eman ta zabal zazu

Universidad | **Euskal Herriko**
del País Vasco | Unibertsitatea

FUNCTIONAL STUDIES OF TIGHT JUNCTION GENES IN CELIAC DISEASE

TESIS DOCTORAL

Amaia Jauregi Miguel

Leioa, 2017

TESIS DOCTORAL

FUNCTIONAL STUDIES OF TIGHT JUNCTION GENES IN CELIAC DISEASE

Amaia Jauregi Miguel

Leioa, 2017

Directores:

Jose Ramon Bilbao Catalá

Ainara Castellanos Rubio

eman ta zabal zazu

Universidad del País Vasco   Euskal Herriko Unibertsitatea

ABBREVIATIONS

AIDS	acquired immune deficiency syndrome
AJ	adherens Junctions
APC	antigen presenting cells
CD	celiac disease
CEGEC	Spanish Consortium for Genetics of Celiac Disease
EGFR	epidermal growth factor receptor
EMA	anti-endomysium autoantibodies
EpCAM	epithelial cell adhesion molecule
ESPGHAN	European Society for Pediatric Gastroenterology, Hepatology and Nutrition
eQTL	expression quantitative trait loci
FBS	fetal bovine serum
GFD	gluten free diet
GWAS	Genome Wide Association Studies
HLA	Human Leucocyte Antigen
IELs	intraepithelial lymphocytes
JAM	junctional adhesion molecule
KIR	killer Immunoglobulin-like receptor
LD	linkage disequilibrium
lncRNA	long non-coding RNA
LY	lucifer Yellow
MHC	Mayor Histocompatibility Complex
NFkB	nuclear kappa B transcription factor
OR	odds ratio
TJ	tight junctions
TLR	toll-like receptor
PCR	polymerase chain reaction
RA	rheumatoid arthritis
RNAseq	RNA sequencing
RT-qPCR	real-time quantitative reverse transcription PCR Abbreviations
SD	standard deviation
SNP	Single Nucleotide Polymorphism
T1D	type 1 diabetes
TEER	transepithelial electrical resistance
TF	transcription factor
TG2	transglutaminase
TGA	anti-tissue transglutaminase autoantibodies
TLR	toll-like receptor

LIST OF ORIGINAL PUBLICATIONS

Jauregi-Miguel A, Fernandez-Jimenez N, Irastorza I, Plaza-Izurieta L, Vitoria JC, Bilbao JR. Alteration of tight junction gene expression in celiac disease. J Pediatr Gastroenterol Nutr. 2014 Jun;58(6):762-7.

Jauregi-Miguel A, Santin I, Romero-Garmendia I, Garcia-Etxebarria K, Sebastian M, Irastorza I, Castellanos-Rubio A, Bilbao JR. The Tight Junction adapter protein MAGI2 locus is involved in celiac disease. In preparation.

PROJECT JUSTIFICATION AND SCOPE

Celiac disease (CD) is a chronic immune mediated disorder with high prevalence. It is believed that prevention will be crucial for the eradication of this disorder, and for that purpose, efficient mechanisms of prediction and early diagnosis need to be developed. In a temporal scale, the presence of clinical symptoms can be considered an advanced stage of the disease-progression process. This active disease stage would be preceded by the presence of immunological markers, such as circulating autoantibodies against tTG (tissue transglutaminase), as well as an enhanced intestinal permeability to gluten reflecting an ongoing immune mediated tissue-destruction process that initiates only among genetically predisposed individuals.

Therefore, it becomes essential to define which genes are involved in disease susceptibility, in order to understand the pathogenic mechanisms underlying CD development and also to provide genetic markers capable of discriminating individuals at risk of this disease, which would allow predictive diagnosis to be performed prior to the activation of the autoimmune response and would improve the selection of candidates for putative immune prevention trials.

In order to dissect the genetics of this complex autoimmune disease, the current project has focused on the search of functional genetic determinants in celiac disease using the genetic approaches and functional studies performed in celiac disease and the potentially pathogenic TJ pathway related to intestinal permeability proposed form those results as starting point.

1. Introduction

1. Celiac disease

Celiac disease (CD; OMIM 212750) or gluten sensitive enteropathy is a complex chronic, immune-mediated inflammatory disorder characterized by flattened villi on the small bowel mucosa, which develops in genetically susceptible individuals in response to ingested gluten and related proteins from wheat, barley and rye.

1.1. Clinical features and diagnosis

Although the clinical picture of CD had been described by Samuel Gee more than 60 years earlier [1], the adverse effects of ingested gluten were not recognized until 1950 [2] . If untreated, classical CD presents with a range of symptoms and signs that can be divided into intestinal features, such as diarrhea, abdominal distension or vomiting, and those caused by malabsorption, like failure to thrive (low weight, lack of fat, hair thinning) or psychomotor impairment (muscle wasting) [3]. Other atypical symptoms are also associated with CD, and include neurological events, dental enamel defects, infertility, osteoporosis, joint symptoms and elevated liver enzyme concentrations [4].

From a histological point of view, when a genetically susceptible person is on a gluten-containing diet, there are gradual changes in the small intestinal mucosa that result in a lesion presenting villous atrophy and hyperplasia of the crypts (Figure 1). The degree and severity of gluten-induced mucosal alterations are classified in the Marsh-Oberhuber scale [5] (originally ranging from type I to type IV), with Marsh 0 reflecting healthy, normal intestinal villi, and Marsh IV indicating hypoplasia of small intestine architecture (Figure 1). Although this classification system has undergone some modifications since then, it still remains a cornerstone for diagnosing CD. In the first stage of the disease, there is an infiltration of the villous epithelium by intraepithelial lymphocytes (IELs) (Marsh I), which is followed by hypertrophic crypts (Marsh II), while the villi are not shortened. In the more advanced stage (Marsh III), crypts are hypertrophic, the *lamina propria* is swollen, and there is either severe partial, subtotal or total villous atrophy. In the most advanced stage (Marsh IV), hypoplasia of small intestine architecture is observed. Together with the damage of the small

intestinal mucosa, CD is characterized by the presence of different gluten-dependent serum autoantibodies, such as anti-endomysium (EMA) or anti-tissue transglutaminase (TGA) autoantibodies among others [6].
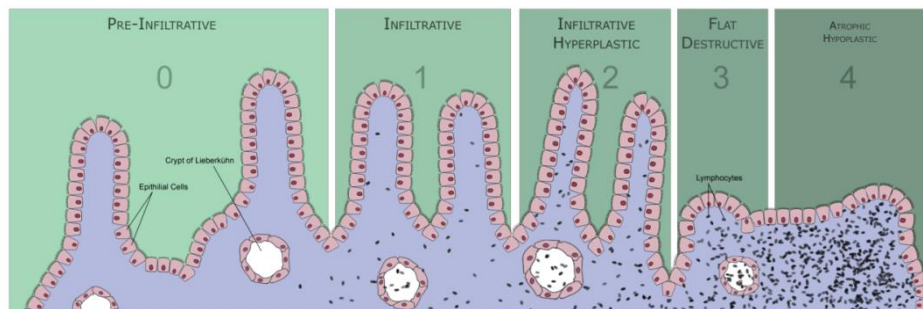
Figure 1: Gluten-induced mucosal changes in different stages in CD according to Marsh classification. Image from www.theglutensyndrome.net.

Taking into account these pathological features, the diagnostic criteria for CD are established by the European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and until recently, have been based on the presence of characteristic histological injuries in a biopsy of small intestine and by positive serologic results, although the latter were not essential [7]. In the diagnostic guidelines published in 2012, the duodenal biopsy can be excluded in symptomatic children with IgA class TGA titers above 10 times the upper limit of normal levels. Additionally, Human Leucocyte Antigen (HLA) genotyping is valuable since CD is very unlikely if risk haplotypes are absent [8].

1.2. Epidemiology

Although in the past CD was considered a comparatively uncommon disorder affecting almost exclusively people of European origin, improvement in screening studies have shown it to be a worldwide distributed common disorder that affects roughly 1-2% of the population [9, 10], being regional differences within Europe (e.g., the prevalence is 0.3% in Germany and 2.4% in Finland) [11, 12].

With the spread of modern Western diet, rich in gluten-containing cereals (especially wheat) to all parts of the world, CD has become a global Public Health problem, and also affects the populations of developing countries [13],

being common particularly in North Africa [14] and the Middle East [15]. Westernization of the diet, changes in wheat production and preparation, increased awareness of the disease, and a combination of these factors are increasing the frequency of CD [16-18]. For example, different investigations have suggested that the incidence of childhood CD may have risen during 1980s and 1990s, and this has been related to infant feeding practices [19]. Environmental risk factors with seasonal patterns, including certain viral infections have also been proposed as risk factors for CD [20].

In terms of gender, females are more commonly affected than males, and among patients presenting the disease during their fertile years, a female to male ratio of almost 3 to 1 has been observed [21].

1.3. Treatment

To date, the only proven treatment for CD is a strict and life-long removal of gluten from the diet, which is achieved by the elimination of wheat, barley, and rye cereal products [3, 22]. Improvement and resolution of symptoms typically occurs within days or weeks, and often precedes normalization of serological markers and of duodenal villous atrophy [23]. Insufficiently treated and untreated patients are predisposed to complications such as short stature, nutritional deficiencies, osteoporosis, secondary autoimmune disorders, malignancies, infertility and poor outcome of pregnancies [24].

Despite treatment effectiveness, complying with gluten free diet (GFD) is difficult and it is thought to decrease quality of life [25]. Additionally, a small subgroup of patients may show non-responsive or refractory CD, having persistent or recurrent symptoms, inflammation of the intestine and villous atrophy despite strict adherence to a GFD [26]. Therefore, the development of a safe, effective and affordable alternative therapy is necessary.

Together with the knowledge of the pathophysiological mechanisms of CD, non-dietary therapies have been developed as an addition or substitute to the GFD [27]. There are drugs in various stages of development and testing, such as larazotide acetate, an oral peptide that modulates tight junctions and prevents passage of gliadin peptides through the epithelial barrier and latiglutenase, an

enzyme preparation that prevents the pathological damage caused by gluten in patients with CD. Related to the inactivation of the immune process in the *lamina propria*, a vaccine consisting of epitopes for gluten-specific CD4-positive T cells has completed phase 1 clinical studies [28].

2. Genetics of celiac disease

Even though the precise inheritance model of CD is still unknown, it has been known for a long time that Genetics participates in the susceptibility to the disease. Evidence of a strong genetic background in disease susceptibility comes from studies on the prevalence of CD in affected families, and especially those comparing twin pairs, in which the proportion of genetic and environmental risk factors in the disease prevalence can be estimated. 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 has allowed scientists to estimate the genetic component of CD, which is higher than what has been described for other immunological complex diseases, such as type 1 diabetes (T1D) (around 30% concordance in monozygotic and 6% in dizygotic twins) [29]. Additionally, concordance rates between sibling pairs and dizygotic twins are almost the same, indicating that the environmental component has a minimum contribution to the risk of developing CD.

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% [30]. The largest portion of the genetic risk to develop CD comes from the presence of HLA-DQ2 and HLA–DQ8 heterodimers from the Major Histocompatibility Complex (MHC) class II genes. However, even if the role of these HLA molecules is essential in the pathogenesis of the disease, their contribution to the heredity is modest. It has been calculated that the classical HLA class II variants alone explain 23% of the CD heritability risk, whereas 5 novel variants in the MHC region reported an additional 18% of

genetic variance [31], which in total explain approximately 40% of CD risk. Consequently, it is clear that there must exist many small effects, conferred by non-HLA susceptibility *loci*.

During the last years, a great effort has been done in order to find additional genetic susceptibility determinants in CD, and approaches include genetic linkage studies, candidate gene association studies and in the last decade, genome-wide association (GWA) and follow-up studies. Using these strategies, several *loci* throughout the genome have been associated with CD; some of them have been firmly established to be involved in the disease while others need further investigation in order to confirm whether they are really contributing to disease susceptibility.

2.1. HLA region and celiac disease

1.1.1. HLA region

HLA genes are located on the 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. 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 [32].

1.1.2. Contribution to the genetic risk and susceptibility genes

As previously pointed out, the HLA region is the most important susceptibility *locus* in CD and explains around 40% (23% classic and 18% novel variants) of the genetic component of the disease. The first evidences supporting the association between HLA and CD where published in 1973 and were detected

using serological methods [33]. Subsequent molecular studies have revealed that the factors directly implicated are the HLA class II genes encoding both HLA-DQ2 and HLA-DQ8 molecules [34, 35] (Figure 2). HLA-DQ2 and HLA-DQ8 variants are in linkage disequilibrium with HLA-DR3 and HLA-DR4, respectively. Thus, we often refer to these risk variants as HLA DR3-DQ2 and HLA DR4-DQ8 haplotypes [34].

The primary function of DQ heterodimer molecules is to present exogenous peptide antigens (in CD, gluten) to helper T cells. Genetic polymorphisms may alter the peptide-binding groove and affect the repertoire of peptides that can be efficiently bound and presented. In CD, 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 (containing DQA1*05 and DQB1*02 alleles that encode the α and β chains of the heterodimer, respectively).

Nearly all of the CD patients negative for the HLA-DQ2 molecule are HLA-DQ8 positive, and have at least one copy of the haplotype containing DQA1*03:01 and DQB1*03:02 alleles [4]. A very small proportion of 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) [35, 36].

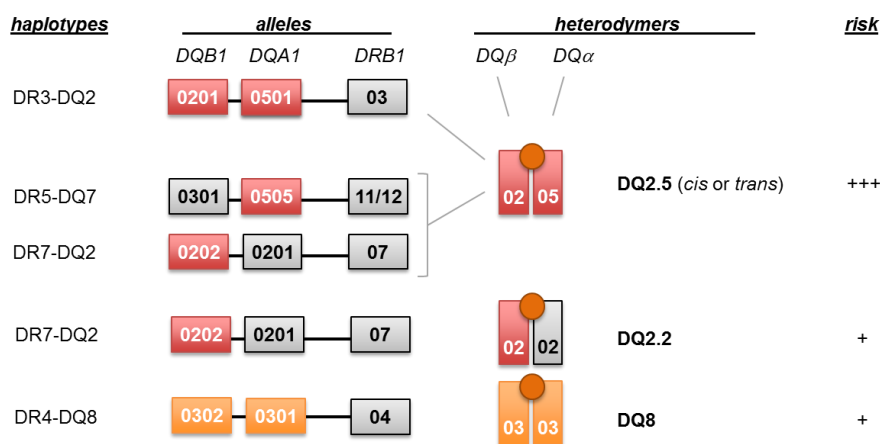**Figure 2: Association of the HLA locus with CD. HLA-DQ2 molecule is the major factor conferring risk to CD. Most celiac patients express the heterodimer HLA-DQ2.5, encoded by the alleles HLA-DQA1*05 (α chain) and HLA-DQB1*02 (β chain), that can be present in cis in the DR3-DQ2 haplotype or in trans, 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. (Adapted from Abadie et al., Annu Rev Immunol (2011)[37])

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 the 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 [38-40]. 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. 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 [31].

Although HLA genes importantly contribute to the genetic susceptibility, the HLA-DQ2 variant is also common in the general population, being present in 20-30% of non-celiac individuals, making it clear that, even though it is very important, it is not sufficient to develop the disease [41].

2.2. Non-HLA susceptibility regions: Genome-wide association studies in CD

Given the fact that HLA 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. Initially, two had been the major strategies used for this aim: 1) linkage studies in affected families and 2) association studies based on population screening. With the exception of the MHC locus, the results of the linkage scans have been

somewhat contradictory. However, apart from HLA (CELIAC1) three chromosomal regions have been repeatedly linked to CD, 5q31-33 (CELIAC2), 2q32 (CELIAC3 containing T lymphocyte regulatory genes CD28, CTLA4 and ICOS) and 19p13 (CELIAC4 containing myosin IXB gene, MYO9XB) [42].

In the past decade, high-throughput genotyping platforms have enabled genome-wide association studies (GWAS), changing the way to approach genetic studies of complex traits and diseases. GWA studies have evolved into a powerful tool that enables researchers to scan a great number of genetic markers (polymorphisms) in large genomic DNA sample sets (case-control), with the aim of finding genetic variants associated with a particular disease.

The two GWAS performed in CD, conducted in 2007 and 2010, analyzed approximately 5,000 patients and 10,000 controls and revealed a total of 26 non-HLA associated regions [43, 44]. One year later 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 13 additional regions associated with CD were identified [45].

Hence, there is a total of 39 non-HLA regions associated with CD, containing 57 independent association signals. 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 non-synonymous variant in the *NCF2* gene (rs17849502) associated with a small increase in risk [46]. Nineteen of the associated regions pinpoint to a single candidate gene, but only 3 of the associated SNPs are linked to protein-altering variants located in exonic regions and some potentially causative genes have been proposed due to the existence of signals near the 5' or 3' regulatory regions (Figure 3).
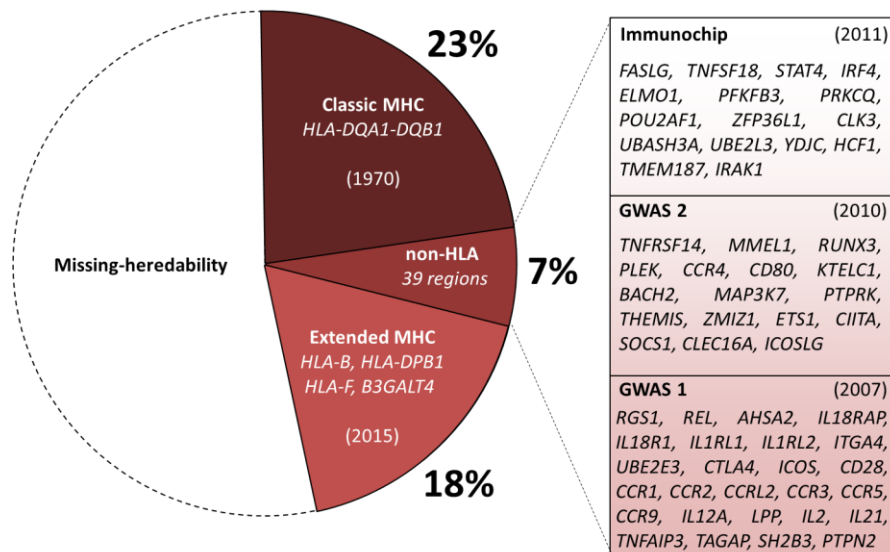
Figure 3: Advances in the Genetics of CD. After the Immunochip study, 39 non-HLA loci have been found to contribute to the genetic risk to develop the disease.

Despite the success of GWAS and follow-up studies in discovering CD susceptible *loci*, these identified variants only explain a small proportion of the genetic contributions to disease and many more remain to be found. Moreover, due to the strong linkage disequilibrium (LD), most of GWAS SNPs are not the causal variants and it is thought that they are merely pointing to associated regions that should be more deeply scanned.

Additionally, the vast majority of SNPs (more than 80%) are located outside protein coding sequences (in gene regulatory or in intergenic regions), so they are assumed to play a regulatory role in the expression of nearby genes or even genes located elsewhere in the genome. Thus, CD associated variants seem to be located in expression quantitative trait *loci* or eQTLs, genomic *loci* that regulate expression levels of mRNAs or proteins.

The recently developed RNA sequencing approach (RNAseq) has allowed the identification of transcribed non-coding regions that could partly explain the unresolved genetic risk in disease [47]. Long noncoding-RNAs (lncRNAs), are RNA molecules longer than 200bp 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 post-transcriptional levels. The analysis of the expression profiles of lncRNAs located in autoimmune disease-associated regions have shown that lncRNA transcripts

are enriched in autoimmune-disease *loci*, suggesting that lncRNAs may be crucial to interpret GWAS findings. Several GWAS SNPs, including some in CD, have been demonstrated to have eQTL-effects on these lncRNAs [48]. A recent functional example is a lncRNA harboring a disease-associated SNP close to *IL18RAP* gene that has been described as a key regulator of genes in the NFKB 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 [49].

In brief, the genomics for CD remains unresolved, with known genetic factors accounting approximately half of the heritability of CD. Termed the 'missing heritability', this critical gap in our knowledge continues to challenge the tremendous efforts made in genomic medicine. Understanding the biological consequences of associated polymorphisms is a very complicated task, which is still far from being fully resolved [50].

3. Pathogenesis of celiac disease

The recent advances in our knowledge on the mechanisms that take part in the development of CD have made it one of the best-understood HLA-linked disorders. However, several pathogenic processes still remain to be described.

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. The most widely-accepted pathogenic model includes altered digestion and transport of gluten across the epithelium. This focuses on adaptive immune mechanisms that depend on stimulation of gluten-reactive CD4+ T cells. 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 healthy, non-celiac individuals. When genetically susceptible individuals are exposed to certain gluten-derived epitopes, the peptides cross the intestinal epithelium into the *lamina propria* where they are deamidated by tissue transglutaminase. Deamidated peptides are presented by HLA-DQ2+ and/or HLA-DQ8+ antigen

presenting cells (APCs) to pathogenic CD4+ T cells, triggering a Th1-mediated response that leads to the infiltration of the epithelial *lamina propria* by inflammatory cells, together with crypt hyperplasia, and villous atrophy [51].

Studies in the last decade have also stressed the role of the innate immune response in the pathogenesis of the disease, and it has been shown that gliadin can also activate a non-T cell mediated response [52, 53] (Figure 4). Gluten has a direct toxic effect on the epithelium, in which the main mediator is IL-15 and this is manifested by the expression of stress molecules in enterocytes and the activation of CD8+ intraepithelial T-cell cytotoxic function.

Some aspects of the pathogenesis still need to be clarified, especially regarding the interaction between gluten and epithelial cells, passage of gluten peptides into the *lamina propria*, TG2 activation, mechanisms that regulate IL-15 expression, and autoantibody production.
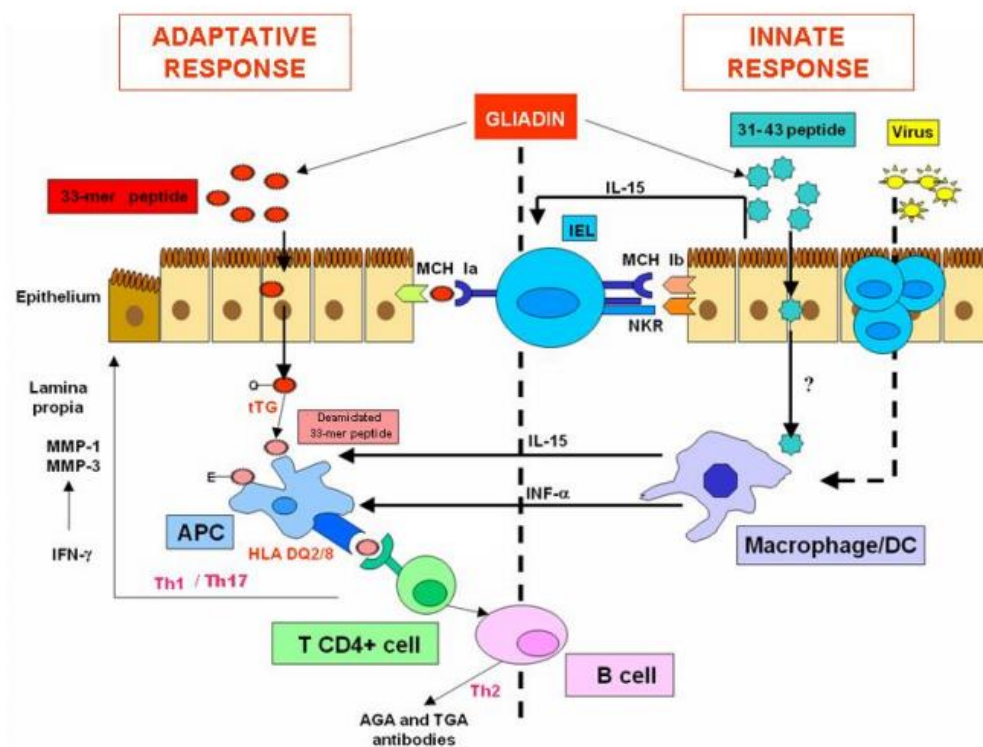

Figure 4: Pathogenic mechanisms of celiac disease: adaptive and innate immune branches. Adapted from Castellanos-Rubio et al., 2010 [54].

