the regulatory potential of 50 CD-associated SNPs detected eQTLs that involved 27 expression-altering SNPs (eSNPs) and 39 unique genes (eGenes). From the 27 eQTLs found, 6 were considered to be thymus-specific eSNPs when compared to a previous study in PBMCs and another 5 tissues from GeneVar (Dubois et al., 2010; Yang et al., 2010). One of the genes that showed thymus-specific regulation, *ZFP36L1*, is an RNA-binding protein that regulates mRNA turnover and is involved in thymocyte development in mice, highlighting once more the importance of investigating eQTLs in all the disease-relevant tissues.

6. CD VARIANTS IN NONCODING REGIONS

Since more than 80% of the GWAS-associated SNPs are located in gene regulatory or in intergenic regions, the search for functional elements that could explain the genetic risk is another active research field.

6.1 Regulatory Elements

To study the impact of common trait GWAS–enhancer variants on gene expression H3K4me1 sites identified by ChIP-seq were analyzed in B lymphoblasts. It was observed that risk SNPs associated with CD often mapped within H3K4me1 sites in these cells and the authors proposed that multiple enhancer variants cooperatively contribute to the alteration of the expression of their gene targets (Corradin et al., 2014). In another study where associated SNPs were crossed with regulatory elements categorized by the ENCODE project, the authors found that CD SNPs are significantly enriched in B-cell-specific enhancer regions (Kumar et al., 2015). The results of the analysis of regulatory elements in these two studies, suggests that B-cell responses play a major role in CD, even if it has been mainly considered as a T-cell driven disease.

6.2 Long Noncoding-RNAs

lncRNAs are RNA molecules longer than 200 bp in length with no protein-coding potential. They have diverse and still not very well-characterized

mechanisms of action by which they regulate gene expression at the transcriptional and posttranscriptional level. When the correlation of SNP genotypes with eQTLs of large intergenic noncoding RNAs (lincRNAs) has been analyzed, more than 100 *cis*-regulated lincRNAs were found. Moreover, the majority of these SNPs seem to be specifically affecting the expression of lincRNAs without influencing the expression of neighboring protein-coding genes. As is the case of coding genes, lincRNA eQTLs are also tissue-specific and the regulatory SNPs had been associated with human complex diseases (Kumar et al., 2013). Analysis of the expression profiles of lncRNAs located in autoimmune disease-associated regions showed that lncRNA transcripts are enriched in autoimmune-disease loci, suggesting that lncRNAs may be crucial to interpret GWAS findings. In the top associated CD GWAS locus on 4q27 there are four protein-coding genes (*KIAA1109*, *ADAD1*, *IL2*, *IL21*) and one lncRNA (*IL21-AS*). Coexpression analyses showed that *IL21* and *IL21-AS* are corregulated in Th1 lymphocytes, suggesting that the lncRNA could be involved in the proper function of the coding gene (Hrdlickova et al., 2014). A study that analyzed the relationship between SNPs associated with different autoimmune diseases and gene expression in blood samples (Ricano-Ponce et al., 2016) found 27 ncRNAs as potential causal genes in 25 autoimmunity loci. The association of the 2q31.3 and 11q23.3 regions to CD was attributed to two lncRNAs (*AC104820.2* and *AP002954.4*) with restricted expression in immune cells. Recently, lncRNA *AC104820.2* has been seen to be upregulated in intestinal biopsies of CD patients (Plaza-Izurieta et al., 2015).

A functional example of lncRNA-mediated regulation of disease-related genes has been found in NFκB pathway, which is known to be constitutively activated in the small intestine of CD (Fernandez-Jimenez et al., 2014; Maiuri et al., 2003). Polymorphisms in genes of the pathway have also been associated with susceptibility to the disease but no correlation of SNP genotypes and gene expression has been demonstrated (Sun and Zhang, 2007; Trynka et al., 2009). Recently, a lncRNA harboring a disease-associated SNP close to *IL18RAP* gene has been described as a key regulator of genes in the NFκB pathway. This lncRNA functions as a scaffold for a protein complex that binds chromatin at the transcription start site, maintaining the expression of certain CD altered inflammatory genes at basal levels. The CD risk allele binds the protein complex less efficiently causing an increase in the expression of inflammatory genes, which in turn will predispose to disease development (Castellanos-Rubio et al., 2016; Fig. 2).

