eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

IDENTIFICATION OF GENETIC AND EPIGENETIC REGULATORS IN CELIAC DISEASE THROUGH COMPUTATIONAL AND EXPERIMENTAL APPROACHES

Doctoral thesis

Irati Romero Garmendia

Supervised by Jose Ramon Bilbao and Nora Fernandez-Jimenez

Leioa, 2019

This work was funded by a predoctoral fellowship from the University of the Basque Country UPV/EHU to Irati Romero (PIF2014/408) and research project grants from the 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 (PI13/01201 and PI16/00258) and the Basque Department of Health (2011/111034). Technical and human support provided by SGIker (UPV/EHU, MICINN, GV/EJ, ERDF and ESF) is gratefully acknowledged.

“Hobe egarri izan eta urik ez
Ura izan eta egarririk ez baino.”
Anari.

Lehenengo eta behin, eskerrak eman nahi dizkiet urte hauetan laborategian topatu ditudan lankide eta lagunei, denak izan baitzarete modu batean edo bestean lan honen parte.

Eskerrik asko nire zuzendari izan diren Buli eta Norari. Buli mila esker Immunogenetics taldean sartzeko eta tesia egiteko aukera emateagatik. Zuri esker ikertzaile karrera hasteko aukera izan dut eta zu gabe guzti hau ez zen posible izango. Nora, mila esker laborategian master ikasle nintzela egin zenidan harrera goxoagatik eta urte hauetan nire lan eta etorkizunarekiko izan duzun arretagatik. Mila esker nire tesiko atzerriko estantzia posible egiteagatik eta bertan eskaini zenidan esperientzia pertsonal zein profesional aberasgarriagatik.

Ainarari, zure alaitasuna eta laguntza ezinbestekoak izan dira tesiko bigarren erdi honetan. Hamar mila gauzetara zabiltzan harren beti zara hamar mila eta bat gauzetan aritzeko kapaza. Eskerrik asko eskaini didazun denboragatik. Izortzeri, eskerrik asko zure laguntza eta babesagatik. Asko ikasi dut zure lan egiteko moduagatik, beti txukun.

Duela gutxi hasi diren Ane eta Maialeni ere eskerrak eman nahi dizkizuet. Aldrebesturik nuen egun bat baino gehiagori buelta eman diozue zuen gogo eta energiekin. Animo bioi, paregabeak zarete!

Eskerrik asko ere laborategia utzi zenuten arren gertu zaudetenoi. Leti, oso pertsona alaia zara, beti zerbait berria ikasteko eta irakasteko prest dagoena. Koldo, mila esker ezagutzen ez nuen mundu bat aurkezteagatik. Lan bikaina egiten duzu, jartoetan jartoena. Eskerrik asko mundu kaotiko horretarako sarrera erakusteagatik eta beti laguntzeko prest azaltzeagatik. Teresa, laborategian zein laborategitik kanpo pasa ditugun momentu on guztiengatik, beti gogoratuko ditugu par artean. Donatella grazie mille per la tua gioia e amicizia, buona fortuna con tutto.

Eskerrak gurutzetako jendeari eta kolaboratzaileak izan diren medikuei. Baita, paziente zein haien familiei beraien eskuzabaltasunagatik. Gure Bunkerreko kideei, bereziki Angela, Idoia eta Nereari, zuen hurbiltasunagatik eta pasatako momentuengatik. Mila esker, Lyonen ezagututako Dr. Herceg eta Epigenetics-eko talde osoari zuen harreragatik eta lehenengo egunetik taldeko parte egiteagatik.

Doktoretzako urte hauetan zehar pertsona asko sentitu ditut ondoan, ezin izango nukeen lan hau gauzatu familia eta lagunen babesik gabe. Eskerrak eman nahi dizkizuet guztioi, aipatuko ditudanei eta aipatu gabe geratuko direnei ere.

Lehenik eta behin eskerrak eman nahi dizkiot nire kuadrilari. Ane, Claudia, Goizane eta Maialen, mila esker azken aldian gutxiago ikusi garen arren beti presente egoteagatik. Zuen animoak oso garrantzitsuak izan dira niretzat. Mila esker azken hilabeteetan urrun egon naizen arren nire lekutxoa gordetzeagatik eta maitatua sentiarazteagatik.