3.1. Gluten

CD is triggered by the exposure to gluten proteins in the diet. Gluten is a mixture of monomeric alcohol soluble glutenins, and polymeric prolamin rich gliadins that is found in the endosperm of cereals like wheat, barley and rye. Gluten proteins are very resistant to degradation by intestinal proteases so that long fragments (10-50 residues) are present in the gut lumen. These fragments are good substrates for the enzyme tissue transglutaminase type 2 (TG2), which can deamidate gluten peptides increasing their ability to bind to HLA-DQ2 or HLA-DQ8 molecules leading to a gluten-specific CD4+ Th1 T-cell response. A unique 33-mer peptide from α-gliadin is the most potent T cell stimulator after its deamidation by tissue transglutaminase. In addition, different *in vitro* studies support the idea that gluten, specifically peptide 31-43 from α-gliadin (31-43 peptide), can also trigger CD8+ T cell mediated innate immune responses in the *lamina propia* and may expand the IEL population independently of MHC presentation [51, 55].

3.2. Transglutaminase

Tissue transglutaminase (TG2) is the key component that explains the activation of the adaptive immune response to gluten because it induces enzymatic modification of gliadin peptides and increases their affinity for the HLA-DQ2 and HLA-DQ8 molecules [56]. This enzyme catalyzes the covalent and irreversible cross-linking of a protein with a glutamine residue to a second protein with a lysine residue [57, 58]. Gluten is rich in prolines and glutamines, and has very few negative residues (necessary to bind to the groove of HLA-DQ2 or HLA-DQ8) so when gluten derived peptides are deamidated by TG2, this reaction 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 [59-61]. TG2 deamidation transforms gliadin peptides from a non-stimulatory molecule into an efficient T-cell antigen capable of evoking a massive secretion of local cytokines, which leads to alterations in enterocyte differentiation and proliferation. Additionally, TG2 mediated crosslinking between gliadin peptides and the enzyme induces the formation of TG2-gliadin complexes that trigger the

production of IgA-class autoantibodies [62]. Whether anti-transglutaminase antibodies participate in the pathogenesis of the typical mucosal lesion of the disease, or only represent a bystander event in CD is still unclear but effects on cell cycle, apoptosis, angiogenesis and intestinal permeability have been reported, suggesting that TG2 antibodies could be pathologically relevant [62, 63].

3.3. Adaptive immunity

Adaptive immunity includes T cell-mediated and humoral immunity and both (arms?) are activated in the small intestinal mucosa of CD patients, with gliadin as the recognized antigen. CD4+ T lymphocytes from the small intestinal mucosa recognize deamidated gliadin peptides bound to HLA-DQ2 and HLA-DQ8 heterodimers on APCs [64, 65]. Gliadin-specific T lymphocytes from celiac mucosa are mainly of the Th1 phenotype and release prevalently proinflammatory cytokines, dominated by IFN-γ [66, 67] but other Th1-inducing cytokines such as interleukin 18 and IFN-α are also increased [68-70]. A different linage of CD4+ T-helper cells that differentiate in the presence of IL6 and TGFβ and produce interleukin 17 cytokine-family members (Th17 lymphocytes) has been identified and seem to be partly responsible for pathogenic effects previously attributed to the IL12/INFγ network [71]. Both Th1 and Th17 responses are present in the active CD lesion, a phenomenon that has also been described in other immune-mediated conditions [72-74].

3.4. Innate immunity

The innate immune response represents the first line of defense against pathogens, and is activated during the first stages of exposure to an infectious agent. In CD, the innate immune system responds to gliadin in a Tαβ-lymphocyte independent manner and contributes to the creation of the proinflammatory environment necessary for subsequent T cell activation in patients carrying HLA-DQ2 or DQ8.

Several studies have implicated MyD88, the major signal transducer of Toll-like receptor 4 (TLR4) on monocytes, macrophages and dendritic cells, and TLR4 itself as the primary receptor for innate responses to cereal proteins [51]. Innate

immune activation of IELs by gluten induces expression of *MICA* on the intestinal epithelium, a stress-induced molecule that interacts with the NKG2D receptor on natural killer, γδT cells and on subsets of CD4+ and CD8+ T cells. Epithelial MICA together with upregulated IL15 leads to the activation of NKG2D on IELs triggering antigen-specific lymphocyte-mediated cytotoxicity and it is capable to activate innate cytotoxic and cytokine production responses in the initial stages of the disease, linking innate with adaptive immunity. Finally, IL21 has emerged as an additional driving force of innate immunity that often acts in concert with IL15 [53, 75].

3.5. Other pathologic pathways

Most of the CD genes identified so far are involved in the immune response, but much less is known about the secondary events that lead to intestinal tissue destruction and nutrient malabsorption. Although the main driving force of the disease is an aberrant autoimmune response to transglutaminase-deamidated gliadin peptides, there are other biological networks that are also altered and contribute to the final anatomical and histological features of overt disease. Those pathogenic mechanisms are still not well understood.

Genomic approaches, such as cDNA microarrays or whole genome expression analyses have identified important dysfunction of several complex biological processes that could participate in CD development (1-4). These type of approaches have been performed in intestinal biopsies [76] [77] [78] and *in vitro* cell models [79, 80]. Although the main pathways associated with CD in microarray studies are mainly those related with immune cells (T and B lymphocytes and macrophages) and increased activity of the antigen processing and presentation pathway, among the identified networks, there were also the cell-cell communication and signaling, the ubiquitin–proteasome system, cell cycle and apoptosis cascade and the extracellular matrix (ECM) receptor interaction networks.

It is thought that the identification and reconstruction of the biological networks that are potentially pathologic will contribute to the understanding of the pathogenesis of CD, and could point towards novel functional candidates that might be responsible for genetic susceptibility and could be potential therapeutic

targets. This work focuses on the cellular communication network, more specifically in the Tight Junction (TJ) pathway, that has been proposed to be involved in the enhanced epithelial permeability observed in CD.

4. Tight Junction pathway

The intestinal epithelial tissue is composed of a single monolayer of contiguous cells lining the gut lumen. Together with the mucus it forms a biologically flexible and environmentally responsive barrier to luminal contents and it fulfills two main functions. First, it acts as a barrier separating the body from the colonic microflora, foreign antigens, microorganisms and their toxins. Second, the intestinal epithelium acts as a selective filter which allows the translocation of essential dietary nutrients, electrolytes and water from the lumen into the circulation via either transcellular or paracellular pathways [81].

Intercellular junctions connecting the epithelial cells regulate the paracellular barrier and are essential for the maintenance of the intestinal barrier function. The intercellelular junctional complex is formed by proteinaceous adhesive contacts and contains three different structures: Tight Junctions (TJ) or *zonula occludens* close to the luminal surface (the most apical junction), Adherent Junctions (AJ) or *zonula adherens* just below TJs and desmosomes or *macula adherens* underneath [82] (Figure 5). Although AJ and desmosomes convey the mechanical linkage between adjacent cells, TJ are responsible for the formation of functional barriers that regulate the passage of cells and solutes through the paracellular space [83].
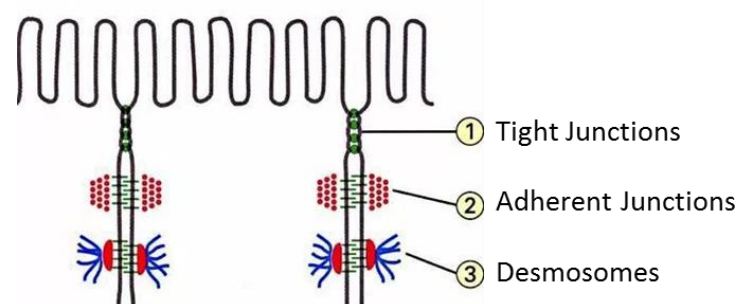

Figure 5. Schematic representation the intercellular junctions in epithelial cells. Adapted from www.pathologystudent.com.

4.1. Structure of tight junctions

Freeze-fracture electron microscopy has revealed TJ as a series of anastomosing membrane strands that completely encircle the cells, resulting in a continuous junctional belt that interconnects neighboring cells [82] (Figure 6, A). Based on these morphological observations and on subsequent functional studies, two main functions have been classically described for TJ; regulation of the flux of ions and solutes or "gate function", and the establishment and maintenance of apico-basal polarity or "fence function" [84]. Thus, TJ serve as a barrier to pathogens and bacterial products while also limiting apical plasma membrane proteins to the lumen-facing domain of the epithelial cell.

At the molecular level, the TJ is a highly diverse structure composed of both transmembrane and cytoplasmic proteins (Figure 6, B). In the membrane domain, TJ is mainly composed of transmembrane proteins occludin [85], and claudins [86]. The claudin family consists of 24 members which can be divided into sealing claudins, that reduce permeability, and pore-forming claudins, that increase permeability [87, 88]. Besides occludin and claudins, there are additional classes of transmembrane proteins including tricellulin and junctional adhesion molecules (JAMs) that function as adhesion proteins and regulate various processes such as leukocyte transmigration [89, 90].

Underlying the membrane domain is the cytoplasmic plaque, a dense plaque of scaffolding molecules that connect the transmembrane proteins to the actin cytoskeleton as well as to different types of signaling proteins [91]. Prominent examples are the Zonula Occludins (ZO-1, -2 and -3) and MAGUK family proteins. These cytosolic adaptor proteins contain multiple protein-interaction domains (PDZ, SH3 and GK) through which they demonstrate affinity for a number of cytoskeletal proteins, signaling molecules and membrane proteins [92, 93]. They can engage in interactions with different proteins such as AJ proteins (a-catenin and afadin/AF-6), polarity proteins (Par3-Par6-aPKC polarity complex), or transcription factors that regulate proliferation (ZONAB/DbpA) [91, 94].

This complex network is thought to be necessary for the correct organization of the integral membrane components of the junction, the regulation of junction

assembly and function, as well as for the transmission of signals to the cell interior [95]. In summary, this molecular signaling apparatus controls the localization and function of transmembrane proteins, and thus the permeability of the intestinal barrier.
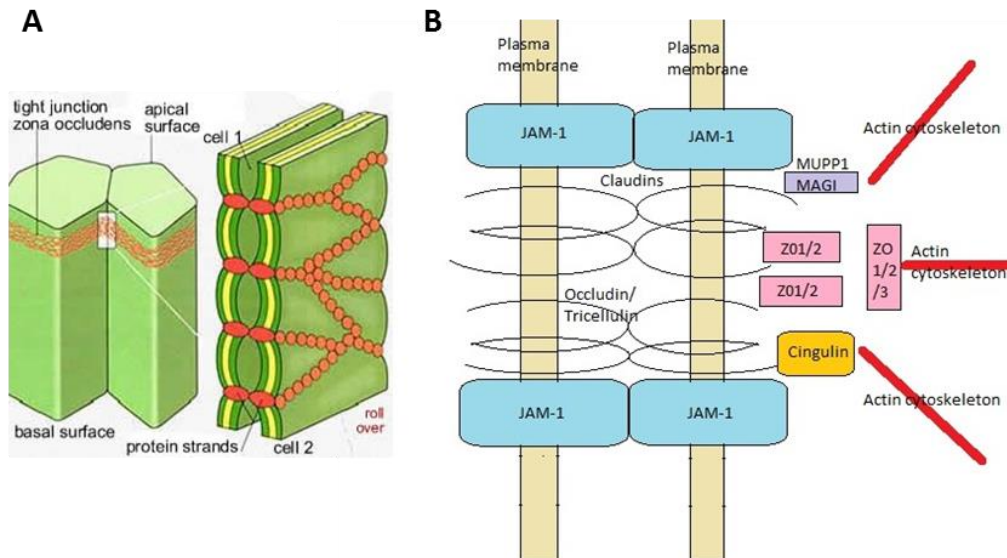
Figure 6. Schematic illustration of the a) ultrastructural and b) molecular basic structural transmembrane components of tight junctions.

4.2. Tight Junctions in celiac disease

Impairment of the epithelial barrier and increased permeability have been shown to be critical determinants in intestinal inflammation and to be involved in a number of gastrointestinal diseases such as Crohn's disease [96], ulcerative colitis [97] and CD [98]. Sixty to eighty percent of CD patients show increased intestinal permeability, leading to an increased entry of luminal antigens/microbes to the mucosal immune environment [99, 100]. It has been suggested that in CD the increased paracellular permeability may contribute to the transepithelial passage of gliadin peptides, facilitating disease development. However, it is still a matter of debate whether TJ contributes or if the passage of gliadin peptides occurs exclusively transcellularly.

Celiac patients are known to have an abnormal TJ structure and an increased intestinal permeability that is often associated with TJ barrier disruption [101,

102]. Ultrastructurally, TJ strands have been shown to be reduced in depth and in number in active CD and to show strand discontinuities. Molecular analysis has shown that there are complex alterations of integral TJ proteins including gene expression alterations, posttranslational modifications such as tyrosine phosphorylation, mislocalization of adaptor molecules and cytoskeletal rearrangement [103]. The abnormal expression and distribution of TJ proteins seem to have a major role in determining proper TJ function. The most remarkable findings described in the small intestine of CD patients are the upreglation of the pore-forming claudin-2 and the downregulation of sealing-related proteins (claudin-3, claudin-5 …), transmembrane occludin and main adaptor ZO-1 [104-106]. Interestingly, together with the data commented above, a TJ-related candidate gene approach showed 2 regulators of epithelial polarity to be associated with CD, *PAR3* (Partition-defective) and *MAGI2* (Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2), suggesting that mechanisms of cell polarity might contribute to TJ integrity [107]. Moreover, changes in cell polarity proteins Par-3 and PP-1 have been associated with altered expression and assembly of TJ proteins such as claudin-2, -3, -5 and -7 and ZO-1, causing paracellular leakage in active CD [104].

A mechanism of modulating TJ permeability in CD has been proposed by Alessio Fasano's group [108]. Zonulin is an endogenous homologue to the ZO toxin of Vibrio cholerae and was shown to be secreted by intestinal epithelial cells and *lamina propria* macrophages upon gliadin exposure, leading to increased paracellular permeability. They proposed zonulin as a regulator of epithelial barrier functions linked to several chronic inflammatory diseases including CD. Zonulin has reached the clinical bedside as larazotide, a drug targeting the zonulin system, which was explored in CD patients as a supplement to a gluten-free diet [109].

Although there is very little information from CD patients prior to the onset of the disease, clinical and experimental studies suggest that altered intestinal barrier function might play an inciting role in the development of CD, by allowing gliadin to cross the intestinal barrier and activate the immune system [110]. The

pathogenic role of intestinal TJ disruption in CD is supported by studies demonstrating that increased intestinal permeability exists prior to disease onset [111], persists in asymptomatic patients who are on a GFD [98] and is also present in a significant proportion of healthy first-degree relatives of patients with CD [112].

In summary, the integrity of the epithelial barrier including normal TJ function is known to be compromised in CD and this seems to be in part affecting the susceptibility to disease. However, mechanisms triggering the barrier defect are still not well understood, neither is the function of the genes that are believed to take part in the impairment of the epithelial barrier in CD. We stress the need to decipher TJ pathway implication in celiac disease pathogenesis by identifying the genetic factors implicated in these phenomena and developing functional studies in order to shed light on the complex genetics and pathogenesis of common disorders.

2. Aims of the study

The present work has three main objectives that aim to contribute to our understanding of the implication of TJ in CD pathogenesis:

I. To determine whether the TJ pathway is altered in CD and to investigate the effects of gliadin on these alterations.

    a. To analyze the expression and coexpression patterns of the TJ-related genes in the intestinal mucosa of celiac patients and controls.

    b. To assess the effects of gliadin in the formation of TJ-dependent intestinal monolayer structures.

II. To characterize the functional contribution of CD-associated TJ-related polymorphisms in the intestinal barrier integrity.

    a. To replicate in an independent population the association results of TJ-related variants previously associated with CD.

    b. To functionally characterize the genes within the associated region and to determine whether disease-associated variants have any influence on candidate gene expression.

III. To create a cellular model with stable modifications in TJ-related genes that are constitutively altered in CD, and to analyze their role in CD pathogenesis.

3. Material and Methods

1. Material

1.1. Subjects

1.1.1. Ethical approval

All the studies performed are part of Research Projects PI13/01201 (within the National Plan for Scientific Research, Development and Technological Innovation 2013-2016, co-financed by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund) and 2011/111034 from the Basque Department of Health, which were approved by the Cruces University Hospital and Basque Autonomous Clinical Trials and Ethics Committees (codes CEIC-E13/20 and PI2013072, respectively). All samples were collected during routine diagnosis endoscopy and after informed consent were obtained from patients or their parents.

1.1.2. Patients and biopsy samples

CD was diagnosed in the Pediatric Gastroenterology Clinic at Cruces University Hospital, according to the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPHGAN) criteria in force at the time of recruitment, including determination of antibodies against gliadin and endomysium (EMA) or tissue transglutaminase (TGA) as well as a confirmatory small bowel biopsy. For the present study, a total of 103 duodenal biopsies from 36 celiac patients and 31 non-celiac patients were collected by endoscopy, and divided into 3 groups:

- Active CD group (36): newly diagnosed CD patients with clinically active disease (positive for CD-associated antibodies and presenting atrophy of intestinal villi with crypt hyperplasia) who were on a non-restricted (gluten-containing) diet at that time.

- Treated CD group (36): normalized CD patients (asymptomatic, antibody-negative and with a recovered intestinal epithelium) from the active group after strict GFD for more than two years.

- Control group (31): non-celiac individuals without intestinal inflammation at the time of endoscopy.

43

Biopsies were snap frozen and stored in liquid nitrogen until use. A summary of clinical, immunological and HLA information of the celiac patients included in the study is available in table 1 (extened data in Suplementary table 1).

Table 1. Clinical, immunological and HLA information of the celiac patients included in the study. Dx: Diagnosis.

Characteristics	Values
Gender	**Female:** 63%; **Male:** %27
Age at Dx (months)	23,1 ± 22,8
Time on GFD (years)	1,97 ± 0,37
TGA at Dx (%POS)	119 ± 87,45
MARSH score at Dx	**3c:** 22; **3b:** 5
HLA type	**DQ2:** 20; **DQ8:** 1; **DQ2/DQ8:** 6

1.2. DNA and RNA samples

DNA samples used for association studies belong to the Spanish Consortium for the Genetics of Celiac Disease (CEGEC). A total of 361 samples were used, of which 270 were from non-celiac individuals and 91 from celiac patients.

1.3. Human cell lines

Human epithelial cell lines were obtained from the American Type Culture Collection (ATTC, Manassas, VA, USA). Colon-derived C2BBe1 [clone of Caco-2] (cat. no. CRL-2102), T84 (cat. no. CCL-248) and HCT116 (CCL-247) cell lines, and embryonic kidney cell line HEK293FT (cat. no. CRL-1573) were used in the present project.

2. Methods

2.1. RNA, DNA and protein isolation from tissues and cultured cells

Frozen biopsy samples were disrupted with disposable plastic pellet pestles. Total RNA and DNA were isolated from tissues and cultured cells using the NucleoSpin RNA kit and NucleoSpin Blood kit respectively (Macherey-Nagel, Düren, Germany) according to the protocol supplied by the manufacturer.

For RNA nuclear fractionation, nuclei of 10^7 cultured cells were isolated using hypotonic lysis buffer (10 mM Hepes, 10 mM KCl, 0.1 mM EDTA and 1% Triton-X-100). After 15 minutes on ice, cells were centrifuged for 15 min at 2500 rpm and the nuclear pellet was used for RNA extraction as above.

The concentration and purity of the RNA and DNA samples were determined by UV spectrophotometry on a NanoDrop 1000 apparatus (Thermo Fisher Scientific, Boston, MA, USA) and samples were stored at -80ºC until use. To obtain total cellular proteins, cultured cells were directly lysed in 2x Laemmli Sample buffer containing 5% 2-Mercaptoethanol (BioRad, Hercules, CA, USA) and stored at -20ºC until use.

2.2. Gene expression analyses (qRT-PCR)

To quantify RNA transcription levels, two-step quantitative real-time polymerase chain reaction (qRT-PCR) technology was used. Samples analyzed belonged to 3 groups: 1) duodenal biopsies, 2) cultured cells and 3) total RNA of different human tissues from the Human Total RNA Master Panel II kit (Clontech, Mountain View, CA, USA).

2.2.1. Candidate Genes and Assays

For the gene expression profiling 52 candidate genes were selected (24 from the TJ pathway and 28 related to the TLR pathway and inflammation signals), using the following criteria:

a) genes that had previously shown genetic association or a significant alteration of expression in CD [54, 104, 107].

b) genes with a central function in the TJ or TLR networks according to the KEGG database (http://www.genome.jp/kegg).

c) genes that have been described in the literature as important functional players in TJ [95, 113].

Primer and probe sets commercially available as TaqMan Assays-on-Demand (Thermo Fisher Scientific, Boston, MA, USA) were purchased for protein-coding gene analyses (Supplementary table 2). Whenever possible, assays with probes that spanned an exon–exon junction were chosen to avoid the detection

of genomic DNA. The expression analysis of non-coding transcripts was performed with SYBR green detection and primers were designed using Primer3 software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Table 2). To confirm the specificity of these primers, BLAST search was performed against public databases to ensure that they only recognized the target of interest.

Table 2. Primer sequences (5' to 3') for non-coding transcript expression analysis. *RPLP0* **was used as a housekeeping control.**

Assay_ID	Amplicon size (bp)	Reverse primer	Forward primer
RP4-587D13.2	172	GGTGCTGGAAATTCATCAGTG	TGACCACATGACTGACACCA
RPLP0	85	GCAGCATCTACAACCCTGAAG	CACTGGCAACATTGCGGAC

2.2.2. Retrotanscription or cDNA synthesis

Total RNA was reverse transcribed with the iScript cDNA Synthesis Kit (BioRad, Hercules, CA, USA). Equal amounts of RNA input and a mixture of oligo(dT) and random primers were used to obtain an uniform representation of all targets analyzed and to avoid the 3' bias of oligo(dT) primers. RT was performed using the following cycling conditions: priming at 25ºC for 5 minutes, retrotranscription at 42ºC for 30 minutes and inactivation at 85ºC for 5 minutes. cDNA samples were stored at -20ºC until use.