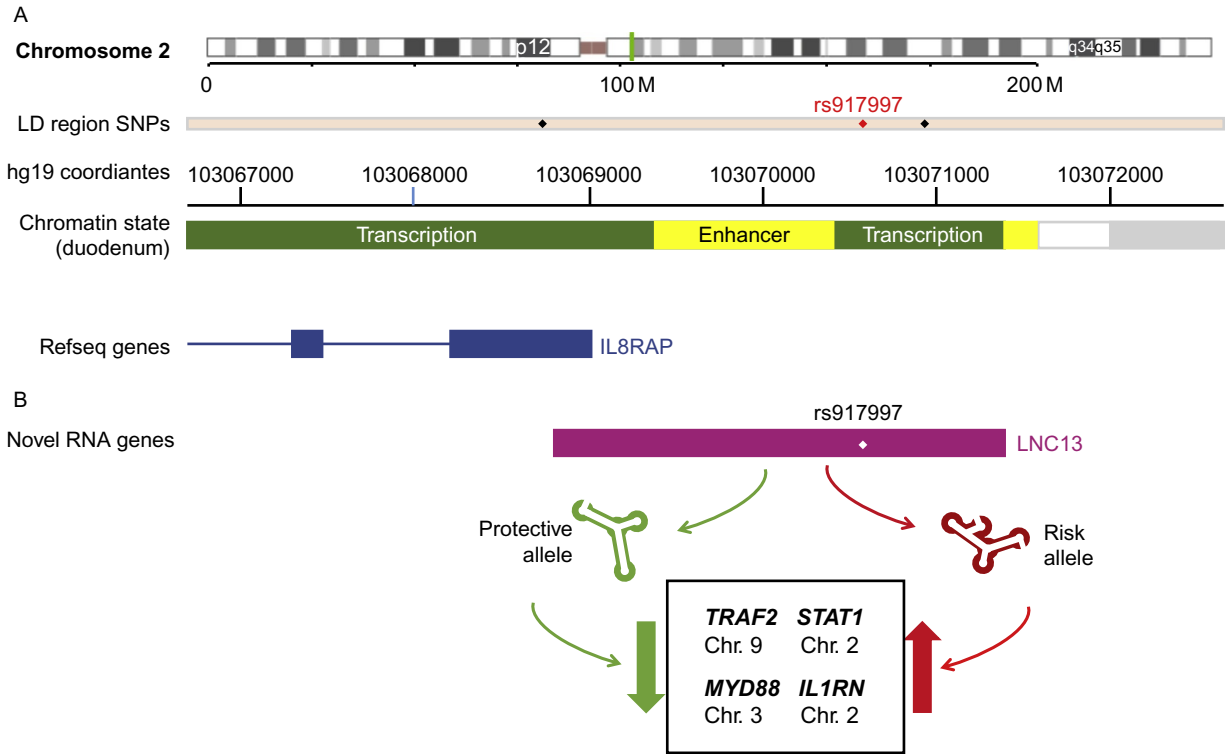

Fig. 2 (A) The recently discovered long noncoding RNA *lnc13* controls the expression of several proinflammatory genes elsewhere in the genome. SNP re917997 modifies its secondary conformation and affects the recruitment of other components of the regulatory complex, so that this SNP is actually a *trans*-eQTL for those genes.

In summary, the availability of the target tissue for research has made it possible to explore the changes in the transcriptome that occur in the celiac-prone gut in different stages of the disease process. This general approach, together with more specific cell-based or stimulus-dependent studies has provided a catalog of altered genes and pathways so that we are now beginning to understand the complexity of the interactions between the environmental trigger, genetic polymorphisms, and gene expression as well as the timeline of the pathogenic processes that occur. The identification of the key genomic players in those events will be crucial for the identification of potential targets for prevention and therapy of this and other disorders of autoimmune etiology that share common genetic roots.

REFERENCES

Aichem, A., Groettrup, M., 2016. The ubiquitin-like modifier FAT10 in cancer development. Int. J. Biochem. Cell Biol. 79, 451–461.