Bilboko lagunei, batez ere azken urteetan nire bizitzan agertu den “Bilbo Peñari”. Talde heterogeneoa, bata besteengandik ikasteko prest dagoena. A Miguel el Murciano Cubano por alegrarnos con sus chistes e historias curiosas; Marc, zuen etxean beti ongietorria sentiarazteagatik; Itzi eta Aritz, asko baloratzen ditut zuen ontasuna eta giro ona jartzeko duzuen ahalmena; Aratz, 10 urte dira ezagutu ginela eta asko pozten nau zu ondoan izaten jarraitzeak, eskerrik asko mila momentutan indarra emateagatik; gracias Shukhrat por tu amabilidad y cariño recibidos durante este tiempo, siempre detallista. Mila esker Txekas zaren laguna izateagatik. Beti sentitu zaitut hurbileko eta momentu askotan izan zara euskarri, askotan konturatu ez zaren arren. Mila esker ostiraletako bisitengatik eta high-ko ekipoko buru izateagatik.

Bereziki eskertu nahiko nituzke Manuelitak, azken urteetan familia izan baikara. Araia, Urreta, Ainara eta Amaia mila esker elkarbizitzan pasa ditugun momentu

alaiengatik. Urreta, etxera etorri zinen lehenengoan argia joan zen arren kolorez bete zenuen etxea. Mila esker izan ditugun elkarrizketa eta barre algarengatik, ez daitezela bukatu! Araia beti eskuzabal, eskerrik asko zure tratu goxoagatik eta etxe-giroa sortzeagatik, kasu honetan ere ez daitezela barreak eta dantzak bukatu! Ainara, badira urte batzuk ezagutu ginela baina denbora pasa ahala geroz eta pertsona interesgarriago bat ezagutzen dut zurekin. Eskerrak eman nahi dizkizut izan zaren lagunagatik, beti ondoan egoteagatik, eta batez ere feminismoaren munduan izan zaren erreferentziagatik. Asko ikasi dut zurekin. Amaia, "siamesaatiempocompleto" azken hilabeteetan herrialde desberdinetan bizi garen arren bata bestearengandik beti bezain gertu jarraitu dugulako. Urte hauetan gertuen izan dudan pertsona izan zara, zentzu askotan, eta eskerrak eman nahi dizkizut nire bizitza aberastu duzun bezala aberasteagatik. Ez nintzateke gaur egun naizena izango zure ekarpenengatik izan ez balitz, eredu zara zentzu askotan. Azken urteetako malko eta algaren testigu izan zara, beti zintzo, entzuteko prest eta pertsonak handi bihurtzeko duzun don horrekin. Lan hau zurea ere bada.

Manu, karreran zehar abentura batzuk bizi izan ditugu eta oraindik ere ez gara nekatu! Eskerrik asko beti gertu egoteagatik. Animo tesiarekin, aurrera! Rodri, me ha encantado tenerte de flatemate y reírme tantas veces contigo, ve eligiendo SPA! Iosu eta Ainhoa, nirekin izan duzuen pazientziagatik eta eman didazuen maitasun eta alaitasunagatik. Guztiak aipatu ezin ditudan arren badira beste asko eskertu nahiko nituzkeenak.

A toda la familia de David por acogerme y hacerme una más de la familia. Sobre todo a Mari Carmen y a Antonio por todo el cariño y generosidad que me habéis ofrecido.

Azkenik nire familia eskertu nahi dut. Aitona eta amona pertsona eredugarriak zarete. Zuek zarete familia lotuta mantentzen duzuenak eta guztiok zaintzen gaituzuenak. Beti onak, gehien behar duenari eskua luzatzen eta jendearen onena

sustatzen. Aita eta ama, nigatik egin duzuen guztiagatik, gehien behar izan zaituztedanetan beti egon zarete ondoan, beti ulerkor eta nire ongizatean pentsatuz. Eskerrik asko pasatu dizkidazuen balioengatik, pasio eta indarragatik. Harro nago zuetaz. Izaro, eskerrik asko nitaz arduratzeagatik eta ditugun jolas eta elkarrizketengatik, asteburuak poztu dizkidazu askotan. Zu ere pertsona indartsua zara, ideiak argi dituena eta horiek defendatzen dituena. Zu ere eredu izan zara askotan.

David, nire bizikide, familia eta lagun minenari. Egunak animatzen dituzu zure umore eta alaitasunarekin, mundua dantzan jartzen duzu eta horrela jarri ninduzun ni ere, dantzan. Eskerrik asko lerro zuzen horretatik kanpo geratzen diren burbuilatxoak sortzeagatik eta errespetuan oinarritzen den harreman polit honengatik. Zure indar, babes eta alaitasuna guztiz beharrezkoak izan dira askotan esan ez dizudan arren. Eskerrik asko nirekin izan duzun pazientziagatik, eta gehien behar izan dudanean eman didazun energia eta bultzadagatik. Beti eman didazu zure onena, bueltan ezer espero gabe. Orain zure txanda da, ondoan izango nauzu nik zu izan zaitudan bezala. Hau hasi besterik ez da egin!