2.2.3. Expression analyses: quantitative PCR (qPCR)

All qPCR expression analyses were performed in duplicate and the expression of the housekeeping gene *RPLPO* (large ribosomal protein) was simultaneously quantified in each experiment and used as an endogenous control of input RNA. Two different systems were used for qPCR depending on the number of samples analyzed simultaneously: Fluidigm BioMark dynamic array system (Fluidigm Corporation, San Francisco, CA, USA) for large numbers of samples and Eco Real-Time PCR System (Illumina, San Diego, CA, USA) for small numbers of samples. After the analyses, the relative expression of each gene was calculated using the accurate Ct method as previously described [114]. The

amount of target, normalized to an endogenous reference and relative to a calibrator, is given as:

$$\text{Relative expression} = 2^{-\Delta\Delta CT}$$

Where ΔΔCt is the difference in the ΔCt values between the experimental and control samples and ΔCt is the difference between the Ct values of the gene of interest and the housekeeping gene.

Fluidigm BioMark dynamic array system

Expression analyses were performed at the Sequencing and Genotyping Unit of the University of the Basque Country (UPV/EHU). First preamplification of reverse transcribed cDNA from each sample was carried out using the TaqMan PreAmp Master Mix (Applied Biosystems, Carlsbad, CA, USA). Conditions were set, following the manufacturer's instructions, at 14 cycles of preamplification. For qPCR analyses, TaqMan Fast Advanced Master Mix (Applied Biosystems, Carlsbad, CA, USA), 2X Assay Loading Reagent (Fluidigm Corporation, San Francisco, CA, USA) and 20X GE Sample Loading Reagent (Fluidigm Corporation, San Francisco, CA, USA) were employed. Five microliters each of 10X Assay Premix (mix of 2.5 µl 20X TaqMan GE Assay and 2.5 µl 2X Assay Loading Reagent) and Sample premix (mix of 2.5 µl 2X TaqMan Fast Advanced Master Mix, 0.25 µl 20X GE Sample Loading Reagent and 2.25 µl cDNA) were used per inlet. Cycling conditions were as follows: polymerase activation at 95ºC for 1 minute and 35 amplification cycles at 95ºC for 5 seconds and at 60ºC for 20 seconds.

Eco Real-Time PCR System

Analyses using TaqMan assays were carried out using TaqMan Universal PCR master mix (Thermo Fisher Scientific, Boston, MA, USA) and following the manufacturer's instructions. Cycling conditions were as follows: polymerase activation at 95ºC for 10 minutes and 40 amplification cycles of 15 seconds at 95°C and 1 minute at 60°C. In SYBR Green detection-based experiments, Quantitect SYBR Green master mix (Qiagen, Hilden, Germany) was used and

the cycling conditions were: 30 seconds at 95 °C, followed by 40 cycles of 15 seconds at 95 °C and 1 minute at 60 °C. Melting curve analysis was performed following the provided protocol to verify the specificity and identity of the PCR products.

2.3. SNP genotyping and association study

Genotyping was performed at the Sequencing and Genotyping Unit of the University of the Basque Country (UPV/EHU) using Dynamic Array Integrated Fluidic Circuits FR48.48 reaction chips in a BioMark HD platform (Fluidigm Corporation, San Francisco, CA, USA). In accordance wuth what had been previously published, 3 SNPs in *MAGI2*, 2 in *PARD3* and 3 in *MYO9B* were selected [107] and specific primers and probes were designed (Table 3). To ensure population homogeneity, identity-by-state analysis and multidimensional scaling were performed before standard χ^2 association analysis using Plink 1.07 [115].

For expression quantitative trait *loci* (eQTL) analyses, SNP rs2691527 was genotyped using predesigned TaqMan Genotyping assay (Thermo Fisher Scientific, Boston, MA, USA; assay. no. c_2577703_10). This SNP was selected because it is in complete LD with the associated SNP rs6962966 (r2 =1). DNA amplification was carried out on an Eco Real-Time PCR System (Illumina, San Diego, CA, USA) in 20 µl reaction volume PCR containing 10 ng DNA, 1X TaqMan assay and 1x TaqMan genotyping Mastermix (Thermo Fisher Scientific, Boston, MA, USA). Cycling conditions were 95°C for 15 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute.

2.4. Cell culture conditions and treatments

2.4.1. Cell culture conditions

Human epithelial cells were cultured at 37°C in a 5% CO_2, humidified atmosphere according to standard mammalian tissue culture protocols. All tissue culture media and supplements (Table 4) were obtained from Lonza

(Lonza group, Basel, Switzerland) except FBS (Biochrom, Cambridge, UK; cat. no. S 0415), which was heat inactivated following the manufactures instructions.

Cells were subcultured after reaching a density of 80 %. Monolayers were detached using 0.25 % Trypsin-EDTA solution (Gibco, Carlsbad, CA, USA; cat. no. 25200056), washed in PBS and seeded in growth medium at low density. For cryopreservation, cryopreservation cells were resuspended in freezing medium, growth medium with %5 DMSO (Sigma-Aldrich, St. Louis, MO, USA), frozen at -1ºC/minute in an isopropanol filled freezing container placed in a -80 °C freezer for 24 hours and afterwards stored in liquid nitrogen. For resuscitation of frozen cells, vials were placed in a water bath at 37°C until thawed before resuspending in growth medium and transfering into culture flasks.

2.4.2. Gliadin (PT-G) preparation for cell stimulation

For *in vitro* cell stimulation, pepsin-trypsin digest of gliadin (PT-G) was prepared as previously described [Bondar C., et al., 2014]. Briefly, 500 mg of gliadin (Sigma-Aldrich, St. Louis, MO, USA; cat. No. G3375) were dissolved in 5 ml 0.2N HCl and incubated with 5 mg of pepsin (Sigma-Aldrich, St. Louis, MO, USA; cat. no. P6887) with continuous agitation at 37º C for 2 hours. After incubation, pH was adjusted to 7.4 with NaOH and the mixture was incubated with 5 mg of trypsin (Sigma-Aldrich, St. Louis, MO, USA; cat. no. T9935) at 37º C for 4 hours. The solution was heated in a boiling water bath for 30 minutes to inactivate the enzymes and the digest was centrifuged at 2000 g for 10 minutes. The supernatant corresponding to PT-G was sterilized by filtering through a 20 µm pore membrane, aliquoted and stored at -80 ºC. An enzymatic digest of BSA (PT-BSA) (Thermo Fisher Scientific, Boston, MA, USA; cat. no. SH30574.03) was prepared in the same way and used as a negative control of stimulation.

Table 3. Designed Fluidigm primers (5' to 3') for genotyping experiment.

ASSAY_ID	SNP_NAME	SNP allele	Reverse primer	Forward primer
GTA0093555	rs6962966	A / G	TGCCCACAAGGATGACCTCA	GGTTGTGTCTTTGAGCCACAT
GTA0093556	rs9640699	A / C	GTAAAGAGTACCTCCAATTGTATGA	GCAGTGGATTCTCTGTACAGC
GTA0093558	rs1496770	C / T	AGAGACAAGAAGACACACATAAAA	ACAAATTCAGATTACTTAATGTCAGTTTGA
GTA0093550	rs10763976	G / A	CGTGTGGTCCACCAGGTAGA	ACATTAGTTCAGGCTACTGACCA
GTA0093551	rs4379776	C / T	TCTTTCCAAAGCTGGAGGCAGA	TCAAATGTGGTGTTACTGCTGG
GTA0093559	rs2305767	C / T	GCATCCTGCTGTGATCTGGG	GGGCTGGCTAGGAATCTACA
GTA0093552	rs1457092	C / A	CATCCACCGGGCACAGAGA	CTGGGCTCGGAGCTGT
GTA0093554	rs2305764	G / A	ACGACCCATGGGATCTCAG	TCTGCACGCGTATGTGT

Table 4. Complete media for cell lines. DMEM= Dulbecco's Modified Eagle's Medium; FBS= Fetal Bovine Serum; Pen-Strep=

Penicillin-Streptomycin; NEAA= Non-Essential Amino Acids.

Cell line	Medium	FBS	Pen-Strep	NEAA
C2BBe1	DMEM (high glucose)	10%	1%	0.1 mM
T84	DMEM: F-12	10%	1%	-
HCT116	DMEM: F-12	10%	1%	-
HEK293FT	DMEM (high glucose)	10%	1%	0.1 mM

2.5. siRNA and plasmid transfection

Dicer-substrate short interfering RNAs (DsiRNAs) against human *MAGI2* (Integrated DNA Technologies, Coralville, IA, USA Cat. no. hs.Ri.MAGI2.13.1 and hs.Ri.MAGI2.13.2) and a non-target control (Cat. no. 51-01-14-03) were used at final concentrations of 10 nM. Reverse transfection using Lipofectamine RNAiMax (Thermo Fisher Scientific, Boston, MA, USA) was performed according to the procedure provided by the manufacturer. Briefly, siRNA and lipid were diluted in Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific, Boston, MA, USA) and incubated for 15 minutes in treated 24-well culture plate (Sigma-Aldrich, St. Louis, MO, USA). C2BBe1 cells ($5x10^4$ cells per well) were suspended in antibiotic-free growth medium and plated directly into the culture vessel containing the siRNA and lipid mixture. The medium was changed after overnight incubation at 37°C in 5% of CO_2. Forty-eight hours after transfection, C2BBe1 cells were stimulated with 1mg/ml PT-G or PT-BSA for 4h. The efficiency of knockdown was confirmed by Western blot assay.

2.6. Immunoblot analysis.

All the reagents used for Immunoblot were purchased from BioRad (BioRad, Hercules, CA, USA) except otherwise specified. Total proteins in Laemmli buffer were denatured at 95º C for 5 minutes and 30 µl were loaded into %12 SDS-page gels and run at 100V for 30 minutes. Subsequently, proteins were transferred to nitrocellulose membranes using the standard protocol (25V – 1.0A – 30 min) in a Trans-Blot Turbo Transfer equipment. After transfer, membranes were blocked for 1 hour in 5% non-fat milk in Tris-buffered saline Tween buffer (TBST). Membranes were incubated overnight at 4º C in the same buffer with primary antibodies against MAGI2 and β-Actin (1:500 and 1:1000 respectively) (ABCAM, Cambridge, UK; cat. no. ab97343 and ab8227 respectively) followed by an incubation of 1 hour with HRP conjugated anti-rabbit IgG (1:2000) and anti-mouse IgG secondary antibodies (1:1000) (Jackson ImmunoResearch Laboratories, Inc; cat No. 111-035-003 and 115-035-062). Proteins were visualized using SuperSignal West Femto Maximum

Sensitivity Substrate (Thermo Fisher Scientific, Boston, MA, USA) on a ChemiDoc MP system.

2.7. Permeability studies

Macromolecular and ionic permeability seem to be quite independent events, as they sustain controversial remodeling under the same conditions [116]. Thus, two methods were used for the assessment of paracellular permeability in C2BBe1 epithelial cell model: 1) the measurement of the transepithelial electrical resistance (TEER) of the monolayer for ionic permeability and 2) the Lucifer Yellow (LY) assay for macromolecular permeability. Cells ($2x10^5$) were seeded in 0.33 cm² cell culture inserts with 0.4 µm pore size (Sigma-Aldrich, St. Louis, MO, USA; cat. no. CLS3470) in a 24-well plate and grown during 17 days for monolayer formation. During that time, TEER and LY assay measurements were routinely performed.

TEER was measured using a Millicell Electrical Resistance System (ERS) Voltohmmeter (Merck Millipore, Billerica, MA, USA), with two electrodes located in the upper and the lower compartment (Figure 1). Measurements were done within 5 min after removing the culture plates from the incubator. The ohmic resistance of a blank (culture insert without cells) was measured in parallel. To obtain the sample resistance, the blank value was subtracted from the total resistance of the sample. The final unit area resistance ($\Omega*cm^2$) was calculated by multiplying the sample resistance by the effective area of the membrane (0.33 cm^2). TEER > 600 $\Omega *cm^2$ was taken as cell confluency value.
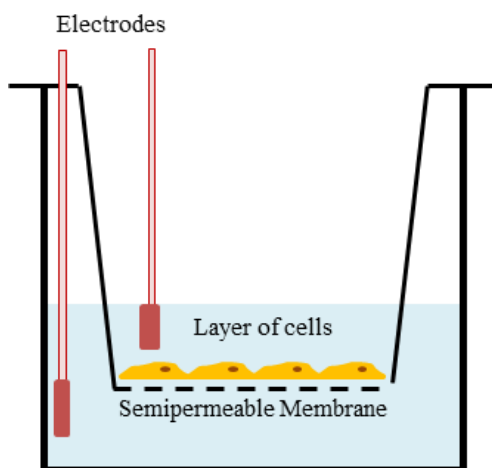

Figure 7. TEER measurement with electrodes. The cellular monolayer is grown in complete media on top of a semipermeable membrane. For electrical measurements, two electrodes are used, with one electrode placed in the upper compartment and the other in the lower compartment. The electrodes are separated by the cellular monolayer.

For Lucifer Yellow (LY) assays, a 100 µM solution of LY (Sigma-Aldrich, St. Louis, MO, USA; cat. no. L0259) was prepared and 0.2 ml were added to each well on the apical side of the inserts. The well, which constitutes the basolateral side, contained 0.9 ml of fresh transport buffer containing HBSS with Ca $^{2+}$, Mg^{2+}, 10 mM HEPES, pH 7.4 (Lonza, Lonza group, Basel, Switzerland). The monolayers were incubated at 37°C in 5% CO2 for 1 hour with 90 rpm shaking. After an hour, 200 µL of the bottom sample were collected and analyzed in a DTX880 fluorescence reader (Beckman Coulter Inc., Brea, CA, USA) with emission set at 530 nm and excitation at 485 nm. Transport buffer was used as a blank and serial dilutions of LY ranging from 100 to 0.2 µM were used to construct a standard curve of the apical-to-basal permeability coefficient (Pc).

$$Pc = (V/(A \times Ci)) \times (Cf/T);$$

where **V** is the volume of the basal chamber (mL), **A** is the area of the membrane insert (cm 2), **Ci** is the initial concentration of the compound (µM or florescence units/ml), **Cf** is the final concentration of the compound (µM or florescence units/ml) and **T** is the assay time (seconds).

2.8. Proliferation assay

All the reagents used in this experiment were purchased from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO, USA). C2BBe1 intestinal cells were seeded at 2×10^4 density in a 6-well plate and proliferation was measured at days 0, 1, 2 and 3. Each day, cells were fixed with 4% paraformaldehyde for 15 minutes and stained with 0.1% crystal violet for 20 minutes. The bound crystal violet was dissolved in 10% acetic acid by shaking for 20 minutes, diluted 1:4 in H_2O and quantified by absorbance at 590nm wavelength in a DTX880 fluorescence reader (Beckman Coulter Inc., Brea, CA, USA). Water was used as a baseline control.

2.9. Wound healing assay

C2BBe1 (50,000) cells were seeded in chambers containing silicone inserts that create a 500 µm thick wall between two chambers (Ibidi, Martinsried, Germany; cat. no. 81176). Inserts were removed when cells had grown to confluency and

cultures were placed in complete medium and photographed in a microscope during 36 hours at 15 minute intervals. The area covered by migrating cells was calculated using ImageJ software.

2.10.   CRISPR-CAS9 technique

CRISPR-Cas9 technique was used to permanently disrupt gene function in different epithelial cell models by deleting the first exon of each of the genes of interest. A schematic representation of the procedure is shown in Figure 7.

2.10.1.   *In silico* sgRNA design

The first exons of *INADL*, *TJP1* and *PPP2R3A* were chosen to be deleted in order to increase the chance of provoking full loss-of-function. Up to 1000bp sequence of the upstrean and downstream flanking regions of each exon were used as input for the online software CRISPR Design Tool32 (http://crispr.mit.edu/), to search for protospacer target sequences that can be cut by Cas9 nuclease. The output included several 20bp target options, with different specificity values, based on a statistical algorithm of off target hits. The best outputs for each gene were chosen to ensure efficiency and avoid off-target mutations (Table 5). Blast analysis of the chosen targets was performed against the human genome to increase specificity.

Table 5. Designed sgRNA oligos for CRISPR-Cas9 mediated gene candidates edition.

		Locus (hg19)		Sequence	PAM	Antiparallel	PAM
INADL	1	chr1:	-62207158	CACCGAAAGACTTTCAGGCGGGGCG	CGG	AAACCGCCCCGCCTGAAAGTCTTTC	GCC
	2	chr1:	-62208992	CACCGACTTAATTCGGCCTGCGTGC	GGG	AAACGCACGCAGGCCGAATTAAGTC	CCC
PPP2R3A	1	chr3:	+135683928	CACCGAATCTGGCGAGTGGTTATAA	GGG	AAACTTATAACCACTCGCCAGATTC	CCC
	2	chr3:	+135684886	CACCGTGCGGTGCGGTGCGTTGCCT	CGG	AAACAGGCAACGCACCGCACCGCAC	GCC
TJP1	1	chr15:	+30115225	CACCGCAGGGCAGCTTGACCCGTT	CGG	AAACAACGGGTCAAGCTGCCCTGC	GCC
	2	chr15:	+30112797	CACCGTCGTTAAATATTACGTTGTA	GGG	AAACTACAACGTAATATTTAACGAC	CCC

2.10.2.   sgRNA cloning

Complementary sgRNA oligos were phosphorylated and annealed by mixing 0.2 nm of each oligo in 10X T4 DNA ligation buffer and 0.5 µl of T4 PNK Polynucleotide kinase (New England Biolabs, Ipswich, MA, USA) in a final volume of 10 µl under the following conditions: incubation at 37ºC for 30 minutes, 95ºC at 5 min and 25ºC for 1 minute. Then, the annealed sgRNAs

were introduced into the sgRNA scaffold site of the px459 plasmid vector (Figure 8) (Addgene, Cambridge, MA, USA). Cloning of the oligonucleotides in the vector consisted of a digestion, followed by a ligation reaction. One microgram of vector was digested with fast digest BbsI enzyme (Thermo Scientific) following the recommended reaction conditions and protocol. The digestion was tested by electrophoresis (150 V - 15 min in 1,5 % gel agarose) and subsequently purified with Nucleospin gel and PCR clean up kit (Macherey-Nagel) following the manufacturer's instructions. Digested vector and annealed oligonucleotides were ligated in 1:3 ratio by T4 DNA ligase (New England Biolabs, Ipswich, MA, USA) in ligation buffer at room temperature for 2 hours.
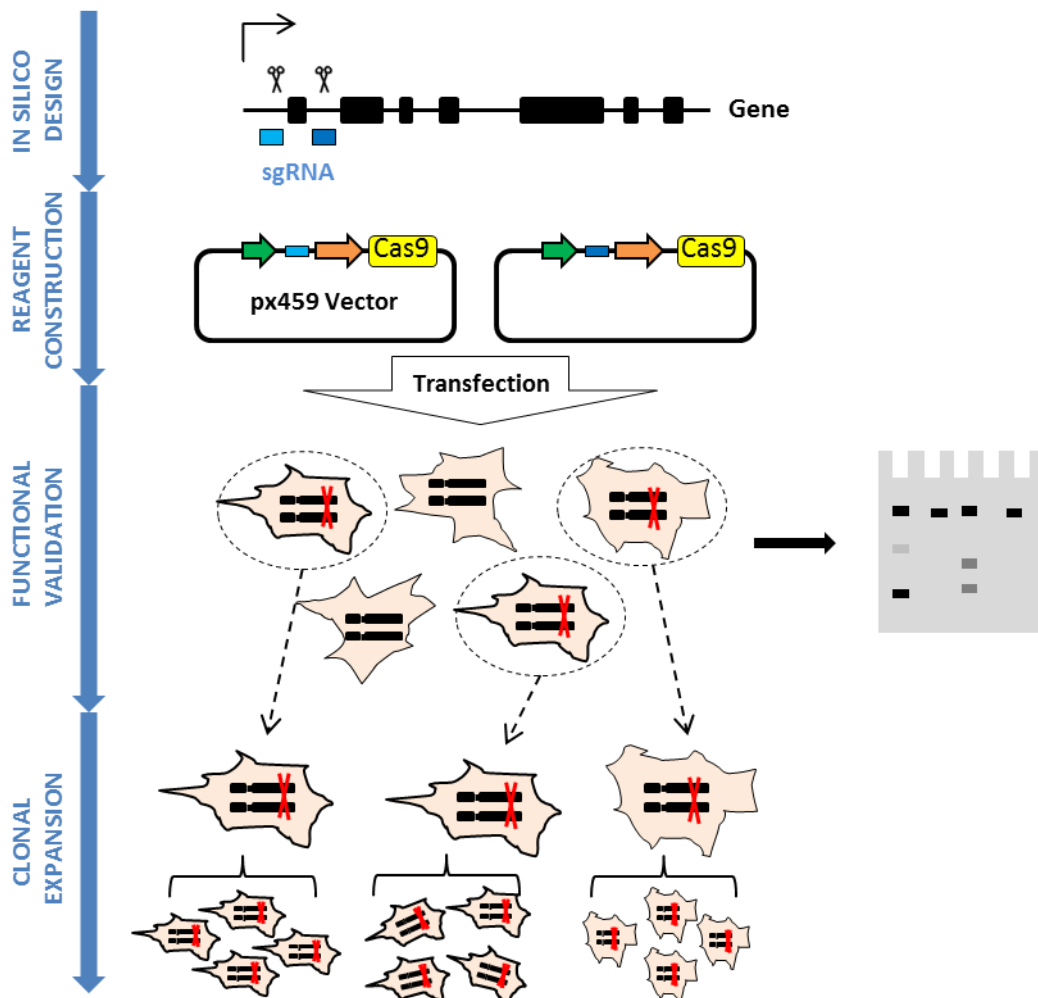

Figure 8. Schematic representation of CRISPR-Cas9 mediated gene-editing mutation of the genes of interest in epithelial cells.
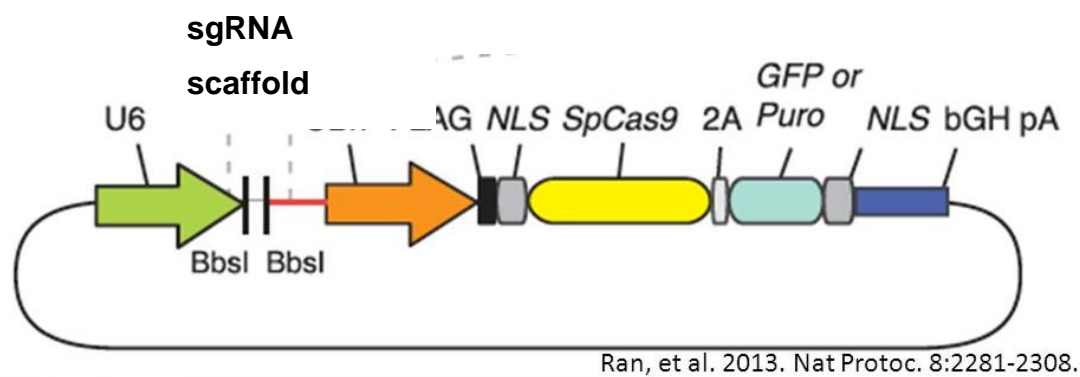
Figure 9. Schematic representation of px459 plasmid.

2.10.3.　　　Bacterial transformation and selection

An aliquot (10 ng) of the ligated construct was transformed into Subcloning Efficiency DH5α Competent Cells (Thermo Fisher Scientific) by heat shock following the supplied protocol. Bacteria were plated and incubated overnight at 37ºC in 50 µg/µl ampicillin-containing LB agar plates. Several colonies were picked and incubated overnight at 37ºC in LB media with antibiotic to check and identify those harboring the right plasmid. For that purpose, plasmid constructs were extracted from bacteria using MiniPrep (MN) according to the manufacturer's instructions and digested with BbsI and BgeI enzymes as previously explained. Visualization of electrophoresic bands was used as a screening for clone selection.