Almeida, R., Ricano-Ponce, I., Kumar, V., Deelen, P., Szperl, A., Trynka, G., Gutierrez-Achury, J., Kanterakis, A., Westra, H.J., Franke, L., Swertz, M.A., Platteel, M., Bilbao, J.R., Barisani, D., Greco, L., Mearin, L., Wolters, V.M., Mulder, C., Mazzilli, M.C., Sood, A., Cukrowska, B., Nunez, C., Pratesi, R., Withoff, S., Wijmenga, C., 2014. Fine mapping of the celiac disease-associated LPP locus reveals a potential functional variant. Hum. Mol. Genet. 23, 2481–2489.

Amundsen, S.S., Viken, M.K., Sollid, L.M., Lie, B.A., 2014. Coeliac disease-associated polymorphisms influence thymic gene expression. Genes Immun. 15, 355–360.

Barone, M.V., Gimigliano, A., Castoria, G., Paolella, G., Maurano, F., Paparo, F., Maglio, M., Mineo, A., Miele, E., Nanayakkara, M., Troncone, R., Auricchio, S., 2007. Growth factor-like activity of gliadin, an alimentary protein: implications for coeliac disease. Gut 56, 480–488.

Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.

Bondar, C., Plaza-Izurieta, L., Fernandez-Jimenez, N., Irastorza, I., Withoff, S., Wijmenga, C., Chirdo, F., Bilbao, J.R., CEGEC, 2014. THEMIS and PTPRK in celiac intestinal mucosa: coexpression in disease and after in vitro gliadin challenge. Eur. J. Hum. Genet. 22, 358–362.

Bracken, S., Byrne, G., Kelly, J., Jackson, J., Feighery, C., 2008. Altered gene expression in highly purified enterocytes from patients with active coeliac disease. BMC Genomics 9, 377.

Buoli Comani, G., Panceri, R., Dinelli, M., Biondi, A., Mancuso, C., Meneveri, R., Barisani, D., 2015. miRNA-regulated gene expression differs in celiac disease patients according to the age of presentation. Genes Nutr. 10, 482.

Castellanos-Rubio, A., Santin, I., Irastorza, I., Castano, L., Carlos Vitoria, J., Ramon Bilbao, J., 2009. TH17 (and TH1) signatures of intestinal biopsies of CD patients in response to gliadin. Autoimmunity 42, 69–73.

Castellanos-Rubio, A., Santin, I., Martin-Pagola, A., Irastorza, I., Castano, L., Vitoria, J.C., Bilbao, J.R., 2010a. Long-term and acute effects of gliadin on small intestine of patients on potentially pathogenic networks in celiac disease. Autoimmunity 43, 131–139.

Castellanos-Rubio, A., Santin, I., Irastorza, I., Sanchez-Valverde, F., Castano, L., Vitoria, J.C., Bilbao, J.R., 2010b. A regulatory single nucleotide polymorphism in the ubiquitin D gene associated with celiac disease. Hum. Immunol. 71, 96–99.

Castellanos-Rubio, A., Fernandez-Jimenez, N., Kratchmarov, R., Luo, X., Bhagat, G., Green, P.H., Schneider, R., Kiledjian, M., Bilbao, J.R., Ghosh, S., 2016. A long non-coding RNA associated with susceptibility to celiac disease. Science 352, 91–95.

Chen, J.Q., Papp, G., Szodoray, P., Zeher, M., 2016. The role of microRNAs in the pathogenesis of autoimmune diseases. Autoimmun. Rev. 15, 1171–1180.

Choi, Y., Kim, J.K., Yoo, J.Y., 2014. NFkappaB and STAT3 synergistically activate the expression of FAT10, a gene counteracting the tumor suppressor p53. Mol. Oncol. 8, 642–655.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., Scacheri, P.C., 2014. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 24, 1–13.

Diosdado, B., Wapenaar, M.C., Franke, L., Duran, K.J., Goerres, M.J., Hadithi, M., Crusius, J.B., Meijer, J.W., Duggan, D.J., Mulder, C.J., Holstege, F.C., Wijmenga, C., 2004. A microarray screen for novel candidate genes in coeliac disease pathogenesis. Gut 53, 944–951.

Diosdado, B., Stepniak, D.T., Monsuur, A.J., Franke, L., Wapenaar, M.C., Mearin, M.L., Koning, F., Wijmenga, C., 2005. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. Am. J. Physiol. Gastrointest. Liver Physiol. 289, G495–500.

du Pre, M.F., Sollid, L.M., 2015. T-cell and B-cell immunity in celiac disease. Best Pract. Res. Clin. Gastroenterol. 29, 413–423.

Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A., Bardella, M.T., van den Berg, L.H., Bockett, N.A., de la Concha, E.G., Dema, B., Fehrmann, R.S., Fernandez-Arquero, M., Fiatal, S., Grandone, E., Green, P.M., Groen, H.J., Gwilliam, R., Houwen, R.H., Hunt, S.E., Kaukinen, K., Kelleher, D., Korponay-Szabo, I., Kurppa, K., MacMathuna, P., Maki, M., Mazzilli, M.C., McCann, O.T., Mearin, M.L., Mein, C.A., Mirza, M.M., Mistry, V., Mora, B., Morley, K.I., Mulder, C.J., Murray, J.A., Nunez, C., Oosterom, E., Ophoff, R.A., Polanco, I., Peltonen, L., Platteel, M., Rybak, A., Salomaa, V., Schweizer, J.J., Sperandeo, M.P., Tack, G.J., Turner, G., Veldink, J.H., Verbeek, W.H., Weersma, R.K., Wolters, V.M., Urcelay, E., Cukrowska, B., Greco, L., Neuhausen, S.L., McManus, R., Barisani, D., Deloukas, P., Barrett, J.C., Saavalainen, P., Wijmenga, C., van Heel, D.A., 2010. Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. 42, 295–302.

Elliott, P.J., Zollner, T.M., Boehncke, W.H., 2003. Proteasome inhibition: a new anti-inflammatory strategy. J. Mol. Med. 81, 235–245.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., De Jager, P.L., Kuchroo, V.K., Epstein, C.B., Daly, M.J., Hafler, D.A., Bernstein, B.E., 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–343.

Fernandez-Jimenez, N., Castellanos-Rubio, A., Plaza-Izurieta, L., Irastorza, I., Elcoroaristizabal, X., Jauregi-Miguel, A., Lopez-Euba, T., Tutau, C., de Pancorbo, M.M., Vitoria, J.C., Bilbao, J.R., 2014. Coregulation and modulation of NFkappaB-related genes in celiac disease: Uncovered aspects of gut mucosal inflammation. Hum. Mol. Genet. 23, 1298–1310.

Gujral, N., Suh, J.W., Sunwoo, H.H., 2015. Effect of anti-gliadin IgY antibody on epithelial intestinal integrity and inflammatory response induced by gliadin. BMC Immunol. 16, 41.

Gutierrez-Achury, J., Zhernakova, A., Pulit, S.L., Trynka, G., Hunt, K.A., Romanos, J., Raychaudhuri, S., van Heel, D.A., Wijmenga, C., de Bakker, P.I., 2015. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nat. Genet. 47, 577–578.

Heap, G.A., Trynka, G., Jansen, R.C., Bruinenberg, M., Swertz, M.A., Dinesen, L.C., Hunt, K.A., Wijmenga, C., Vanheel, D.A., Franke, L., 2009. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC Med. Genomics 2, 1.

Hrdlickova, B., Kumar, V., Kanduri, K., Zhernakova, D.V., Tripathi, S., Karjalainen, J., Lund, R.J., Li, Y., Ullah, U., Modderman, R., Abdulahad, W., Lahdesmaki, H., Franke, L., Lahesmaa, R., Wijmenga, C., Withoff, S., 2014. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. Genome Med. 6, 88.

Hunt, K.A., Zhernakova, A., Turner, G., Heap, G.A., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L.C., Ryan, A.W., Panesar, D., Gwilliam, R., Takeuchi, F., McLaren, W.M., Holmes, G.K., Howdle, P.D., Walters, J.R., Sanders, D.S., Playford, R.J., Trynka, G., Mulder, C.J., Mearin, M.L., Verbeek, W.H., Trimble, V., Stevens, F.M., O'Morain, C., Kennedy, N.P., Kelleher, D., Pennington, D.J., Strachan, D.P., McArdle, W.L., Mein, C.A., Wapenaar, M.C., Deloukas, P., McGinnis, R., McManus, R., Wijmenga, C., van Heel, D.A., 2008. Newly identified genetic risk variants for celiac disease related to the immune response. Nat. Genet. 40, 395–402.