2.10.4.　　　Cell line editing

C2BBe1, T84, HTC116 and HEK293FT cells were reverse transfected with 100 ng of plasmid construct using XTremHP DNA reagent (Invitrogen) in 24-well plates at a density of 100,000 cells per well, following the manufacturer's protocols. After overnight incubation, 10 ug/ml puromycin were added to the cells that were further incubated for another 48 hours for positive selection of transfected cells and enrichment of cell populations with gene editing. For mutational analysis, common PCR was conducted from cell lysates to amplify the target region using flanking primers. Wild-type and truncated genomic fragments were resolved by gel electrophoresis. Cell lysis was performed in 300

µl of prepared Lysis Buffer (100 mM Tris.Cl pH8.0, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) and 5 µl of 20 mg/ml Proteinase K. Samples were incubated for 2 hours at 60ºC and then enzymes were inactivated at 95°C for 10 minutes.

2.10.5.   Clonal expansion

Single cell isolation was performed in a 96 well plate by seeding cells at low density (1 cell per five wells). Isolated cells were expanded until getting healthy colonies and mutational analyses was performed as previously described to select the colonies harboring the mutation. For each colony and each mutated gene, the percent of deletion was quantified at RNA level through real-time qPCR (RT-PCR) (see 3.2.2.3.).

2.11.   Analysis of Results.

Statistical analyses were carried out using GraphPad Prism version 5.0 software (GraphPad Software Inc, La Jolla, CA) to assess differences among groups. Nonparametric paired (Wilcoxon matched-pairs test) and unpaired (Mann-Whitney) tests were performed for the comparison of the expression levels between active and treated patients and between the disease groups and control subjects. Coexpression was assessed by Spearman's correlation test. Heatmaps were constructed using FiRe v2.2 Macro for MS Excel available from the University of Fribourg at http://www.unifr.ch/plantbio/FiRe/main.html. Only 2-tailed $P < 0.05$ were considered statistically significant.

4. Results

1. TJ pathway analysis in CD

1.1. Expression profile

To create the overall picture of TJ pathway function, gene expression profiling of the duodenal mucosa was performed in active CD (n=15), treated CD patients (n=15) and non-CD controls (n=15). When expression levels from active CD patients were compared to those of controls, results revealed that 9 of the 23 genes analyzed were differentially expressed (6 upregulated and 3 downregulated) (Figure 10). In the case of treated patients on GFD, *PPP2R3A* was the only gene that remained significantly altered, being underexpressed when compared to controls.

Comparison of relative expression levels between the groups of patients (active CD *vs.* treated CD) showed that 6 of the 9 altered genes mentioned above followed the same trend as when compared with controls (Figure 11). Additionally, the expression of *CLDN1* and *CDSA* was increased, whereas *CRB3* and *GNAI1* were downregulated in active disease compared to GFD.

We doubled the sample size to 30 samples per group extending the analysis to 15 additional samples. The combined analysis confirmed previous alterations and identified several additional changes (Table 6). When CD patients were compared to controls, expression analysis showed that in active CD, 15 out of 23 genes were altered. Regarding to treated CD patients, most of the genes were expressed at the same level as in controls except for *PPP2R3A*, *INADL* and *TJP1,* which were constitutively downregulated (full data in Supplementary table 3).
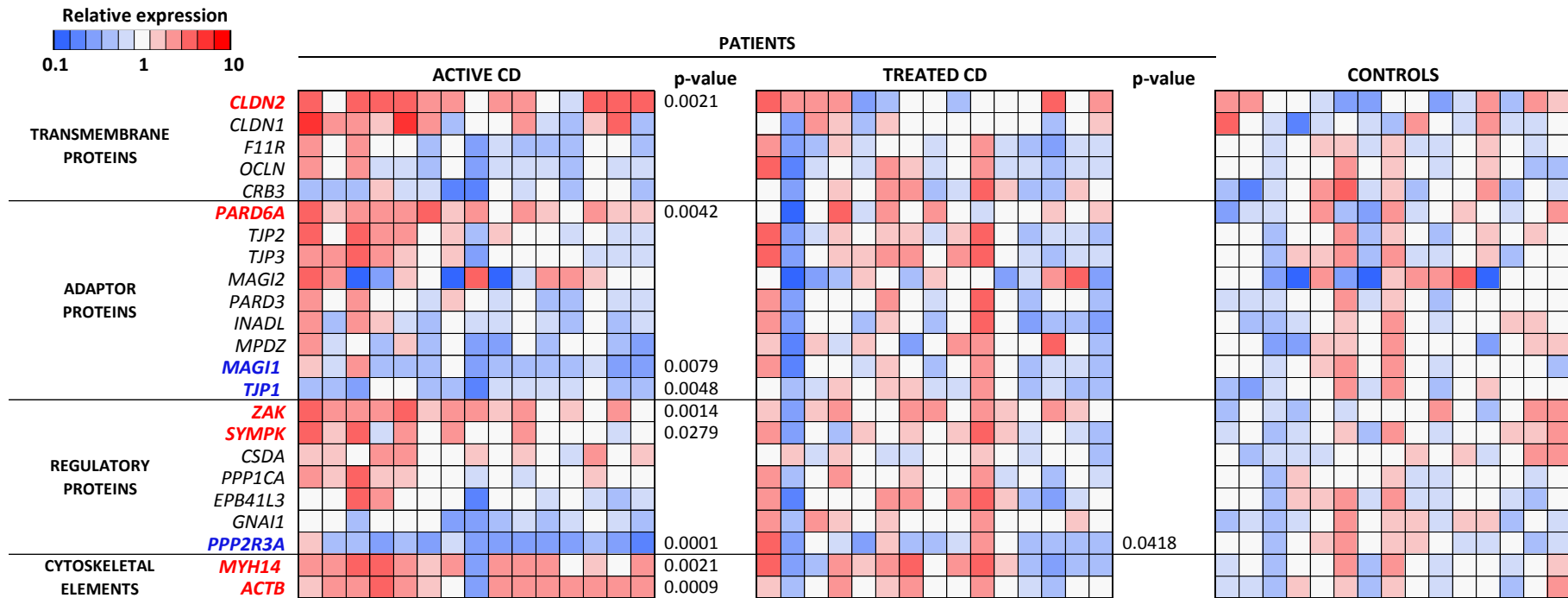
Figure 10. Heat maps of candidate gene expression levels. Each column represents an individual. In the case of active CD and treated CD, since they are paired, the results of the same individual are showed in the same column order. Each line represents a candidate gene, grouped by protein function. The color scale shows the relative expression to the average of control samples for each sample and gene. P-values were determined by Mann-Whitney U test and only significant differences to the control group are shown (p < 0.05). Genes that were altered in active disease compared to controls are shown in red (overexpressed) or in blue (underexpressed).

Figure 11. Relative gene expression of genes that were altered in active CD compared to treated CD (Wilcoxon signed-rank test, p<0.05). The graph shows the mean and the standard deviation of the fold change between active CD and treated CD groups. Red and blue colors represent overexpression and underexpression, respectively.
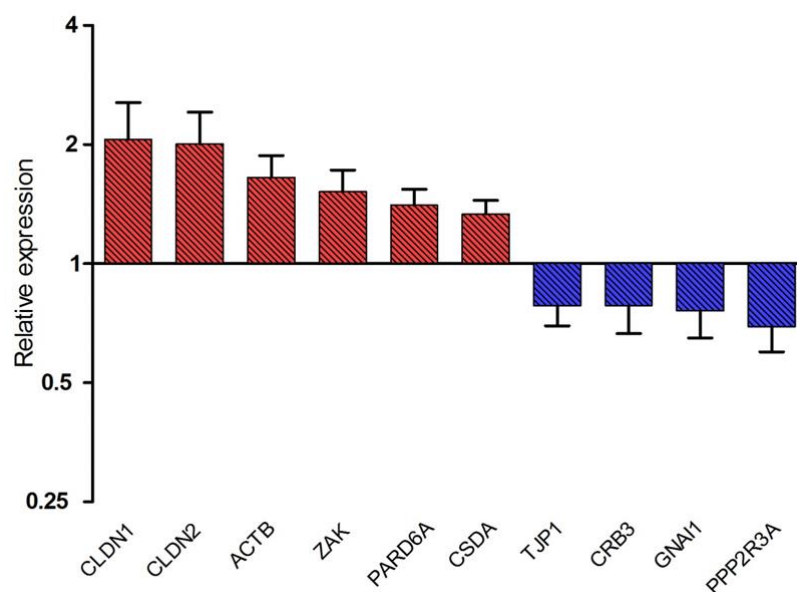
Table 6. Extended analysis of gene expression profiling in active CD (n=30), treated CD (n=30) and control (n=30) samples. Candidate genes are grouped by protein function and fold changes of significant expression alterations are shown (p-values < 0.05). Genes altered are shown in red for those that are overexpressed and in blue for those underexpressed.

		CD / Controls		Active CD /
		Active	**Treated**	**Treated CD**
TRANSMEMBRANE PROTEINS	*CLDN2*	1.62		1.65
	CLDN1	1.56		1.82
	F11R	0.82		
	OCLN	0.83		
	CRB3			
ADAPTOR PROTEINS	*PARD6A*	1.79		1.62
	TJP2			
	TJP3			
	MAGI2			
	PARD3			
	INADL	0.75	0.79	
	MPDZ	0.64		
	MAGI1	0.63		0.70
	TJP1	0.62	0.78	0.79
REGULATORY PROTEINS	*ZAK*			1.28
	SYMPK	1.25		1.35
	CSDA			1.16
	PPP1CA	1.31		1.23
	EPB41L3			
	GNAI1	0.77		0.68
	PPP2R3A	0.46	0.83	0.56
CYTOSKELETAL ELEMENTS	*MYH14*	1.96		1.68
	ACTB	1.77		1.47

1.2. Coexpression analysis

Coexpression studies were performed to clarify whether the alterations observed in expression are affecting the whole TJ network (Supplementary table 2). The analysis showed loss of *CRB3, TJP1, GNAI1,* and *ACTB* coexpression pattern in active CD, whereas these genes maintained certain degree of coexpression in the GFD and control groups (Figure 12). On the contrary, *CLDN2* and *CLDN1* were coexpressed and positively correlated with other 7 genes in patients with active CD, but this coexpression pattern was not found in patients treated for CD and in the control group.
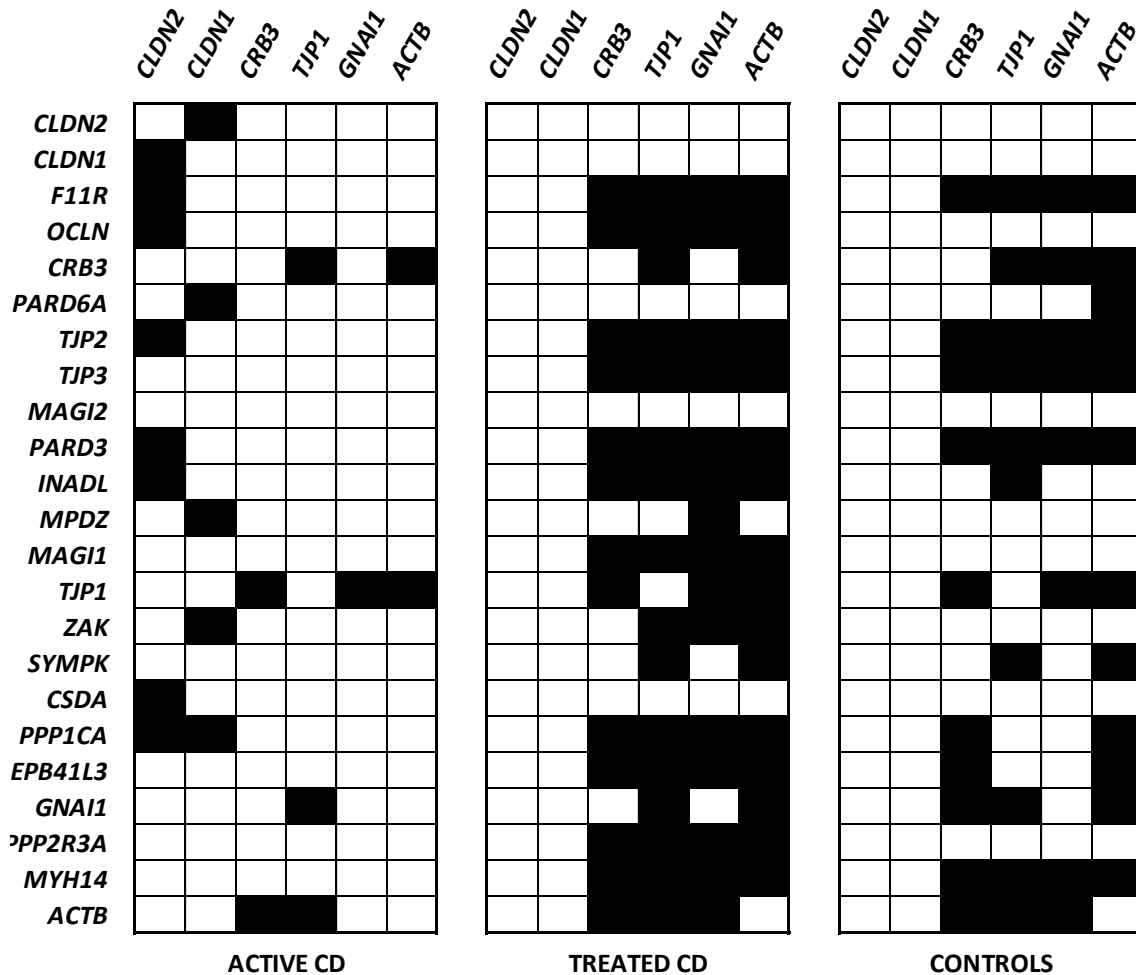

Figure 12. Correlation matrix showing patterns of coexpression among candidate genes in patients with active CD, treated CD and non-CD controls. Black squares represent significant correlation between the corresponding pair of genes (Spearman p-value <0.05).

1.3. TJ structure integrity

Intestinal epithelial cells form a monolayer that serves as a barrier and controls paracellular flux of molecules. In order to demonstrate that TJ regulation in disease is not altered only at gene expression level, but also at the structural level, epithelial cell model C2BBe1 was used to analyze the effects of immunogenic gluten fragments (gliadin) on monolayer integrity and formation.

The ability of C2BBe1 cell line to form monolayers was assessed by measuring the electrical resistance (TEER) and lucifer yellow (LY) paracellular permeability during 17 days (Figure 4 and 5). Under regular experimental conditions (control), cells reached TEER value indicative of monolayer formation at day 7 (819.18 ± 317.01 Ω*cm2) confirming that monolayer is formed in a week. TEER values continued increasing until reaching a maximum on day 13 (1526.8 ± 22.83 Ω*cm^2). Afterwards, values stayed around 1500 Ω cm^2 in all measurements. On the other hand, while monolayer was being formed, paracellular permeability decreased over time until reaching a minimum on day 11 ($1.67 \times 10^{-7} \pm 2.96 \times 10^{-8}$ ml/cm^2*sec) and aferwards remained stable.

We analyzed whether pepsin-trypsin digested gliadin (PT-G) would have destructive effects on the formation of the epithelial monolayers and/or on those already formed. In the first case, PT-G (or PT-BSA as a negative control) was added during the whole duration of the experiment (Figure 13). PT-G was able to disrupt monolayer formation and to increase the paracellular permeability of C2BBe1 cells, and this effect was dependent of the PT-G concentration. When cells were stimulated with PT-G 500 µg/ml, they achieved the basal value indicative of monolayer formation at day 7 (803.22 ± 171.11 $\Omega \times$ cm^2) at the same time as in the PT-BSA and control, but then, values decreased and were maintained below the basal line until the end of the experiment (Figure 13, A). Similar results were obtained when the PT-G concentration was increased to 1 mg/ml (Figure 4b), cells achieved the basal value at day 7 (568.15 ±46.68 $\Omega \times$ cm^2) and then values remained around the basal line.
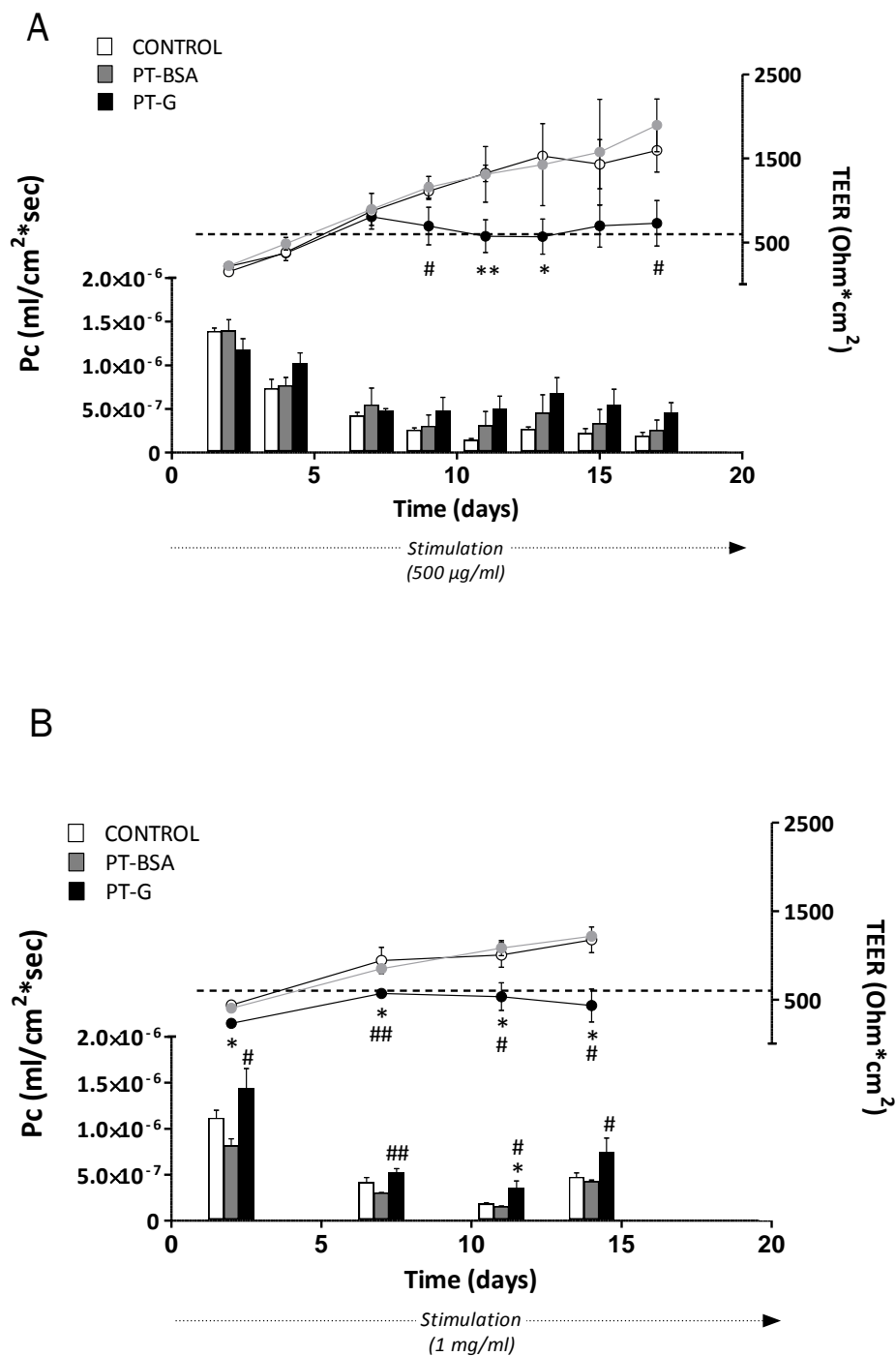
Figure 13.

Gliadin effects on monolayer formation and barrier permeability in intestinal C2BBe1 cells. TEER values are shown on the right Y axis and LY permeability coefficient (PC) on the left axis. Stimulation with PT-G and PT-BSA was done at different concentrations a) 500 µg/ml and b) 1mg/ml. The line at TEER=600 $\Omega \times cm^2$ indicates the basal value of formed monolayer and the discontinuous arrow indicates the stimulation time. Mean values and standard error of the mean (n=3) are shown. Differences to the control group are shown by asterisk and differences to the PT-BSA negative control group are shown by hash (* or [#] p-value 0.01 < 0.05; **or [##] p-value 0.001 < 0.01).

66
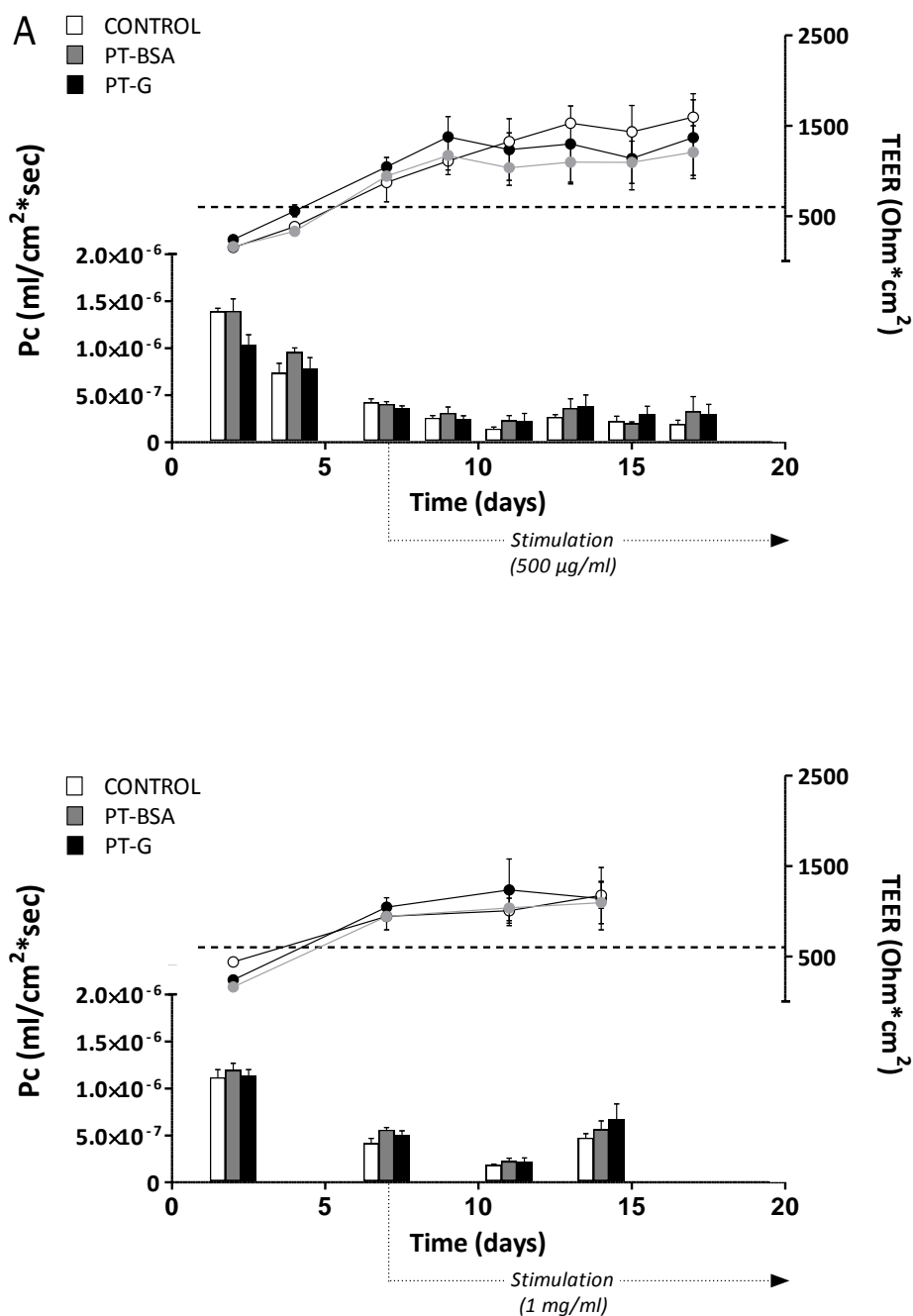
Figure 14. Gliadin effects on formed monolayers in intestinal C2BBe1 cells. TEER values are shown on the right Y axis and LY permeability coefficient (PC) on the left axis. Stimualtion with PT-G and PT-BSA was done at different concentrations a) 500 µg/ml and b) 1mg/ml. The line at TEER=600 $\Omega \times cm^2$ indicates the basal value of formed monolayer and the discontinous arrow indicates the stimulation time. Mean values and standard error of the mean (n=3) are shown. Differences to the control group are shown by asterisk and differences to the PT-BSA negative control group are shown by hash (* or [#] p-value 0.01 < 0.05; **or [##] p-value 0.001 < 0.01).