Hunt, K.A., Mistry, V., Bockett, N.A., Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O., Capon, F., Compston, A., Gough, S.C., Jostins, L., Kong, Y., Lee, J.C., Lek, M., MacArthur, D.G., Mansfield, J.C., Mathew, C.G., Mein, C.A., Mirza, M., Nutland, S., Onengut-Gumuscu, S., Papouli, E., Parkes, M., Rich, S.S., Sawcer, S., Satsangi, J., Simmonds, M.J., Trembath, R.C., Walker, N.M., Wozniak, E., Todd, J.A., Simpson, M.A., Plagnol, V., van Heel, D.A., 2013. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 498, 232–235.

Husby, S., Koletzko, S., Korponay-Szabo, I.R., Mearin, M.L., Phillips, A., Shamir, R., Troncone, R., Giersiepen, K., Branski, D., Catassi, C., Lelgeman, M., Maki, M., Ribes-Koninckx, C., Ventura, A., Zimmer, K.P., ESPGHAN Working Group on Coeliac Disease Diagnosis, ESPGHAN Gastroenterology Committee, European Society for Pediatric Gastroenterology, Hepatology, and Nutrition, 2012. European Society for Pediatric Gastroenterology, hepatology, and nutrition guidelines for the diagnosis of coeliac disease. J. Pediatr. Gastroenterol. Nutr. 54, 136–160.

Jin, J., Chou, C., Lima, M., Zhou, D., Zhou, X., 2014. Systemic sclerosis is a complex disease associated mainly with immune regulatory and inflammatory genes. Open Rheumatol. J. 8, 29–42.

Juuti-Uusitalo, K., Maki, M., Kaukinen, K., Collin, P., Visakorpi, T., Vihinen, M., Kainulainen, H., 2004. cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. J. Autoimmun. 22, 249–265.

Juuti-Uusitalo, K., Maki, M., Kainulainen, H., Isola, J., Kaukinen, K., 2007. Gluten affects epithelial differentiation-associated genes in small intestinal mucosa of coeliac patients. Clin. Exp. Immunol. 150, 294–305.

Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D.V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vosa, U., Hofker, M.H., Fehrmann, R.S., Fu, J., Withoff, S., Metspalu, A., Franke, L., Wijmenga, C., 2013. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet. 9, e1003201.

Kumar, V., Gutierrez-Achury, J., Kanduri, K., Almeida, R., Hrdlickova, B., Zhernakova, D.V., Westra, H.J., Karjalainen, J., Ricano-Ponce, I., Li, Y., Stachurska, A., Tigchelaar, E.F., Abdulahad, W.H., Lahdesmaki, H., Hofker, M.H., Zhernakova, A., Franke, L., Lahesmaa, R., Wijmenga, C., Withoff, S., 2015. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. Hum. Mol. Genet. 24, 397–409.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. Nature 435, 834–838.

Lundin, K.E., Scott, H., Hansen, T., Paulsen, G., Halstensen, T.S., Fausa, O., Thorsby, E., Sollid, L.M., 1993. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J. Exp. Med. 178, 187–196.

Magni, S., Buoli Comani, G., Elli, L., Vanessi, S., Ballarini, E., Nicolini, G., Rusconi, M., Castoldi, M., Meneveri, R., Muckenthaler, M.U., Bardella, M.T., Barisani, D., 2014. miRNAs affect the expression of innate and adaptive immunity proteins in celiac disease. Am. J. Gastroenterol. 109, 1662–1674.

Maiuri, M.C., De Stefano, D., Mele, G., Fecarotta, S., Greco, L., Troncone, R., Carnuccio, R., 2003. Nuclear factor kappa B is activated in small intestinal mucosa of celiac patients. J. Mol. Med. (Berl) 81, 373–379.

Matysiak-Budnik, T., Candalh, C., Cellier, C., Dugave, C., Namane, A., Vidal-Martinez, T., Cerf-Bensussan, N., Heyman, M., 2005. Limited efficiency of prolyl-endopeptidase in the detoxification of gliadin peptides in celiac disease. Gastroenterology 129, 786–796.

Molberg, O., Kett, K., Scott, H., Thorsby, E., Sollid, L.M., Lundin, K.E., 1997. Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. Scand. J. Immunol. 46, 103–108.

Myhr, C.B., Hulme, M.A., Wasserfall, C.H., Hong, P.J., Lakshmi, P.S., Schatz, D.A., Haller, M.J., Brusko, T.M., Atkinson, M.A., 2013. The autoimmune disease-associated SNP rs917997 of IL18RAP controls IFNgamma production by PBMC. J. Autoimmun. 44, 8–12.

Nanayakkara, M., Lania, G., Maglio, M., Kosova, R., Sarno, M., Gaito, A., Discepolo, V., Troncone, R., Auricchio, S., Auricchio, R., Barone, M.V., 2013a. Enterocyte proliferation and signaling are constitutively altered in celiac disease. PLoS One 8, e76006.

Nanayakkara, M., Kosova, R., Lania, G., Sarno, M., Gaito, A., Galatola, M., Greco, L., Cuomo, M., Troncone, R., Auricchio, S., Auricchio, R., Barone, M.V., 2013b. A celiac cellular phenotype, with altered LPP sub-cellular distribution, is inducible in controls by the toxic gliadin peptide P31-43. PLoS One 8, e79763.

Parmar, A., Greco, D., Venalainen, J., Gentile, M., Dukes, E., Saavalainen, P., 2013. Gene expression profiling of gliadin effects on intestinal epithelial cells suggests novel non-enzymatic functions of pepsin and trypsin. PLoS One 8, e66307.

Plaza-Izurieta, L., Castellanos-Rubio, A., Irastorza, I., Fernandez-Jimenez, N., Gutierrez, G., CEGEC, Bilbao, J.R., 2011. Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes. J. Med. Genet. 48, 493–496.

Plaza-Izurieta, L., Fernandez-Jimenez, N., Irastorza, I., Jauregi-Miguel, A., Romero-Garmendia, I., Vitoria, J.C., Bilbao, J.R., 2015. Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease. Eur. J. Hum. Genet. 23, 1100–1105.

Quinn, E.M., Coleman, C., Molloy, B., Dominguez Castro, P., Cormican, P., Trimble, V., Mahmud, N., McManus, R., 2015. Transcriptome analysis of CD4+ T cells in coeliac disease reveals imprint of BACH2 and IFNgamma regulation. PLoS One 10, e0140049.

Ricano-Ponce, I., Zhernakova, D.V., Deelen, P., Luo, O., Li, X., Isaacs, A., Karjalainen, J., Di Tommaso, J., Borek, Z.A., Zorro, M.M., Gutierrez-Achury, J., Uitterlinden, A.G., Hofman, A., van Meurs, J., Consortium, B., Lifelines Cohort, S., Netea, M.G., Jonkers, I.H., Withoff, S., van Duijn, C.M., Li, Y., Ruan, Y., Franke, L., Wijmenga, C., Kumar, V., 2016. Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. J. Autoimmun. 68, 62–74.

Roychoudhuri, R., Hirahara, K., Mousavi, K., Clever, D., Klebanoff, C.A., Bonelli, M., Sciume, G., Zare, H., Vahedi, G., Dema, B., Yu, Z., Liu, H., Takahashi, H., Rao, M., Muranski, P., Crompton, J.G., Punkosdy, G., Bedognetti, D., Wang, E., Hoffmann, V., Rivera, J., Marincola, F.M., Nakamura, A., Sartorelli, V., Kanno, Y., Gattinoni, L., Muto, A., Igarashi, K., O'Shea, J.J., Restifo, N.P., 2013. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. Nature 498, 506–510.

Sollid, L.M., Markussen, G., Ek, J., Gjerde, H., Vartdal, F., Thorsby, E., 1989. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J. Exp. Med. 169, 345–350.

Spurkland, A., Sollid, L.M., Polanco, I., Vartdal, F., Thorsby, E., 1992. HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. Hum. Immunol. 35, 188–192.

Stoven, S., Murray, J.A., Marietta, E.V., 2013. Latest in vitro and in vivo models of celiac disease. Expert Opin. Drug Discov. 8, 445–457.

Sun, X.F., Zhang, H., 2007. NFKB and NFKBI polymorphisms in relation to susceptibility of tumour and other diseases. Histol. Histopathol. 22, 1387–1398.

Thomas, K.E., Sapone, A., Fasano, A., Vogel, S.N., 2006. Gliadin stimulation of murine macrophage inflammatory gene expression and intestinal permeability are MyD88-dependent: role of the innate immune response in Celiac disease. J. Immunol. 176, 2512–2521.