Regarding the paracellular permeability (measured as permeability coefficient - Pc), when cells were stimulated with PT-G 500 µg/ml, there were no differences compared to unstimulated and PT-BSA stimulated controls. However, a trend towards higher permeability was observed in PT-G stimulated cells (Figure 13, A). When the PT-G concentration was increased to 1 mg/ml, the effect in monolayer permeability was significantly higher (Figure 13, B). PT-G stimulated cells showed a differentially higher paracellular permeability during all the experiments compared to PT-BSA negative control and at day 11 compared to the control.

On the other hand, to analyze the effect of PT-G in formed monolayers, PT-G (and PT-BSA) were added once monolayers had been formed, at day 7 (Figure 14). In this case, disruption of the monolayer or the paracellular permeability were not observed. Cells stimulated with PT-G showed the same values as non-treated and PT-BSA treated controls. This result was obtained with both PT-G concentrations used.

2. Identification of TJ-related CD susceptibility genetic factors

2.1. Association analysis

A total of 361 subjects (270 CD and 91 Controls) from the Spanish population were genotyped and association studies performed for 8 TJ-related polymorphisms previously associated with CD in other populations: 3 SNPs in *MAGI2*, 2 SNPs in *PARD3* and 3 SNPs in *MYO9B* (Table 7). We detected an association between rs6962966 in the *MAGI2 locus* and CD risk. rs6962966 was in Hardy-Weinberg's equilibrium in the control group. The presence of the minor allele increased the risk to develop CD 1.88-fold. No other SNPs in *MAGI2, PARD3* or *MYO9B* were associated with CD.

2.2. Associated SNP characterization

2.2.1. Genomic context: characterization of *MAGI2* and RP4-587D13.2

In order to investigate the possible effects of the associated SNP on CD pathogenesis, we analyzed its genomic context. The variant rs6962966 is located in an intronic region of the *MAGI2* (a membrane-associated guanylate kinase) genomic locus, which includes 13 protein coding transcripts. This intronic region also harbors *RP4-587D13.2*, a long RNA transcript represented by a cluster of ESTs (expressed sequence tag) covering 721 bp located 3.9 kb upstream of rs6962966 (Figure 15, A). The function of this transcript is unknown, but based on its low coding potential; *RP4-587D13.2* is assigned to the non-coding RNA (ncRNA) class according to coding potential calculator program (cpc.cbi.pku.edu.cn/).

As there is no information about the possible effects of the rs6962966 variant on nearby genes, we proceeded to characterize the region. First of all, expression analysis of the nearby genes in different human tissues showed that the expression of *MAGI2* and *RP4-587D13.2* was positively correlated across different human tissues (Spearman r=0.933; p-value 0.0007). Expression was detected in all the tissues studied including intestine (the CD target tissue), although both genes were predominantly expressed in brain (Figure 15, B and C). Subcellular localization analysis of the *RP4-587D13.2* revealed that this transcript is predominantly nuclear (Figure 15, D).

Table 7. Association analyses of single nucleotide polymorphisms in TJ-related candidate genes in CD. Associated SNP is shown in bold.

| SNP	Position (Hg38)	Gene	Minor allele	Frequency in CD	Frequency in controls	p-value	Odds-ratio	(CI 95%)
rs6962966	**chr7: 78174806**	***MAGI2***	**A**	**0.521**	**0.3661**	**0.0029**	**1.88**	**(1.24-2.87)**
rs9640699	chr7: 78366115	*MAGI2*	A	0.3613	0.3056	0.1763	1.29	(0.89-1.85)
rs1496770	chr7: 78629694	*MAGI2*	T	0.4699	0.4341	0.4022	1.16	(0.82-1.62)
rs10763976	chr10: 34275364	*PARD3*	G	0.4713	0.456	0.7246	1.06	(0.76-1.49)
rs4379776	chr10: 34328092	*PARD3*	T	0.3811	0.4286	0.2583	0.82	(0.58-1.16)
rs2305767	chr19: 17183487	*MYO9B*	C	0.3547	0.3956	0.3311	0.85	(0.59-1.19)
rs1457092	chr19: 17193427	*MYO9B*	A	0.394	0.3889	0.9042	1.02	(0.72-1.45)
rs2305764	chr19: 17203024	*MYO9B*	A	0.4147	0.4333	0.671	0.93	(0.65-1.32)
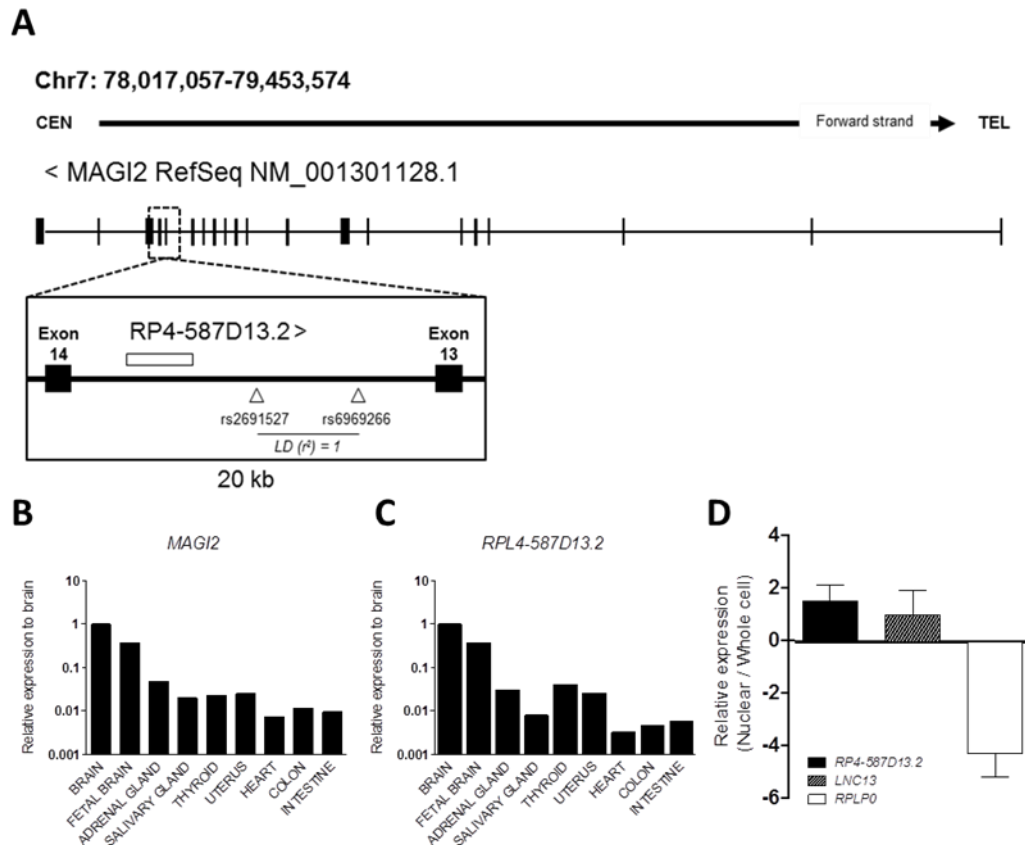
Figure 15. Characterization of the associated region. A) Exon-intron structure of MAGI2 gene and genomic annotations of a 20-kb region surrounding SNP rs6962966. B) *MAGI2* **and C)** *RPL4-587D13.2* **lncRNA expression profiles in 9 human tissues. Values were represented relative to the tissue with highest expression levels (brain). D)** *RPL4-587D13.2* **subcellular localization. Values are represented as relative amount of nuclear transcript compared to whole cell amount.** *LNC13* **and** *RPLP0* **genes are used as nuclear and cytoplasmic controls respectively. The graph shows the means and the standard error of the mean of three independent experiments.**

2.2.2. *MAGI2* and *RP4-587D13.2* expression in CD

In order to determine whether the genes located in the genomic context of associated rs6962966 are altered in CD, we analyzed the expression at the mRNA level of both *MAGI2* and *RP4-587D13.2* in active (n=15) and treated (n=15) CD patients and healthy controls (n=15). Both *MAGI2* and *RP4-587D13.2* were expressed at lower levels in the duodenum of active CD patients compared to non-CD controls (Figure 16, A and B). When patients on gluten free diet (GFD) were compared to controls, both genes remained significantly underexpressed. The correlation between *MAGI2* and RP4-

71

587D13.2 previously observed in tissues (Figure 15) was also observed in duodenal biopsies (Figure 16, C).
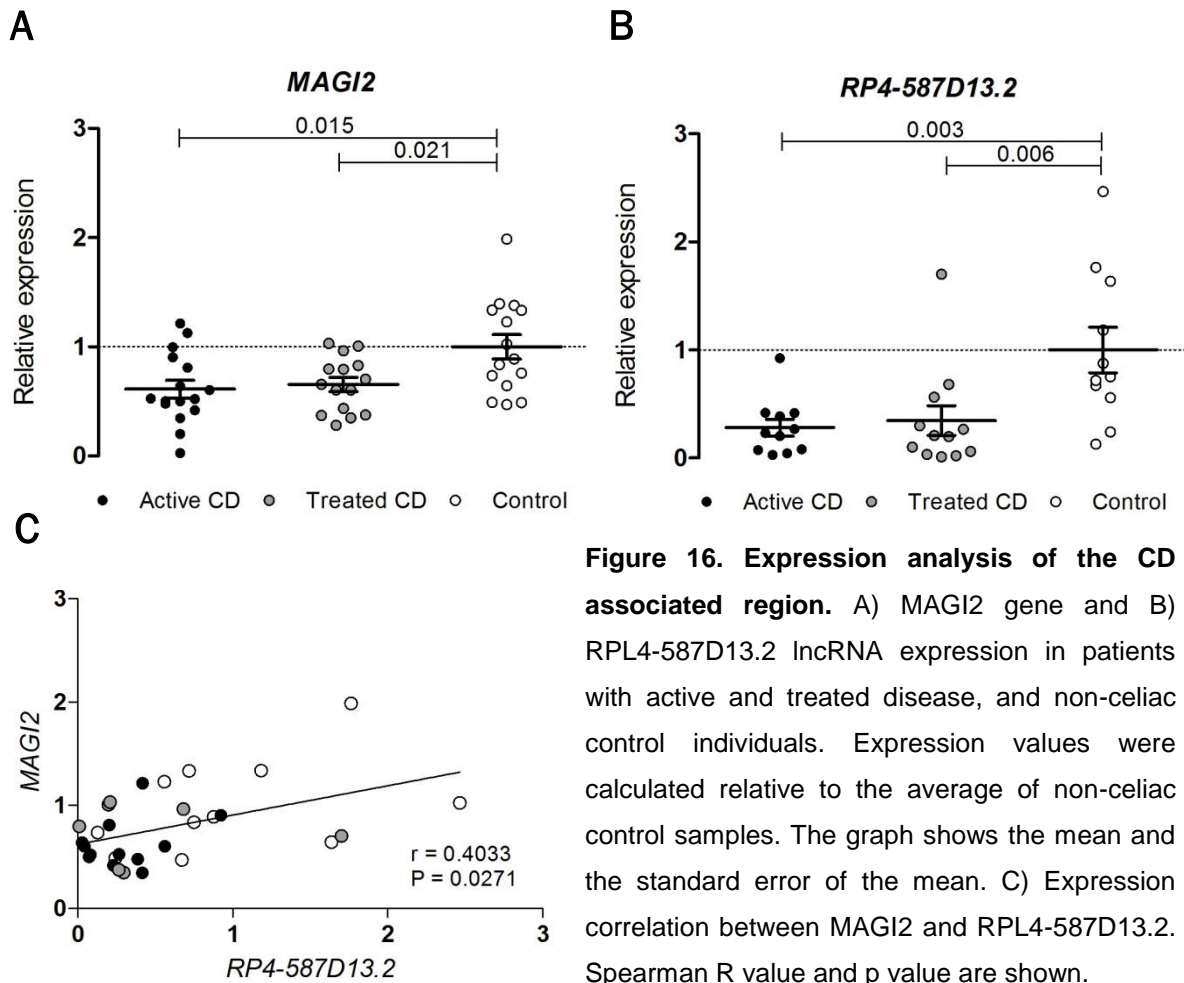
Figure 16. Expression analysis of the CD associated region. A) MAGI2 gene and B) RPL4-587D13.2 lncRNA expression in patients with active and treated disease, and non-celiac control individuals. Expression values were calculated relative to the average of non-celiac control samples. The graph shows the mean and the standard error of the mean. C) Expression correlation between MAGI2 and RPL4-587D13.2. Spearman R value and p value are shown.

As the expression levels of *MAGI2* and *RPL4-587D13.2* were altered in disease, an expression quantitative trait *loci* (eQTL) analysis was performed to test whether the associated SNP is affecting the expression of the nearby genes. No significant association was found in duodenal biopsies of CD patients or controls (Figure 17).

Figure 17. eQTL analysis for A) *MAGI2* and B) *RP4-587D13.2* in duodenal biopsies from patients with CD (with active or treated disease) and non-CD controls. Scatter plots show relative expression to control mean depending on the rs2691527 genotype (presence or absence of minor allele T).

2.2.3. Effects of *MAGI2* silencing and gliadin stimulation
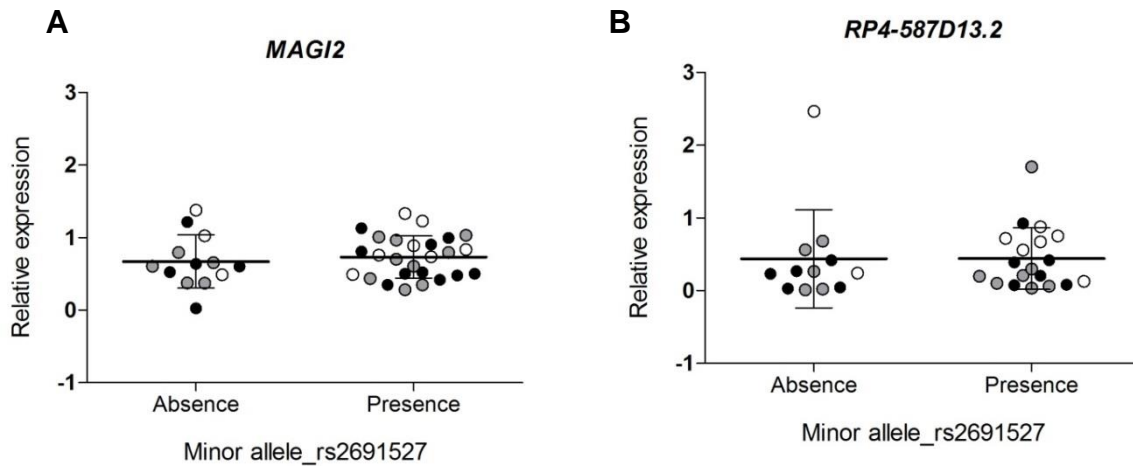
To better understand the functional significance of the downregulation of *MAGI2* in the TJ pathway, a simplified celiac intestinal model was created by knocking down the *MAGI2* gene in C2BBe1 intestinal cells. After silencing, cells were stimulated with gliadin (PT-G) and the expression of a panel of 22 TJ-related genes was analyzed. The effects of PT-G stimulation alone and *MAGI2* silencing alone were also analyzed.

The efficiency of the siRNA mediated *MAGI2* knockdown was higher than %75 (Figure 18, A). While PT-G stimulation alone had little impact on TJ-related genes, only *F11R* and *TICAM2* genes were altered, *MAGI2* silencing provoked alterations of *F11R, PARD6A, PPP2R3A, TICAM2* and *ZAK* expression (Figure 18, B). Interestingly, those alterations were more pronounced when silencing was combined with PT-G stimulation, showing enhanced effects on *ZAK, PARD6A* and *TICAM2*, and provoking additional alterations in *GNAI1, CLDN2* and *ACTB*.
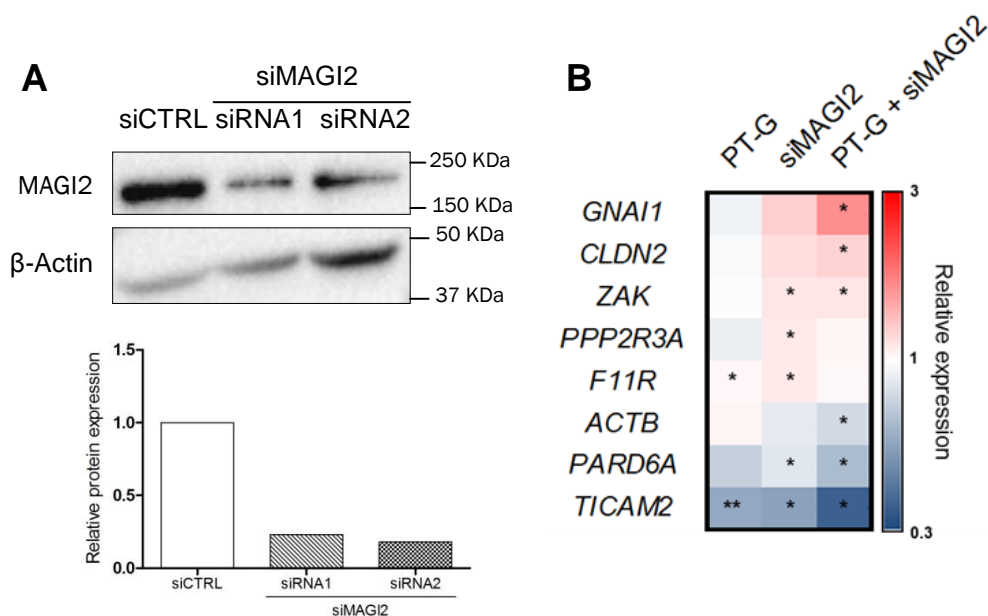
Figure 18. **Effects of *MAGI2* silencing and gliadin stimulation in TJ related genes expression A) Western Blot and the relative protein expression of siRNA-mediated measured with MAGI2 knockdown in C2BBe1 cells. B) Heatmap of candidate genes that presented alterations under different conditions. Expression was calculated relative to the average of control samples (silenced with siCTRL and stimulated with PT-BSA). Each column represents treatment category and each line represents a candidate gene. The color scale on the right shows the relative expression to the control average for each sample and gene. Statistically significant differences to the control group are shown by asterisk (* p value < 0.05; ** p value < 0.01). siCTRL= non-targeting siRNA control; siMAGI2= *MAGI2* targeting siRNA; PT-G= gliadin digested with pepsin and trypsin.**

3. Gene editing of constitutively altered TJ genes

To elucidate the implication that constitutively downregulated genes identified in chapter 1 (*INADL*, *PPP2R3A* and *TJP1*) could have in the pathogenesis of CD, we performed functional studies in mammalian cells using CRISPR/Cas9 gene editing technology.

3.1. Validation of the deletion

Deletions were assessed by PCR with primers flanking the deleted region in epithelial cell lines (C2BBe1, HCT116 and HEK293FT) that had been transfected with Cas9 nuclease and specific sgRNAs (Figure 19). Deletion of *INADL* was obtained in C2BBe1, HCT116 and HEK293FT cells, whereas deletion of *PPP2R3A* was only detected in HEK293FT cells. In the case of *TJP1* gene, deletion was not found in any of the cell lines analyzed.
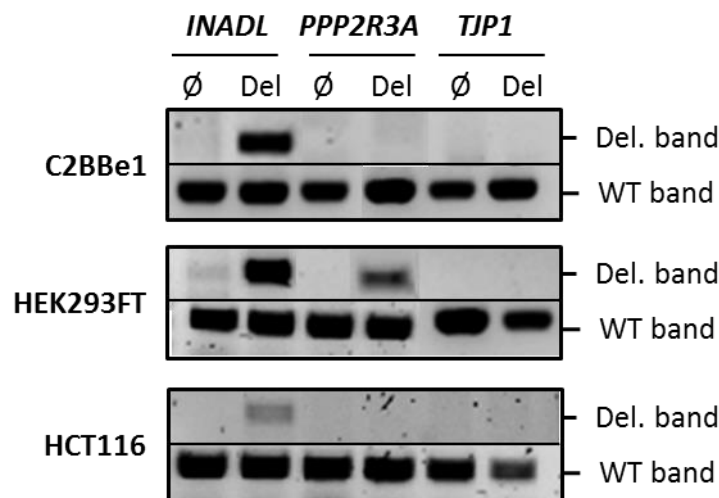
Figure 19. Deletion of exon 1 of *INADL*, *PPP2R3A* and *TJP1* genes in human cell lines using CRISPR/Cas9 gene editing technology. For each studied human cell line, non-edited lines (Ø) and edited lines (Del) for *INADL*, *PPP2R3A* and *TJP1* genes are shown. Primers outside expected deletion regions were used to determine the targeted deletion and primers inside the deletion were used to detect wild type (WT) clones as a control. Del. band= deletion band; WT band= Wild type band.