Trynka, G., Zhernakova, A., Romanos, J., Franke, L., Hunt, K.A., Turner, G., Bruinenberg, M., Heap, G.A., Platteel, M., Ryan, A.W., de Kovel, C., Holmes, G.K., Howdle, P.D., Walters, J.R., Sanders, D.S., Mulder, C.J., Mearin, M.L., Verbeek, W.H., Trimble, V., Stevens, F.M., Kelleher, D., Barisani, D., Bardella, M.T., McManus, R., van Heel, D.A., Wijmenga, C., 2009. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. Gut 58, 1078–1083.

Trynka, G., Wijmenga, C., van Heel, D.A., 2010. A genetic perspective on coeliac disease. Trends Mol. Med. 16, 537–550.

Trynka, G., K. A. Hunt, N. A. Bockett, J. Romanos, V. Mistry, A. Szperl, S. F. Bakker, M. T. Bardella, L. Bhaw-Rosun, G. Castillejo, E. G. de la Concha, R. C. de Almeida, K. R. Dias, C. C. van Diemen, P. C. Dubois, R. H. Duerr, S. Edkins, L. Franke, K. Fransen, J. Gutierrez, G. A. Heap, B. Hrdlickova, S. Hunt, L. Plaza Izurieta, V. Izzo, L. A. Joosten, C. Langford, M. C. Mazzilli, C. A. Mein, V. Midah, M. Mitrovic, B. Mora, M. Morelli, S. Nutland, C. Nunez, S. Onengut-Gumuscu, K. Pearce, M. Platteel, I. Polanco, S. Potter, C. Ribes-Koninckx, I. Ricano-Ponce, S. S. Rich, A. Rybak, J. L. Santiago, S. Senapati, A. Sood, H. Szajewska, R. Troncone, J. Varade, C. Wallace, V. M. Wolters, A. Zhernakova, Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC), Control, B. K. Thelma, B. Cukrowska, E. Urcelay, J. R. Bilbao, M. L. Mearin, D. Barisani, J. C. Barrett, V. Plagnol, P. Deloukas, C. Wijmenga, and D. A. van Heel. 2011. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 43: 1193–1201.

Tuckova, L., Novotna, J., Novak, P., Flegelova, Z., Kveton, T., Jelinkova, L., Zidek, Z., Man, P., Tlaskalova-Hogenova, H., 2002. Activation of macrophages by gliadin fragments: isolation and characterization of active peptide. J. Leukoc. Biol. 71, 625–631.

Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M., Tang, Z., Ruan, Y., Collins, F.S., Sartorelli, V., O'Shea, J.J., 2015. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature 520, 558–562.

van Heel, D.A., Franke, L., Hunt, K.A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M.C., Barnardo, M.C., Bethel, G., Holmes, G.K., Feighery, C., Jewell, D., Kelleher, D., Kumar, P., Travis, S., Walters, J.R., Sanders, D.S., Howdle, P., Swift, J., Playford, R.J., McLaren, W.M., Mearin, M.L., Mulder, C.J., McManus, R., McGinnis, R., Cardon, L.R., Deloukas, P., Wijmenga, C., 2007. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat. Genet. 39, 827–829.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.

Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zhernakova, A., Zhernakova, D.V., Veldink, J.H., Van den Berg, L.H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., Hoen, P.A.C., Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Volker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B.M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, J.B., Franke, L., 2013. Systematic identification of *trans*-eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243.

Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P., Dermitzakis, E.T., 2010. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26, 2474–2476.

Young, A.R., Narita, M., Ferreira, M., Kirschner, K., Sadaie, M., Darot, J.F., Tavare, S., Arakawa, S., Shimizu, S., Watt, F.M., Narita, M., 2009. Autophagy mediates the mitotic senescence transition. Genes Dev. 23, 798–803.

Zhernakova, A., Elbers, C.C., Ferwerda, B., Romanos, J., Trynka, G., Dubois, P.C., de Kovel, C.G., Franke, L., Oosting, M., Barisani, D., Bardella, M.T., Finnish Celiac Disease Study, G., Joosten, L.A., Saavalainen, P., van Heel, D.A., Catassi, C., Netea, M.G., Wijmenga, C., 2010. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am. J. Hum. Genet. 86, 970–977.