As genome editing is not 100% efficient (WT band present in all the cases), we isolated single cells containing the edition for single colony expansion. PCR was performed to determine the presence of the deletion in the clonal lines (Figure 11). We were able to subclone the HEK293FT cells getting 3 clones for the *INADL* deletion and 5 clones for *PPP2R3A* deletion. All edited clonal cell lines were heterozygous for the deletion, showing both the deletion band and WT band. We had difficulties in subcloning the intestinal cell lines and did not get any clonal cell line.
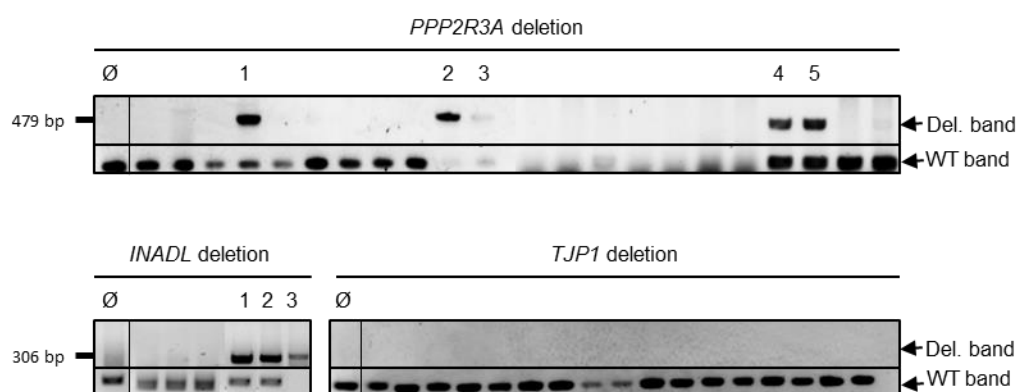
Figure 20. Targeted deletion of human *INADL* and *PPP2R3A* with CRISPR/Cas9 in subcloned HEK293FT cell lines. Clonal cell lines harboring the targeted deletion were identified by PCR (numbered in the graph). Primers outside expected deletion regions were used to determine the targeted deletion (Del.) and primers inside the deletion were used to detect wild type (WT) clones as a control.

Expression analysis confirmed differences in edited cells comparing to wild type clonal cell lines transfected with an empty vector. Both genes, *INADL* and *PPP2R3A* were significantly silenced in HEK293FT edited clonal cell lines (Figure 12).
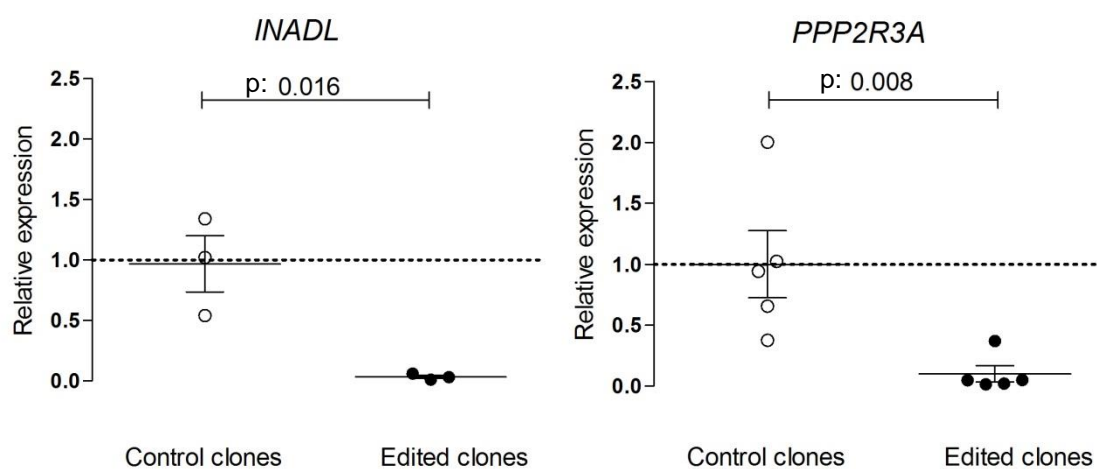
Figure 21. Expression analysis of HEK293FT clonal cell lines with CRISPR/Cas9 mediated *INADL* and *PPP2R3A* mutations. Expression values were calculated relative to the average of empty vector transfected control clonal cells. The graph shows the means, the standard error of the mean and p-value.

3.2. Cell functional studies

We tested the functional effects of the deletions on the HEK293FT clonal cell lines. As *INADL* gene function is related to cell polarization and proliferation, we investigated the impact of *INADL* deletion on cell proliferation rate and cellular migration on the generated clonal cell lines.

A significant increase in cell proliferation rate was observed in *INADL* deleted cells compared to controls (Figure 22), especially at the exponential phase of

proliferation (3rd and 4th days), and was less pronounced when *INADL* edited cells reached confluency.
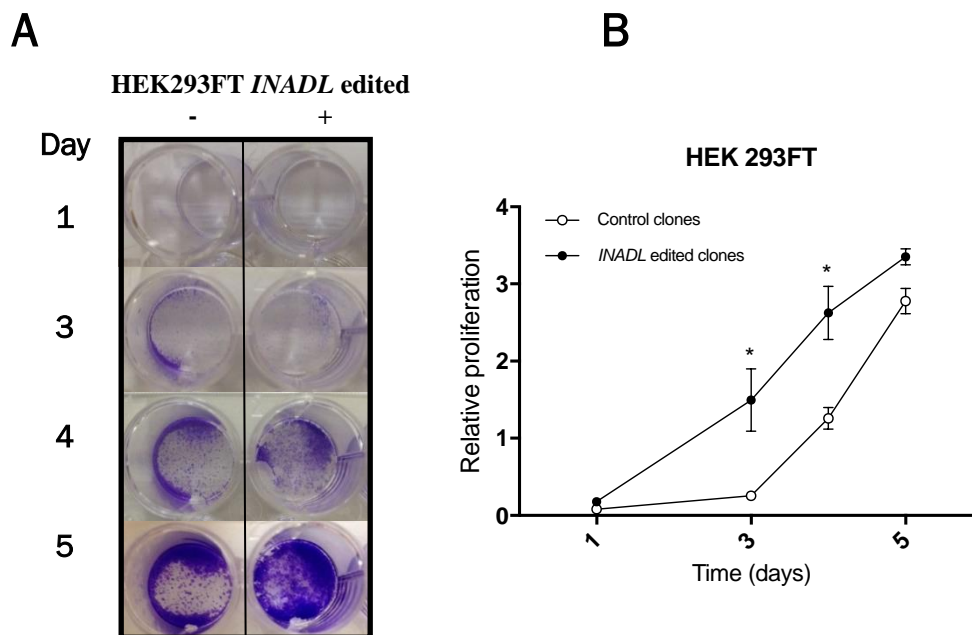
Figure 22. Proliferation assay in *INADL* deleted HEK293FT clonal cell lines. A) Time-course cell proliferation assay of *INADL* deleted and control non-deleted cells plated in a 24-well cell culture dish, determined by crystal violet staining for live cells. A Representative experiment of three independent replicates performed. B) Relative proliferation rate of HEK293FT cells. All assays were performed in triplicate, and the proliferation of CRISPR/Cas9 edited cells is compared to that of unedited cells. Values are shown as the mean and standard error of the mean (n=3).

On the other hand, a wound healing assay showed a decreased migration ability of *INADL* edited HEK293FT cells compared to controls (empty vector) (Figure 23). Cells with *INADL* mutation seemed to need a longer time to fill the scratched area although differences could not be statistically assessed (Figure 23). In the representative phase contrast photographs, a differential timing for healing could be observed. While control sample wound was completely filled 21 hours after wound generation, *INADL* edited cells took a longer time to fill in the wound area, indicating a defect in migration.
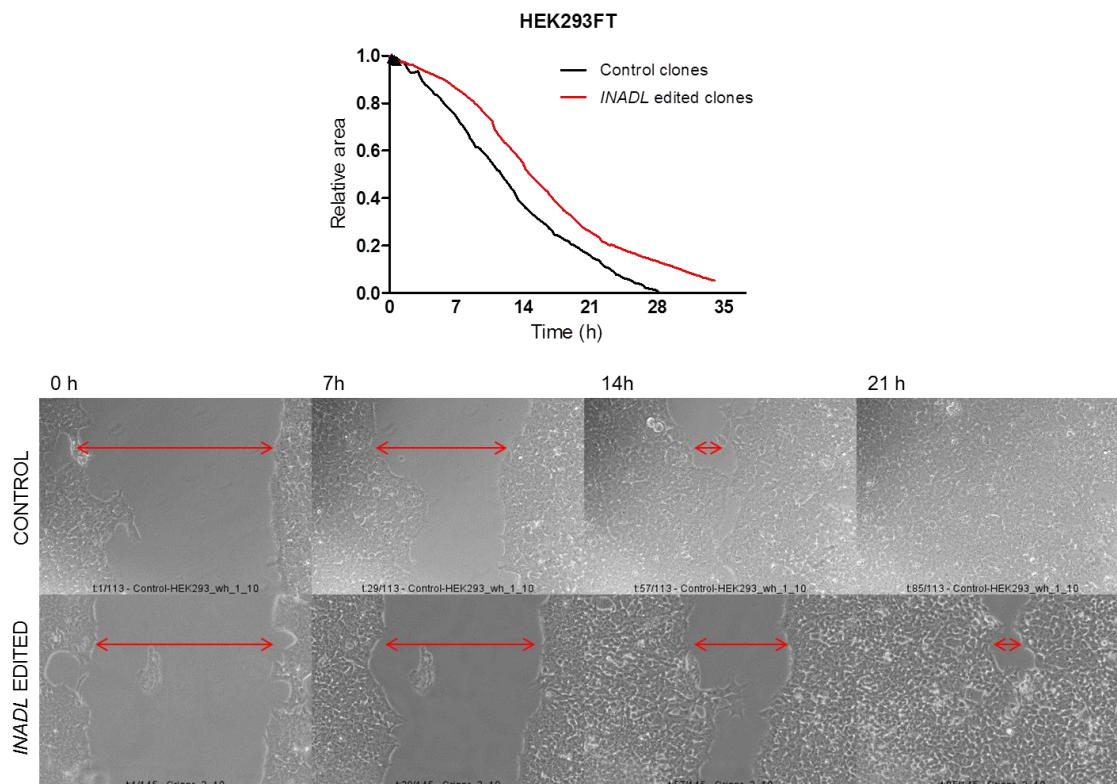
Figure 23. Wound healing assay of *INADL* edited HEK2293FT clones A) Graph shows the quantification of the wounded area. Cells were monitored during 36 hours and images were captured every 15 minutes. The migration was determined by the rate of cells filling the scratched area and the result is represented as mean of three independent areas of the same wound healing assay. b) Phase contrast photographs of the cultures taken at 0 h (immediately after scratching) and at the indicated time intervals show the wound closure by cells with *INADL* deletion (upper panels) and unedited controls (bottom panels). A representative experiment of the wound healing assay is shown. The red line points the length of the scratched area at the different time points.

5. Discussion

DISCUSSION

Identification and characterization of biologically active or perturbed pathways is important to understand the changes that characterize different cellular states such as disease. Gene expression profiling is an excellent approach to create a global picture of cellular function. In the present project, 15 out of the 23 TJ-related genes analyzed were seen to be differentially expressed in active CD intestinal biopsies, highlighting a global deregulation of the pathway as a consequence of gluten-induced inflammation. Besides, correlation studies suggest that the alterations observed are also influencing the interaction among the genes in this pathway.

Several of the alterations observed in our analysis have been previously shown to contribute to the epithelial barrier impairment in CD and other associated intestinal disorders such as IBD, intestinal infections or food allergy, highlighting the role of TJ related gene expression alterations in the pathogenesis of the gastrointestinal tract [103, 117]. The upregulation of cytoskeletal components observed is probably related to the gliadin-induced cytoskeleton reorganization that alters TJ protein distribution and promotes TJ disruption [118, 119]. *In vitro* and *in vivo* animal studies have demonstrated that inflammatory cytokines (such as IFNγ and TFNα) act through the phosphorylation of cytoskeleton, which together with the downregulation of adaptor protein ZO-1 (encoded by *TJP1*) and transmembrane proteins such as occluding and F11R provokes a destabilization of the TJ complexes and an increase permeability to macromolecules [120, 121]. The augmented permeability is also due to the overexpression of pore-forming claudins, especially claudin-2 (*CLDN2*), that severely affects the barrier integrity in disease [104, 105]. *CLND2* overexpression depends on disease status, being higher in severe atrophy. We observed the disease characteristic overexpression of *CLDN1* and CLDN2 in our study. Moreover, we found that these two genes coexpress with other gene in the pathway only in active disease. So, our results present the upregulation of epithelial transcytosis represented by claudins as an additional cause of

paracellular leakage, together with the complete deregulation and disorganization of the TJ pathway.

We also found other alterations that had not been previously reported related to polarity (*PARD6A* and *INADL*) , cell proliferation (*PPP2R3A*) and other signaling molecules (*CSDA* and *GNAI1*). This reflects a global perturbation of the pathway that possibly leads to the well-known disrupted TJ strand alterations in CD. Those alterations could have a relevant impact on the function of the pathway as a correct polarization of the cell is essential for the tight junction formation [122]. Further studies would be necessary to elucidate the functional implication of each of the altered genes.

After removal of gluten from the diet, the expression of all of the genes except *INADL, PPP2R3A* and *TJP1* reverted to levels comparable to non-CD tissue, suggesting that most of the altered genes are functionally involved in the pathogenesis of the disease, but not in the genetic susceptibility. This is in concordance with the hypothesis that gliadin induces zonulin-pathway mediated disassembly of intercellular TJ, leading to a rapid intestinal permeability increase [108, 123] which would allow paracellular gliadin uptake and an enhanced inflammatory reaction. However, the mechanisms underlying the loss of intestinal barrier function that allow gliadin peptides to traffic through the gut are largely unknown.

Although it has been proven that gliadin peptides induce epithelial barrier disruption, there is growing evidence that the leaky gut is both a cause (or an enhancer) and a consequence of the disease. The lack of an animal model reproducing all features of CD has limited the study of the epithelial barrier in the early stages of disease. However, the use of *in vitro* approaches including a variety of cell line models and organ cultures of the celiac patient small-bowel mucosal biopsies has enabled to generate some knowledge on this field [124]. In these culture systems, gluten induces different effects on gene expression that can be quantified and the ability of epithelial cells to form monolayers through TJ unions permits studies concerning permeability. For example, Rauhavirta et al. found that the expression of some intestinal TJ proteins was already altered in early developing CD, when the mucosal architecture is still

normal [125]. We used C2BBe1 intestinal cell monolayers as a model to understand the increase of epithelial permeability characteristic of the disease. Using this model, we confirmed that digested gliadin inhibits the formation of TJ structures but does not affect the integrity of the already formed epithelial monolayers (regardless the gliadin concentration). It seems that gliadin stimulation leads to a disruption of the C2BBe1 monolayer only when the epithelial barrier integrity is compromised.

If we accept our experiment as a valid model to study what happens at the epithelial monolayer in early-stages of CD, we could conclude that an impaired epithelial integrity in the duodenal mucosa before the onset of disease may allow gliadin peptides to affect the formation of those structures through paracellular flux of gluten, triggering an inflammatory response and resulting in TJ dysfunction. Our findings support the idea that compromised epithelial barrier function is probably an early event before the gluten-driven pathogenesis of CD leading to mucosal damage in CD patients. *In vivo* intestinal permeability studies will help to solve the functional significance of our findings.

Apart from the augmented permeability, polymorphisms of TJ-related genes *MAGI2*, *PARD3* and *MYO9B* have been previously associated to risk of gastrointestinal disorders like CD and IBD pointing to a genetic susceptibility to the intestinal barrier impairment. Attempts to replicate those associations in different cohorts have sometimes failed, highlighting the heterogeneity and complexity of common human disorders [126, 127]. The strength of the associated signals is relatively small, and the ability to detect significant association often depends on minute allele frequency differences across populations, which could account, at least in part, for some of the negative results. In our cohort, we have been able to replicate the association with rs6969266 (an intronic variant in *MAGI2*) previously identified in Dutch and British cohorts [107].

In order to investigate the possible implication of this intronic variant in CD pathogenesis, we scrutinized the genomic region around rs6969266 and found that, apart from *MAGI2,* there is also an intronic lncRNA of unknown biological

function (*RP4-587D13.2*). Recent studies have evidenced that intronic regions are key sources of regulatory ncRNAs and several long regulatory intergenic ncRNAs have been functionally characterized [128] discovering their ability to regulate gene expression using diverse mechanisms. The widespread expression of RP4-587D13.2, together with nuclear localization and downregulation in CD, gives clues on its functionality. Interestingly, the concordance in the expression profile of the intronic transcript and *MAGI2*, across the human tissues and in celiac disease, indicates that *RP4-587D13.2* could be processed from the same pre-mRNA and could be functionally related to *MAGI2*, as has been previously evidenced for other intronic lncRNAs [129].

Not only do *MAGI2* and *RP4-587D13.2* seem to be related but their expression is also silenced in the duodenum of CD patients. Moreover, this reduced expression persists in the CD gut even after more than two years on a GFD and an apparent recovery from intestinal atrophy, suggesting a genetic background of this alteration. The altered expression together with the presence of an associated variant, points to a dysfunction that is present before the clinical onset of the disease, as it has been suggested previously [130]. Although the associated variant does not seem to have any direct effect on the reduction of transcription, evidence suggests that the region is functionally altered and genetically implicated in disease, being an interesting region for functional characterization.

To further address the biological function of the associated region, we mimicked the CD scenario, repressing *MAGI2* translation and stimulating intestinal cells with PT-G. The altered expression of several TJ-related genes after *MAGI2* silencing, suggests that *MAGI2* gene is necessary for a correct function of the pathway, and that the decrease in its expression observed in the disease could be affecting the whole network. Interestingly, in the absence of *MAGI2* expression, gliadin stimulation augmented the effects observed in some of the genes. This supports that the alteration of the TJ pathway has a role in the development of the disease, as its dysfunction triggers an increased response to ingested gluten that probably leads to an enhanced intestinal permeability.

As commented above, the genetic association with 2 regulators of epithelial polarity, Par-3 and Magi2, suggest that epithelial polarity may play a role in the TJ disruption and the consequent CD barrier defect being one of the principal causes of the alterations in permeability present in CD [107]. Additionally, in the present study, we found for the first time that 3 genes are constitutively downregulated pointing also to a possible genetic effect: 1) *INADL*, encoding the PATJ component of a CRB-PALS1-PATJ polarity complex, 2) *PPP2R3A*, encoding a component of a negative cell cycle regulator complex and *TJP1*, encoding ZO-1, the main scaffold protein in TJ pathway.

Taking into account the pivotal role of these genes in the TJ pathway, we believed that the intrinsic downregulation could have a strong impact in the early impairment of the epithelial barrier in CD. This insight raises the question of how mechanisms of cell polarity might be contributing to TJ integrity. In order to answer those questions, as siRNA technology doesn't resemble the constitutive downregulation of these genes, we used CRISPR/Cas9 gene editing technology in order to obtain precise and permanent downregulation of these genes in epithelial cell lines. This technique enables long-term studies in the edited cell lines which allow studying phenotypes. Some studies using CRISPR/CAS9 to study in TJ-related genes started to be done [131, 132] and this work is a major advance in the field, generating clonal mutant HEK293FT cell lines for the genes *INADL* and *PPP2R3A*.

Evolutionarily conserved cell polarity complexes Par3-Par6-aPKC and PATJ-PALS-CRB regulate tight junction formation and directional migration in epithelial cells but the mechanism by which these polarity proteins assemble at the leading edge of migrating epithelial cells remains unclear [133, 134]. PATJ, encoded by *INADL,* plays a central role in tight junction biogenesis and epithelial polarity by providing a link between the lateral and apical components [135].

Here, we show that the protein PATJ is required for directional migration by using a wound-induced migration assay. Directed cell migration plays a key role in wound healing and immune system function as cells have to determine

the correct migration direction to the wound area and form protrusions to move. During wound healing, the apical-basal polarity of epithelial cells is disrupted and polarity protein complexes are redistributed [136]. Dan du et al. found that during migration, TJ protein occludin recruits polarity proteins aPKC-Par3 and PATJ to the leading edge, in order to stabilize polarized migration in response to external stimuli. Knockdown of occludin blocks wound closure, disrupting the accumulation of aPKC-Par3 and PATJ at the leading edge, and results in a disorganized microtubule network and reduced cell protrusion [122]. PATJ recruitment could be via association with aPKC as a complex or via ZO-3-occludin because ZO-3 was reported to interact with both of them [137]. Although the mechanism is still unknown, our data indicate that PATJ recruitment is essential for polarized epithelial migration as its knockdown reduces cell migration capacity.

Taking all into account, we believe that PATJ could be regulating the migration of epithelial cells by assisting in the recruitment of aPKC and Par3 to the leading edge and controlling the microtubule and MTOC reorientation, which are crucial for directional migration. This idea is supported by an experiment where the orientation of the microtubule-organizing center was disrupted in kidney cells in wich PATJ was silenced [10]. Thus, PATJ depletion probably affects the interaction might promote or sustain the polarity complex at the leading edge, resulting in an incorrect microtubule elongation and the consequent reduction in migration.

Interestingly, the *INADL*-mutated HEK293FT cells showed higher proliferative activity, suggesting that changes in the apico-basal polarity can have an acute influence on the control of cell proliferation. Apico-basal polarity and cell-cell adhesion are tightly interconnected and the mechanism by which loss of cell polarity in epithelial affects cell proliferation has been a topic of discussion. It is known that disturbance of the epithelium, such as wounding, can stimulate proliferative activity, whereas epithelial cell-cell communication regulates suppression of proliferation when the final organ size has been achieved [138, 139]. In CD, proliferation of crypt enterocytes is an early alteration of the intestinal mucosa induced by gluten in susceptible individuals [140]. In fact, the

celiac intestine is characterized by an inversion of the differentiation/proliferation program of the tissue, with a progressive reduction in the differentiated compartment that reaches complete atrophy of the villi, and an increase of the proliferative compartment, with crypt hyperplasia [77, 141] .

Different microarray studies in CD duodenal biopsies as well as three-dimensional epithelial cell differentiation models have revealed that genes with a role in proliferation and differentiation pathways take part in CD pathogenesis, including *PPP2R3A* [77, 141].

Taking our results into account, we speculate that the sustained downregulation of *INADL* could be implicated in the alteration of proliferation that is seen in the disease and have an important role in disease pathogenesis, affecting cell migration and proliferative functions.

In general conclusion, we have demonstrated that TJ pathway is overall altered in CD, affecting both fence and gate functions of these structures. On the one hand, we saw that TJ barrier disruption is probably an early event in the pathogenesis, being the epithelial layer damaged before the onset of the disease. We also discovered that gliadin peptides have a toxic effect in impaired epithelial monolayers enhancing the permeability of the layer which could be one of the causes for the observed intestinal permeability in CD patients. On the other hand, we found that some TJ-related genes, most of them related to cell polarity, could be implicated in the susceptibility to CD, suggesting a key role in disease development. Finally, we created an epithelial cells with stable modifications that are a good model for functional studies of constitutive candidate genes.

6. Conclusions

CONCLUSIONS

I. There is an overall disruption of the expression and coexpression profiles of the TJ pathway genes in CD.

    a. The majority of gene expression alterations observed in active CD tissue seems a consequence of the disease, since they revert to normal after GFD. However, the constitutive downregulation of TJP1, INADL and PPP2R3A suggest a primary event of genetic origin.

    b. Intestinal cell monolayers are a useful model to study early events in epithelial barrier integrity. The ability of gliadin to disrupt the formation of TJ structures supports the idea of barrier defect as an early event that could precede the autoimmune response.

II. TJ pathway regions harbor CD associated polymorphisms but their functional characterization is still unclear stressing the complex functional relationship between candidate genes and associated SNPs.

    a. This study confirms the association of rs6969266 with CD in a Spanish population. The rest of the polymorphisms could not be replicated.

    b. The rs6969266 region contains *MAGI2* and RP4-XX lncRNA which are express in a coordinated manner in several human tissues, including duodenum mucosa. Both are downregulated in CD patients suggesting their involvement in the disease. However, rs6969266 genotype does not seem to have a direct effect on their expression.

III. We have developed stable HEK293FT cell lines harboring constitutive deletions in CD related TJ genes *INADL* and *PPP2R3A.* The former shows increased proliferation and decreased migration resembling what has been described in the celiac epithelium. CRISPR-Cas9 based technology is an efficient tool for the detailed functional analysis of genes involved in complex diseases, and will be useful to clarify the physiopathology of those diseases.

7. Bibliography

1. Dowd, B. and J. Walker-Smith, *Samuel Gee, Aretaeus, and the coeliac affection.* Br Med J, 1974. **2**(5909): p. 45-7.
2. K, W.D.W., et al., *Coeliakie.* 1950, MD Thesis.
3. Feighery, C., *Fortnightly review: coeliac disease.* BMJ, 1999. **319**(7204): p. 236-9.
4. Mäki, M. and P. Collin, *Coeliac disease.* Lancet, 1997. **349**(9067): p. 1755-9.
5. Marsh, M.N., *Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue').* Gastroenterology, 1992. **102**(1): p. 330-54.
6. Stenman, S.M., et al., *Secretion of celiac disease autoantibodies after in vitro gliadin challenge is dependent on small-bowel mucosal transglutaminase 2-specific IgA deposits.* BMC Immunol, 2008. **9**: p. 6.
7. Mäki, M., *The humoral immune system in coeliac disease.* Baillieres Clin Gastroenterol, 1995. **9**(2): p. 231-49.
8. Husby, S., et al., *European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(1): p. 136-60.
9. Catassi, C., *[The global village of celiac disease].* Recenti Prog Med, 2001. **92**(7-8): p. 446-50.
10. Rostom, A., J.A. Murray, and M.F. Kagnoff, *American Gastroenterological Association (AGA) Institute technical review on the diagnosis and management of celiac disease.* Gastroenterology, 2006. **131**(6): p. 1981-2002.
11. Dubé, C., et al., *The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review.* Gastroenterology, 2005. **128**(4 Suppl 1): p. S57-67.
12. Mustalahti, K., et al., *The prevalence of celiac disease in Europe: results of a centralized, international mass screening project.* Ann Med, 2010. **42**(8): p. 587-95.
13. Cataldo, F. and G. Montalto, *Celiac disease in the developing countries: a new and challenging public health problem.* World J Gastroenterol, 2007. **13**(15): p. 2153-9.
14. Murray, J.A., et al., *Trends in the identification and clinical features of celiac disease in a North American community, 1950-2001.* Clin Gastroenterol Hepatol, 2003. **1**(1): p. 19-27.
15. Dalgic, B., et al., *Prevalence of celiac disease in healthy Turkish school children.* Am J Gastroenterol, 2011. **106**(8): p. 1512-7.
16. Collin, P., et al., *High incidence and prevalence of adult coeliac disease. Augmented diagnostic approach.* Scand J Gastroenterol, 1997. **32**(11): p. 1129-33.
17. Bodé, S. and E. Gudmand-Høyer, *Incidence and prevalence of adult coeliac disease within a defined geographic area in Denmark.* Scand J Gastroenterol, 1996. **31**(7): p. 694-9.
18. Manzel, A., et al., *Role of "Western diet" in inflammatory autoimmune diseases.* Curr Allergy Asthma Rep, 2014. **14**(1): p. 404.
19. Ivarsson, A., et al., *Epidemic of coeliac disease in Swedish children.* Acta Paediatr, 2000. **89**(2): p. 165-71.

20. Plot, L. and H. Amital, *Infectious associations of Celiac disease.* Autoimmun Rev, 2009. **8**(4): p. 316-9.

21. Thomas, H.J., et al., *Contribution of histological, serological, and genetic factors to the clinical heterogeneity of adult-onset coeliac disease.* Scand J Gastroenterol, 2009. **44**(9): p. 1076-83.

22. Di Sabatino, A. and G.R. Corazza, *Coeliac disease.* Lancet, 2009. **373**(9673): p. 1480-93.

23. Murray, J.A., et al., *Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease.* Am J Clin Nutr, 2004. **79**(4): p. 669-73.

24. Lerner, A., *New therapeutic strategies for celiac disease.* Autoimmun Rev, 2010. **9**(3): p. 144-7.

25. Samasca, G., et al., *Gluten-free diet and quality of life in celiac disease.* Gastroenterol Hepatol Bed Bench, 2014. **7**(3): p. 139-43.

26. Rubio-Tapia, A. and J.A. Murray, *Classification and management of refractory coeliac disease.* Gut, 2010. **59**(4): p. 547-57.

27. Freeman, H.J., *Non-dietary forms of treatment for adult celiac disease.* World J Gastrointest Pharmacol Ther, 2013. **4**(4): p. 108-12.

28. Goel, G., et al., *Epitope-specific immunotherapy targeting CD4-positive T cells in coeliac disease: two randomised, double-blind, placebo-controlled phase 1 studies.* Lancet Gastroenterol Hepatol, 2017. **2**(7): p. 479-493.

29. Sollid, L.M. and E. Thorsby, *HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis.* Gastroenterology, 1993. **105**(3): p. 910-22.

30. Greco, L., et al., *The first large population based twin study of coeliac disease.* Gut, 2002. **50**(5): p. 624-8.

31. Gutierrez-Achury, J., et al., *Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease.* Nat Genet, 2015. **47**(6): p. 577-8.

32. Horton, R., et al., *Gene map of the extended human MHC.* Nat Rev Genet, 2004. **5**(12): p. 889-99.

33. Ludwig, H., et al., *[Association of HL-A1 and HL-A8 with childhood celiac disease].* Z Immunitatsforsch Exp Klin Immunol, 1973. **146**(2): p. 158-67.

34. Sollid, L.M., et al., *Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer.* J Exp Med, 1989. **169**(1): p. 345-50.

35. Spurkland, A., et al., *HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7.* Hum Immunol, 1992. **35**(3): p. 188-92.

36. Karell, K., et al., *HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease.* Hum Immunol, 2003. **64**(4): p. 469-77.

37. Abadie, V., et al., *Integration of genetic and immunological insights into a model of celiac disease pathogenesis.* Annu Rev Immunol, 2011. **29**: p. 493-525.

38. van Belzen, M.J., et al., *Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients.* Genes Immun, 2004. **5**(3): p. 215-20.

39. Ploski, R., et al., *On the HLA-DQ(alpha 1*0501, beta 1*0201)-associated susceptibility in celiac disease: a possible gene dosage effect of DQB1*0201.* Tissue Antigens, 1993. **41**(4): p. 173-7.

40. Lundin, K.E., et al., *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, 1993. **178**(1): p. 187-96.

41. Sollid, L.M., *Coeliac disease: dissecting a complex inflammatory disorder.* Nat Rev Immunol, 2002. **2**(9): p. 647-55.

42. Gujral, N., H.J. Freeman, and A.B. Thomson, *Celiac disease: prevalence, diagnosis, pathogenesis and treatment.* World J Gastroenterol, 2012. **18**(42): p. 6036-59.

43. van Heel, D.A., et al., *A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21.* Nat.Genet., 2007. **39**(7): p. 827-829.

44. Dubois, P.C., et al., *Multiple common variants for celiac disease influencing immune gene expression.* Nat.Genet., 2010. **42**(4): p. 295-302.

45. Trynka, G., et al., *Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.* Nat.Genet., 2011. **43**(12): p. 1193-1201.

46. Hunt, K.A., et al., *Negligible impact of rare autoimmune-locus coding-region variants on missing heritability.* Nature, 2013. **498**(7453): p. 232-5.

47. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

48. Kumar, V., et al., *Human disease-associated genetic variation impacts large intergenic non-coding RNA expression.* PLoS Genet, 2013. **9**(1): p. e1003201.

49. Castellanos-Rubio, A., et al., *A long noncoding RNA associated with susceptibility to celiac disease.* Science, 2016. **352**(6281): p. 91-5.

50. Farh, K.K., et al., *Genetic and epigenetic fine mapping of causal autoimmune disease variants.* Nature, 2015. **518**(7539): p. 337-43.

51. Schuppan, D., Y. Junker, and D. Barisani, *Celiac disease: from pathogenesis to novel therapies.* Gastroenterology, 2009. **137**(6): p. 1912-33.

52. Maiuri, L., et al., *Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease.* Lancet, 2003. **362**(9377): p. 30-7.

53. Hüe, S., et al., *A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease.* Immunity, 2004. **21**(3): p. 367-77.

54. Castellanos-Rubio, A., et al., *Long-term and acute effects of gliadin on small intestine of patients on potentially pathogenic networks in celiac disease.* Autoimmunity, 2010. **43**(2): p. 131-139.

55. Jabri, B. and L.M. Sollid, *Tissue-mediated control of immunopathology in coeliac disease.* Nat.Rev.Immunol., 2009. **9**(12): p. 858-870.

56. Arentz-Hansen, H., et al., *The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase.* J Exp Med, 2000. **191**(4): p. 603-12.

57. Folk, J.E. and P.W. Cole, *Transglutaminase: mechanistic features of the active site as determined by kinetic and inhibitor studies.* Biochim Biophys Acta, 1966. **122**(2): p. 244-64.

58. Folk, J.E. and S.I. Chung, *Transglutaminases.* Methods Enzymol, 1985. **113**: p. 358-75.

59. Ráki, M., et al., *Surface expression of transglutaminase 2 by dendritic cells and its potential role for uptake and presentation of gluten peptides to T cells.* Scand J Immunol, 2007. **65**(3): p. 213-20.

60. van de Wal, Y., et al., *Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity.* J Immunol, 1998. **161**(4): p. 1585-8.

61. Molberg, O., et al., *Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease.* Nat Med, 1998. **4**(6): p. 713-7.

62. Caputo, I., et al., *Tissue transglutaminase in celiac disease: role of autoantibodies.* Amino Acids, 2009. **36**(4): p. 693-9.

63. Lindfors, K., K. Kaukinen, and M. Mäki, *A role for anti-transglutaminase 2 autoantibodies in the pathogenesis of coeliac disease?* Amino Acids, 2009. **36**(4): p. 685-91.

64. Mazzarella, G., et al., *An immunodominant DQ8 restricted gliadin peptide activates small intestinal immune response in in vitro cultured mucosa from HLA-DQ8 positive but not HLA-DQ8 negative coeliac patients.* Gut, 2003. **52**(1): p. 57-62.

65. Lundin, K.E., et al., *T lymphocyte recognition of a celiac disease-associated cis- or trans-encoded HLA-DQ alpha/beta-heterodimer.* J Immunol, 1990. **145**(1): p. 136-9.

66. Nilsen, E.M., et al., *Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma.* Gut, 1995. **37**(6): p. 766-76.

67. Troncone, R., et al., *Majority of gliadin-specific T-cell clones from celiac small intestinal mucosa produce interferon-gamma and interleukin-4.* Dig Dis Sci, 1998. **43**(1): p. 156-61.

68. Monteleone, G., et al., *Role of interferon alpha in promoting T helper cell type 1 responses in the small intestine in coeliac disease.* Gut, 2001. **48**(3): p. 425-9.

69. León, A.J., et al., *Interleukin 18 maintains a long-standing inflammation in coeliac disease patients.* Clin Exp Immunol, 2006. **146**(3): p. 479-85.

70. Steinman, L., *A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage.* Nat Med, 2007. **13**(2): p. 139-45.

71. Castellanos-Rubio, A., et al., *TH17 (and TH1) signatures of intestinal biopsies of CD patients in response to gliadin.* Autoimmunity, 2009. **42**(1): p. 69-73.

72. Harris, K.M., A. Fasano, and D.L. Mann, *Monocytes differentiated with IL-15 support Th17 and Th1 responses to wheat gliadin: implications for celiac disease.* Clin Immunol, 2010. **135**(3): p. 430-9.

73. Monteleone, I., et al., *Characterization of IL-17A-producing cells in celiac disease mucosa.* J Immunol, 2010. **184**(4): p. 2211-8.

74. Sjöström, H., et al., *Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition.* Scand J Immunol, 1998. **48**(2): p. 111-5.

75. Fina, D., et al., *Interleukin 21 contributes to the mucosal T helper cell type 1 response in coeliac disease.* Gut, 2008. **57**(7): p. 887-92.

76. Juuti-Uusitalo, K., et al., *cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples.* J Autoimmun, 2004. **22**(3): p. 249-65.

77. Diosdado, B., et al., *A microarray screen for novel candidate genes in coeliac disease pathogenesis.* Gut, 2004. **53**(7): p. 944-51.

78. Kumar, V., et al., *Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels.* Hum Mol Genet, 2015. **24**(2): p. 397-409.

79. Parmar, A., et al., *Gene Expression Profiling of Gliadin Effects on Intestinal Epithelial Cells Suggests Novel Non-Enzymatic Functions of Pepsin and Trypsin.* PLoS One, 2013. **8**(6): p. e66307.

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

81. Suzuki, T., *Regulation of intestinal epithelial permeability by tight junctions.* Cell Mol Life Sci, 2013. **70**(4): p. 631-59.

82. Cereijido, M., L. G onzalez-Mariscal , and G. Contreras, *Tight Junction: Barrier Between Higher Organisms and Environment.* Physiology. p. 72-75.

83. Turner, J.R., *Intestinal mucosal barrier function in health and disease.* Nat Rev Immunol, 2009. **9**(11): p. 799-809.

84. Zihni, C., et al., *Tight junctions: from simple barriers to multifunctional molecular gates.* Nat Rev Mol Cell Biol, 2016. **17**(9): p. 564-80.

85. Furuse, M., et al., *Occludin: a novel integral membrane protein localizing at tight junctions.* J Cell Biol, 1993. **123**(6 Pt 2): p. 1777-88.

86. Kotler, B.M., J.E. Kerstetter, and K.L. Insogna, *Claudins, dietary milk proteins, and intestinal barrier regulation.* Nutr Rev, 2013. **71**(1): p. 60-5.

87. Mineta, K., et al., *Predicted expansion of the claudin multigene family.* FEBS Lett, 2011. **585**(4): p. 606-12.

88. Krause, G., J. Protze, and J. Piontek, *Assembly and function of claudins: Structure-function relationships based on homology models and crystal structures.* Semin Cell Dev Biol, 2015. **42**: p. 3-12.

89. Ikenouchi, J., et al., *Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells.* J Cell Biol, 2005. **171**(6): p. 939-45.

90. Martìn-Padura, I., et al., *Junctional adhesion molecule, a novel member of the immunoglobulin superfamily that distributes at intercellular junctions and modulates monocyte transmigration.* J Cell Biol, 1998. **142**(1): p. 117-27.

91. Matter, K. and M.S. Balda, *Epithelial tight junctions, gene expression and nucleo-junctional interplay.* J Cell Sci, 2007. **120**(Pt 9): p. 1505-11.

92. Guillemot, L., et al., *The cytoplasmic plaque of tight junctions: a scaffolding and signalling center.* Biochim Biophys Acta, 2008. **1778**(3): p. 601-13.

93. Fanning, A.S. and J.M. Anderson, *Zonula occludens-1 and -2 are cytosolic scaffolds that regulate the assembly of cellular junctions.* Ann N Y Acad Sci, 2009. **1165**: p. 113-20.

94. McCaffrey, L.M. and I.G. Macara, *Signaling pathways in cell polarity.* Cold Spring Harb Perspect Biol, 2012. **4**(6).

95. Matter, K. and M.S. Balda, *Signalling to and from tight junctions.* Nat Rev Mol Cell Biol, 2003. **4**(3): p. 225-36.

96. Hollander, D., et al., *Increased intestinal permeability in patients with Crohn's disease and their relatives. A possible etiologic factor.* Ann Intern Med, 1986. **105**(6): p. 883-5.

97. Büning, C., et al., *Increased small intestinal permeability in ulcerative colitis: rather genetic than environmental and a risk factor for extensive disease?* Inflamm Bowel Dis, 2012. **18**(10): p. 1932-9.

98. Schulzke, J.D., et al., *Epithelial tight junction structure in the jejunum of children with acute and treated celiac sprue.* Pediatr Res, 1998. **43**(4 Pt 1): p. 435-41.

99. Edelblum, K.L. and J.R. Turner, *The tight junction in inflammatory disease: communication breakdown.* Curr Opin Pharmacol, 2009. **9**(6): p. 715-20.

100. Hollander, D., *Intestinal permeability, leaky gut, and intestinal disorders.* Curr Gastroenterol Rep, 1999. **1**(5): p. 410-6.

101. DeMeo, M.T., et al., *Intestinal permeation and gastrointestinal disease.* J Clin Gastroenterol, 2002. **34**(4): p. 385-96.

102. Balda, M.S., et al., *Functional dissociation of paracellular permeability and transepithelial electrical resistance and disruption of the apical-basolateral intramembrane diffusion barrier by expression of a mutant tight junction membrane protein.* J Cell Biol, 1996. **134**(4): p. 1031-49.

103. Schumann, M., et al., *Celiac Disease: Role of the Epithelial Barrier.* Cell Mol Gastroenterol Hepatol, 2017. **3**(2): p. 150-162.

104. Schumann, M., et al., *Cell polarity-determining proteins Par-3 and PP-1 are involved in epithelial tight junction defects in coeliac disease.* Gut, 2012. **61**(2): p. 220-8.

105. Szakál, D.N., et al., *Mucosal expression of claudins 2, 3 and 4 in proximal and distal part of duodenum in children with coeliac disease.* Virchows Arch, 2010. **456**(3): p. 245-50.

106. Montalto, M., et al., *Immunohistochemical analysis of ZO-1 in the duodenal mucosa of patients with untreated and treated celiac disease.* Digestion, 2002. **65**(4): p. 227-33.

107. Wapenaar, M.C., et al., *Associations with tight junction genes PARD3 and MAGI2 in Dutch patients point to a common barrier defect for coeliac disease and ulcerative colitis.* Gut, 2008. **57**(4): p. 463-7.

108. Sturgeon, C. and A. Fasano, *Zonulin, a regulator of epithelial and endothelial barrier functions, and its involvement in chronic inflammatory diseases.* Tissue Barriers, 2016. **4**(4): p. e1251384.

109. Khaleghi, S., et al., *The potential utility of tight junction regulation in celiac disease: focus on larazotide acetate.* Therap Adv Gastroenterol, 2016. **9**(1): p. 37-49.

110. Groschwitz, K.R. and S.P. Hogan, *Intestinal barrier function: molecular regulation and disease pathogenesis.* J Allergy Clin Immunol, 2009. **124**(1): p. 3-20; quiz 21-2.

111. Hall, E.J. and R.M. Batt, *Abnormal permeability precedes the development of a gluten sensitive enteropathy in Irish setter dogs.* Gut, 1991. **32**(7): p. 749-53.

112. Vogelsang, H., et al., *Screening for celiac disease in first-degree relatives of patients with celiac disease by lactulose/mannitol test.* Am J Gastroenterol, 1995. **90**(10): p. 1838-42.

113. Tsukita, S., M. Furuse, and M. Itoh, *Multifunctional strands in tight junctions.* Nat Rev Mol Cell Biol, 2001. **2**(4): p. 285-93.

114. Martin-Pagola, A., et al., *MICA response to gliadin in intestinal mucosa from celiac patients.* Immunogenetics, 2004. **56**(8): p. 549-554.

115. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

116. Dye, J.F., et al., *Cyclic AMP and acidic fibroblast growth factor have opposing effects on tight and adherens junctions in microvascular endothelial cells in vitro.* Microvasc Res, 2001. **62**(2): p. 94-113.

117. Laukoetter, M.G., P. Nava, and A. Nusrat, *Role of the intestinal barrier in inflammatory bowel disease.* World J Gastroenterol, 2008. **14**(3): p. 401-7.

118. McCarthy, K.M., et al., *Inducible expression of claudin-1-myc but not occludin-VSV-G results in aberrant tight junction strand formation in MDCK cells.* J Cell Sci, 2000. **113 Pt 19**: p. 3387-98.

119. Itoh, M., et al., *Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins.* J Cell Biol, 1999. **147**(6): p. 1351-63.

120. Furuse, M., et al., *Direct association of occludin with ZO-1 and its possible involvement in the localization of occludin at tight junctions.* J Cell Biol, 1994. **127**(6 Pt 1): p. 1617-26.

121. Mitic, L.L., C.M. Van Itallie, and J.M. Anderson, *Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: lessons from mutant animals and proteins.* Am J Physiol Gastrointest Liver Physiol, 2000. **279**(2): p. G250-4.

122. Du, D., et al., *The tight junction protein, occludin, regulates the directional migration of epithelial cells.* Dev Cell, 2010. **18**(1): p. 52-63.

123. Thomas, K.E., et al., *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, 2006. **176**(4): p. 2512-21.

124. Fasano, A. and T. Shea-Donohue, *Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases.* Nat Clin Pract Gastroenterol Hepatol, 2005. **2**(9): p. 416-22.

125. Rauhavirta, T., et al., *Impaired epithelial integrity in the duodenal mucosa in early stages of celiac disease.* Transl Res, 2014. **164**(3): p. 223-31.

126. Amundsen, S.S., et al., *Association analysis of MYO9B gene polymorphisms with celiac disease in a Swedish/Norwegian cohort.* Hum Immunol, 2006. **67**(4-5): p. 341-5.

127. Chen, Y.Q., et al., *Lack of Association between MYO9B Gene Polymorphisms and Susceptibility to Coeliac Disease in Caucasians: Evidence from a Meta-Analysis.* Immunol Invest, 2016. **45**(5): p. 396-405.

128. Qu, Z. and D.L. Adelson, *Evolutionary conservation and functional roles of ncRNA.* Front Genet, 2012. **3**: p. 205.

129. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature, 2009. **458**(7235): p. 223-7.

130. Bjarnason, I. and T.J. Peters, *In vitro determination of small intestinal permeability: demonstration of a persistent defect in patients with coeliac disease.* Gut, 1984. **25**(2): p. 145-50.

131. Hopcraft, S.E. and M.J. Evans, *Selection of a hepatitis C virus with altered entry factor requirements reveals a genetic interaction between the E1 glycoprotein and claudins.* Hepatology, 2015. **62**(4): p. 1059-69.

132. Van Itallie, C.M., et al., *A complex of ZO-1 and the BAR-domain protein TOCA-1 regulates actin assembly at the tight junction.* Mol Biol Cell, 2015. **26**(15): p. 2769-87.

133. Goldstein, B. and I.G. Macara, *The PAR proteins: fundamental players in animal cell polarization.* Dev Cell, 2007. **13**(5): p. 609-22.

134. Shin, K., V.C. Fogg, and B. Margolis, *Tight junctions and cell polarity.* Annu Rev Cell Dev Biol, 2006. **22**: p. 207-35.

135. Shin, K., S. Straight, and B. Margolis, *PATJ regulates tight junction formation and polarity in mammalian epithelial cells.* J Cell Biol, 2005. **168**(5): p. 705-11.

136. Ridley, A.J., et al., *Cell migration: integrating signals from front to back.* Science, 2003. **302**(5651): p. 1704-9.

137. Roh, M.H., et al., *The carboxyl terminus of zona occludens-3 binds and recruits a mammalian homologue of discs lost to tight junctions.* J Biol Chem, 2002. **277**(30): p. 27501-9.

138. Bryant, P.J. and P. Simpson, *Intrinsic and extrinsic control of growth in developing organs.* Q Rev Biol, 1984. **59**(4): p. 387-415.

139. Johnston, L.A. and P. Gallant, *Control of growth and organ size in Drosophila.* Bioessays, 2002. **24**(1): p. 54-64.

140. Marsh, M.N., *Clinical and pathological spectrum of coeliac disease.* Gut, 1993. **34**(12): p. 1740; author reply 1741.

141. Juuti-Uusitalo, K., et al., *Gluten affects epithelial differentiation-associated genes in small intestinal mucosa of coeliac patients.* Clin Exp Immunol, 2007. **150**(2): p. 294-305.
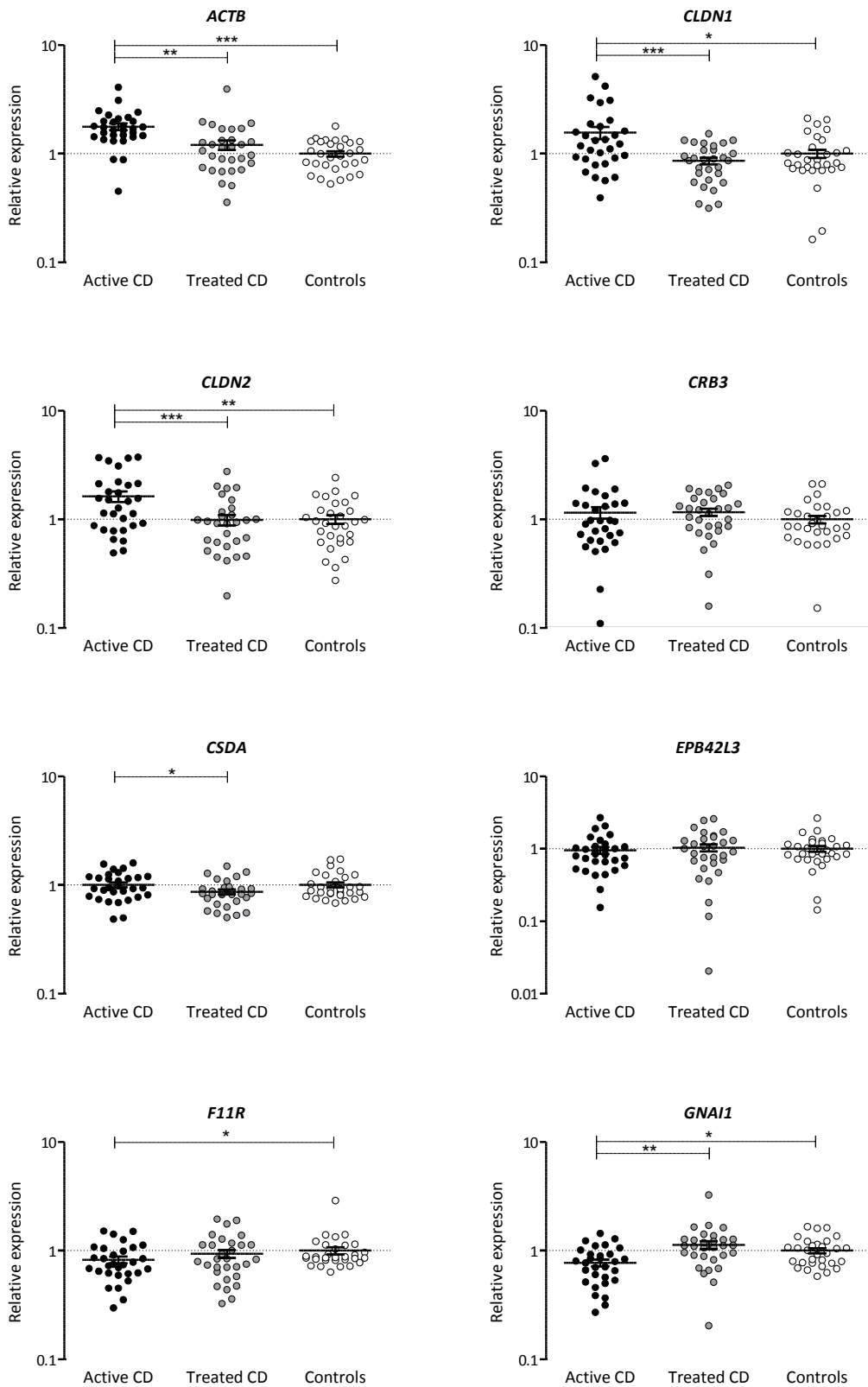
8. Supplementary

Table S1. Clinical, immunological and HLA information of the celiac patients included in the study. F: female; M: male; Dx: Diagnosis.

	GENDER	AGE AT DX (month)	tTG at Dx	Marsh at Dx	DR	HLA
1	F	11,0	62	3c	DR3/DRX	DQ2
2	F	96,0	110	3c	DR3/DR7	DQ2
3	F	18,0	37	3a/b	DR3/DRX	DQ2
4	M	13,0	131	3c	DR3/DR7	DQ2
5	F	12,0	104	3a/b	DR3/DR3	DQ2
6	F	12,0	106	3b	DR3/DRX	DQ2
7	F	41,0	76	3c	DR3/DR4	DQ2/DQ8
8	F	15,0	116	3c	DR3/DR7	DQ2
9	M	22,0	113	3c	DR3/DR7	DQ2
10	F	9,0	122	3c	DR3/DR7	DQ2
11	F	20,0	127	3c	DR4/DR7	DQ2/DQ8
12	M	8,0	125	3c	DR3/DR4	DQ2/DQ8
13	M	0,0	105	3c	DR3/DRX	DQ2
14	F	22,0	103	3b/c	DR5/DR7	DQ2
15	F	6,0	113	3c	DR3/DR4	DQ2/DQ8
16	M	6,0	84	3c	DR3/DR7	DQ2
17	F	17,0	113	3c	DR4/DR5	DQ8
18	F	20,0	49	3c	DR3/DRX	DQ2
19	F	16,0	103	3c	DR3/DRX	DQ2
20	M	24,0	77	3c	DR3/DR3	DQ2
21	M	26,0	109	3c	DR3/DR7	DQ2
22	F	30,0	49	3a/b	DR5/DR7	DQ2
23	M	12,0	108	3c	DR3/DR3	DQ2
24	M	19,0	100	3c	DR3/DRX	DQ2
25	M	34,0	278	3c	DR3/DR7	DQ2
26	F	96,0	102	3c	DR3/DR4	DQ2/DQ8
27	F	20,0	501	3c	DR3/DR4	DQ2/DQ8

Table S2. TaqMan assays information.

	Gene symbol	Assay code	Encoding protein	Protein family	Pathway
1	*CLDN1*	Hs00221623_m1	Claudin 1	**Transmembrane proteins**	
2	*CLDN2*	Hs00252666_s1	Claudin 2		
3	*CRB3*	Hs01548179_m1	Crumbs homolog 3 (*Drosophila*)		
4	*F11R*	Hs00170991_m1	Junctional adhesion molecule 1 (JAM-1)		
5	*OCLN*	Hs00170162_m1	Occludin		
6	*INADL*	Hs00195106_m1	Pals1-Associated Tight Junction Protein (PATJ)	**TJ Adaptor Proteins**	**TJ**
7	*MAGI1*	Hs00191026_m1	Membrane associated guanylate kinase 1		
8	*MAGI2*	Hs00202321_m1 Hs01111613_m1	Membrane associated guanylate kinase 2		
9	*MPDZ*	Hs01075090_m1	Multiple PDZ domain protein (MUPP1)		
10	*PARD3*	Hs00969077_m1	Partitioning defective 3 homolog (*C. elegans*)		
11	*PARD6A*	Hs00180947_m1	Partitioning defective 6 homolog α (*C. elegans*)		
12	*TJP1*	Hs01551861_m1	Zonula occludens 1 (ZO-1)		
13	*TJP2*	Hs00910543_m1	Zonula occludens 2 (ZO-2)		
14	*TJP3*	Hs00274276_m1	Zonula occludens 3 (ZO-3)		
15	*CSDA*	Hs01124964_m1	ZO-1-Associated Nucleic Acid-Binding (ZONAB)	**TJ Regulators**	
16	*EPB41L3*	Hs00202360_m1	Erythrocyte membrane protein band 4.1-like 3		
17	*GNAI1*	Hs01053353_m1	G protein, alpha inhibiting activity polypeptide 1		
18	*PPP1CA*	Hs00267568_m1	Protein phosphatase 1, catalitic subunit, α		
19	*PPP2R3A*	Hs01097014_m1	Protein phosphatase 2, regulatory subunit B,α		
20	*SYMPK*	Hs00191361_m1	Symplekin		
21	*ZAK*	Hs00370447_m1	Sterile α motif & leucine zipper containing kinase		
22	*ACTB*	Hs01060665_g1	Actin, beta	**Cytoskeletal Filaments**	
23	*MYH14*	Hs00226855_m1	Myosin, heavy chain 14, non-muscle		
24	*MYO9B*	Hs00188109_m1	Myosin IXB		
25	*MYD88*	Hs01573837-g1	Myeloid Differentiation Primary Response 88	TLR adaptor proteins	**TLR**
26	*TICAM1*	Hs00706140-s1	Toll Like Receptor Adaptor Molecule 1		
27	*TICAM2*	Hs04189225_m1	Toll Like Receptor Adaptor Molecule 2		
28	*TIRAP*	Hs00364644_m1	TIR Domain Containing Adaptor Protein		
29	*TOLLIP*	Hs01553188_m1	Toll Interacting Protein		
30	*IRAK1*	Hs00155570_m1	Interleukin 1 Receptor Associated Kinase 1	TLR Regulators	
31	*IRF5*	Hs00158114_m1	Interferon Regulatory Factor 5		
32	*IRF7*	Hs01014809-g1	Interferon Regulatory Factor 7		

33	*TRAF6*	Hs00371512-g1	TNF Receptor Associated Factor 6	
34	*MAP3K7*	Hs00177373_m1	Mitogen-Activated Protein Kinase Kinase Kinase 7	
35	*AKT1*	Hs00178289_m1	AKT Serine/Threonine Kinase 1	
36	*RAC1*	Hs01902432-s1	Ras-Related C3 Botulinum Toxin Substrate 1	
37	*RAC2*	Hs01032884_m1	Ras-Related C3 Botulinum Toxin Substrate 2	
38	*MALT1*	Hs01120052_m1	MALT1 Paracaspase	
39	*NFKB1*	Hs00765730_m1	Nuclear Factor Kappa B Subunit 1	NF-kappaB Signaling
40	*REL*	Hs00968436_m1	REL Proto-Oncogene, NF-KB Subunit	
41	*STAT3*	Hs01047580_m1	Signal Transducer And Activator Of Transcription 3	
42	*CCL5*	Hs00982282_m1	C-C Motif Chemokine Ligand 5	
43	*CXCL1*	Hs00605382-gh	C-X-C Motif Chemokine Ligand 1	
44	*CXCL10*	Hs00171042_m1	C-X-C Motif Chemokine Ligand 10	
45	*CXCL11*	Hs04187682-g1	C-X-C Motif Chemokine Ligand 11	
46	*IFNA1*	Hs04189288-g1	Interferon Alpha 1	Inflammatory response
47	*IFNB1*	Hs01077958-s1	Interferon Beta 1	
48	*IL1B*	Hs00174097_m1	Interleukin 1 Beta	
49	*IL6*	Hs00174131_m1	Interleukin 6	
50	*IL15*	Hs01003716_m1	Interleukin 15	
51	*TNF*	Hs00174128_m1	Tumor Necrosis Factor	
52	*RPLP0*	4333761F	Ribosomal Protein Lateral Stalk Subunit P0	Endogenous control

INADL

MAGI1

MAGI2

MPDZ

MYH14

OCLN

PARD3

PARD6A
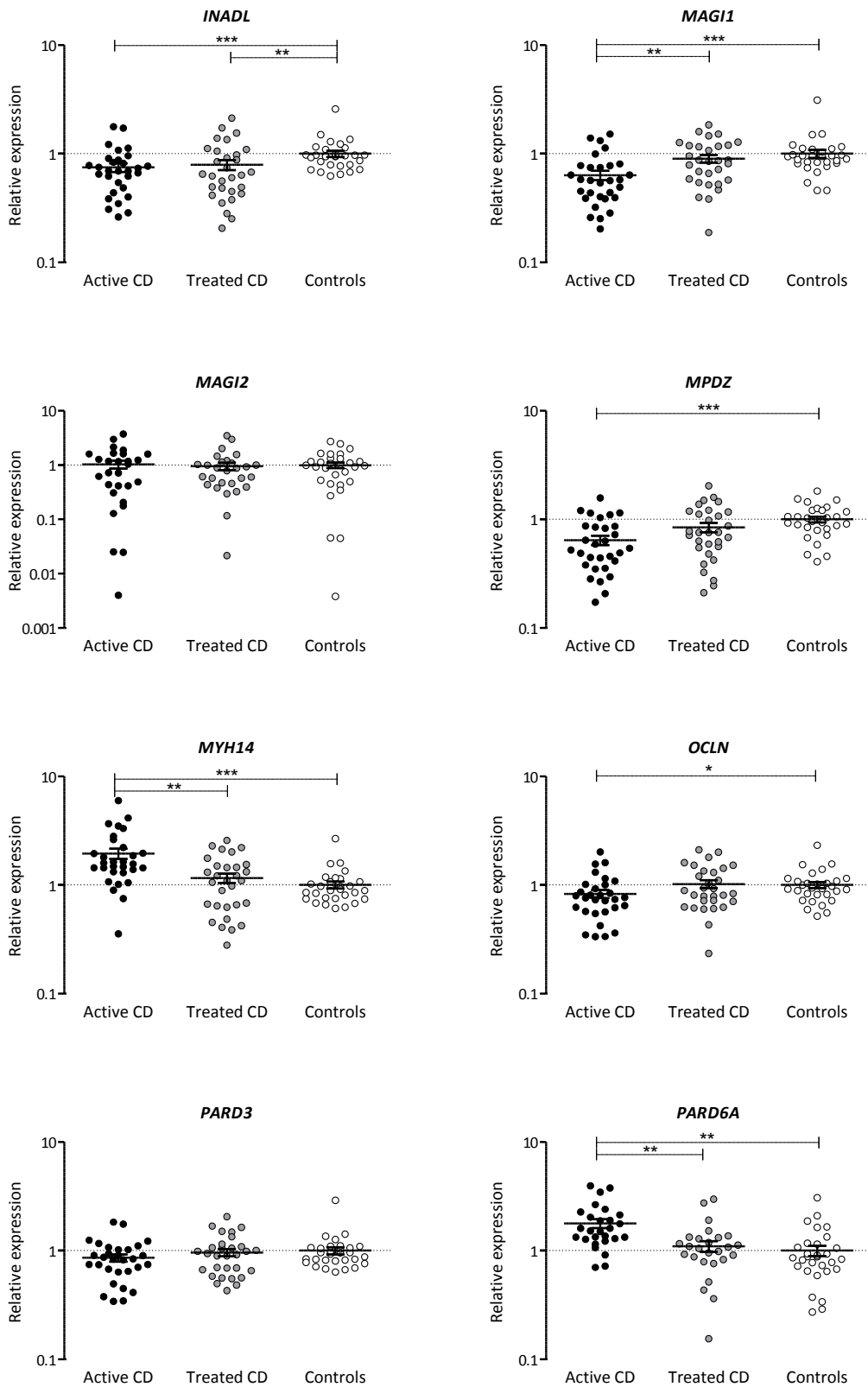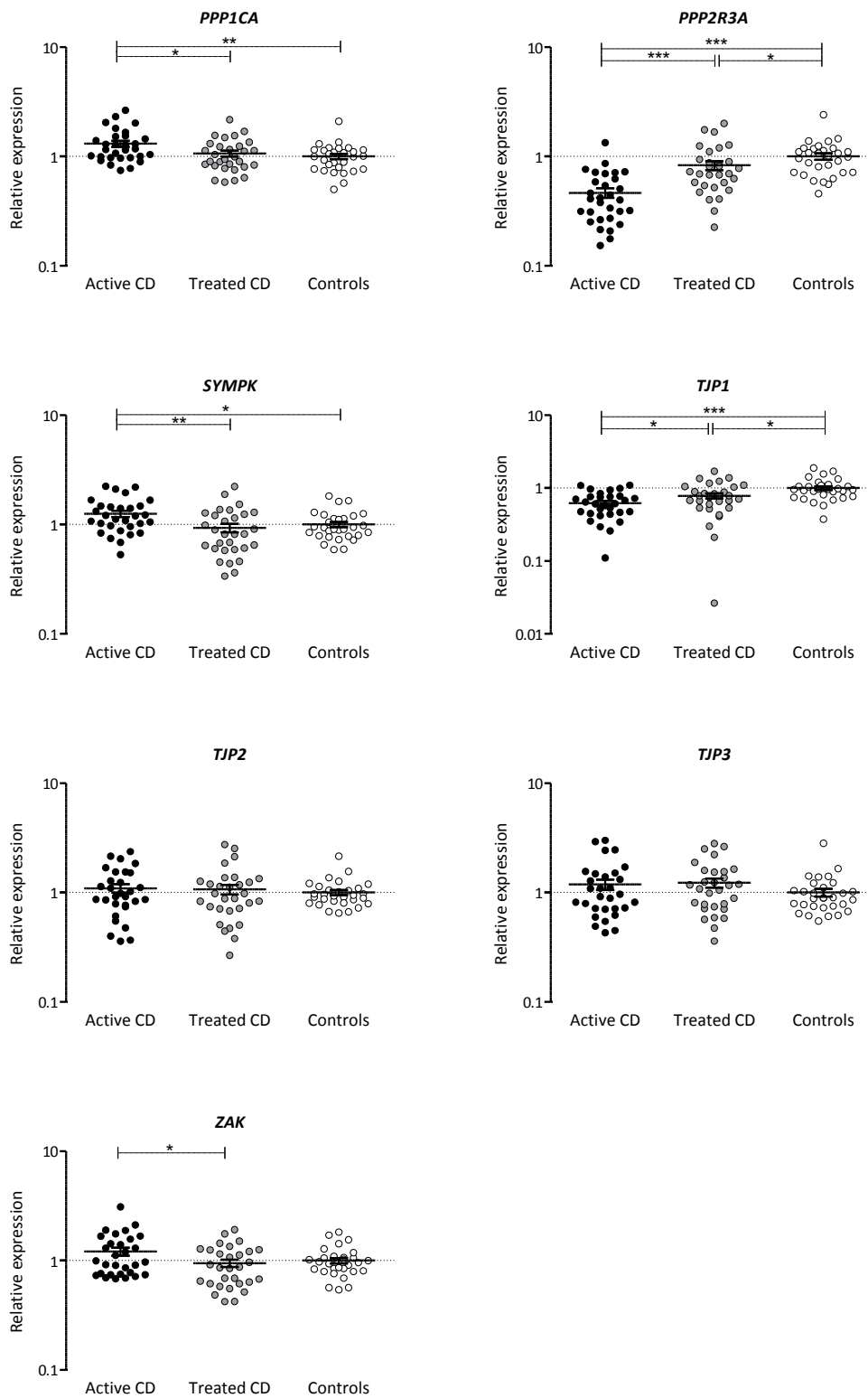
Supplementary figure 1. Relative gene expression of TJ related genes in active CD, treated CD and controls. Expression values were calculated relative to the average of non-celiac control samples. The graph shows the means and the standard error of the mean. Differences among groups are shown by asterisk (* p-value 0.01 < 0.05; ** p-value 0.001 < 0.01).
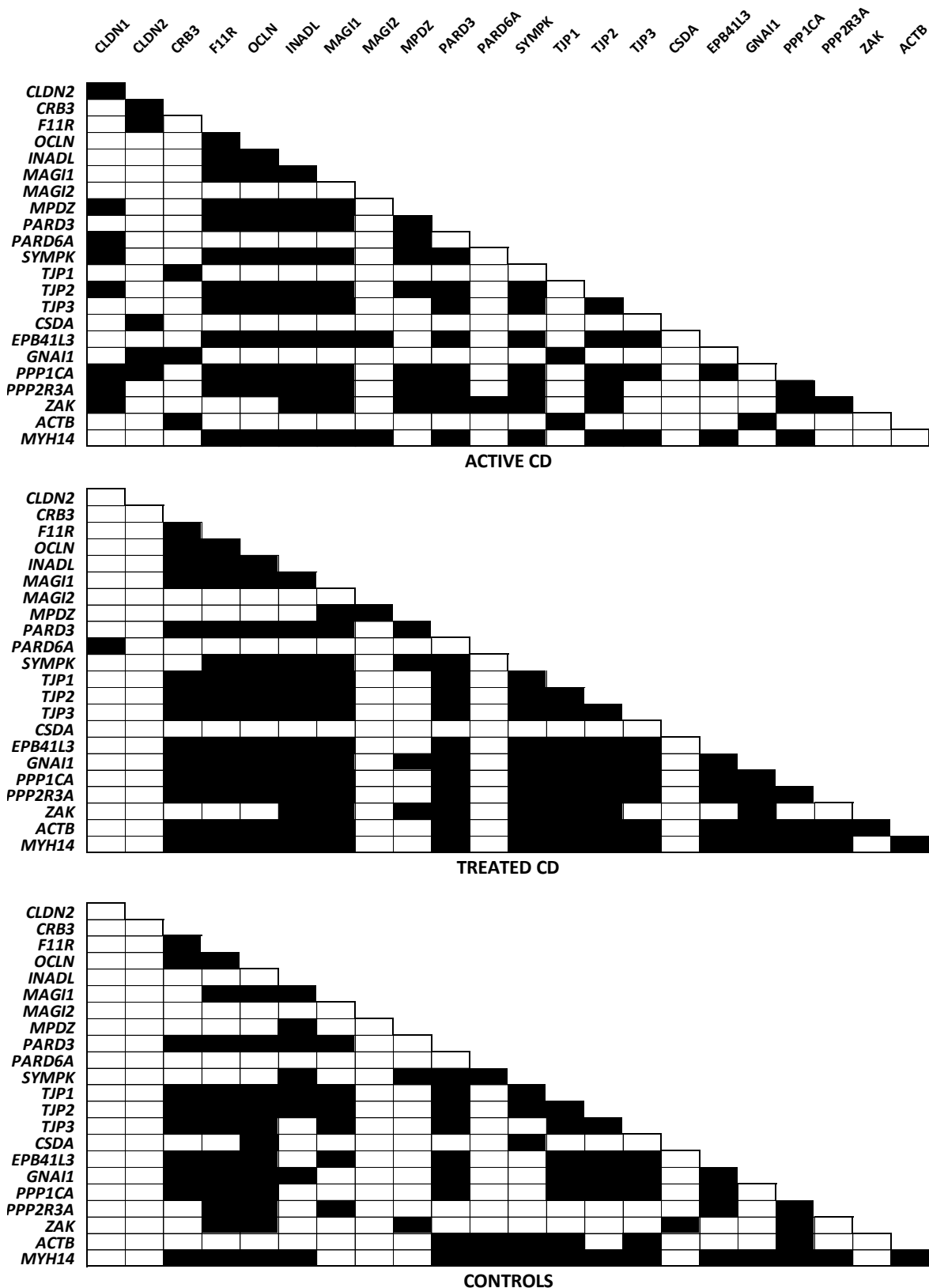
Supplementary figure 2. Correlation matrix depicting patterns of coexpression among all candidate genes in patients with active CD, treated CD and non-CD controls. Black squares represent significant correlation between the corresponding pair of genes (Spearman p-value <0.05).

111