ABBREVIATIONS

3D	Three-dimensional
AGA	Anti-Gliadin Antibodies
APC	Antigen Presenting Cells
CD	Celiac Disease
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation-sequencing
CTCF	CCCTC-binding factor protein
DC	Dendritic Cell
DCGL	Differential Co-expression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data
DCGs	Differentially Co-expressed Genes
DHSs	DNase I hypersensitive sites
DMEM	Dulbecco's Modified Eagle's Medium
DMMT	DNA Methyltransferase
DMP	Differentially Methylated Position
DMRs	Differentially Methylated Regions
EAE	Experimental Autoimmune Encephalomyelitis
EGF	Epidermal Growth Factor
EMA	anti-endomysium autoantibodies
eQTL	Expression Quantitative Trait *Locus*
ESPGHAN	European Society for Pediatric Gastroenterology, Hepatology and Nutrition
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FPKM	Fragments Per Kilobase Million
GEO	Gene Expression Omnibus
GFD	Gluten Free Diet
GO	Gene Ontology
GWA	Genome Wide Association
HLA	Human Leucocyte Antigen
IBD	Inflammatory Bowel Disease
IELS	Intraepithelial Lymphocytes
IFN	Interferon
LD	Linkage Disequilibrium
lncRNA	Long non-coding RNA
MHC	Mayor Histocompatibility Complex

miRNA	microRNA
MMP	Matrix Metalloproteinase
ncRNA	Non-coding RNA
NEAA	Non-Essential Amino Acids
NGS	Next-Generation Sequencing
NK	Natural Killer
NKR	NK Receptor
PAM	Protospacer Adjacent Motif
PCR	Polymerase Chain-Reaction
PMA	Phorbol 12-Myristate 13-Acetate
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
PT-BSA	Pepsin-Trypsin-digested Bovine Serum Albumin
PT-G	Pepsin-Trypsin-digested Gliadin
qRT-PCR	Quantitative Real-Time Polymerase Chain-Reaction
RCD	Refractory CD
RISC	RNA-Induced Silencing Complex
RNA-seq	RNA sequencing
ROS	Reactive Oxygen Species
RPMI	Roswell Park Memorial Institute
SNP	Single Nucleotide Polymorphism
SRA	Sequence Read Archive
T1D	Type 1 Diabetes
TADs	Topologically Associating Domains
TBST	Tris-Buffered Saline with 0.05% Tween
TF	Transcription Factor
TFBS	TF-Binding Sites
TG2	Tissue Transglutaminase type 2
TGA	Anti-Tissue Transglutaminase Autoantibodies
TLR	Toll-Like Receptors
tTG	Tissue Transglutaminase
UCSC	University of California Santa Cruz
WGCNA	Weighted Correlation Network Analysis
WR	Working Reagent
WT	Wild Type

GLOSSARY for *Whole genome co-expression in celiac disease*

Experiments

- **Acute experiment:** genome expression study of the response to an *in vitro* gliadin-challenge. The experiment consisted of Human U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, USA) data from duodenal biopsies taken from 10 patients on gluten free diet (GFD) that were cut into two portions that were incubated with or without 10 µg/ml gliadin for 4 hours. Data were downloaded from EBI Array Express database (http://www.ebi.ac.uk/microarray-as/ae/) experiment code E-MEXP-1823 (Castellanos-Rubio et al., 2008).

- **Long-term experiment:** whole genome expression study of the response to a chronic, dietary exposure to gliadin. The experiment consisted of Human U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, USA) data from duodenal biopsies taken from 9 active celiac disease (CD) patients (who were on a gluten-containing diet at that time) and 9 CD patients on GFD for more than two years. Data were downloaded from EBI Array Express database (http://www.ebi.ac.uk/microarray-as/ae/) experiment code E-MEXP-1828 (Castellanos-Rubio et al., 2008).

Samples

- **Gliadin-challenged samples** for the acute experiment; GFD patients' biopsies incubated with the addition of gliadin.
- **Unchallenged samples** for the acute experiment; GFD patients' biopsies incubated without the addition of gliadin.
- **Active samples** for the long-term experiment; biopsies of active CD patients who were on a gluten-containing diet at the time of endoscopy.

- **GFD samples** for the long-term experiment; biopsies of CD patients on a GFD for more than 2 years.

Comparisons

- **Gliadin-challenged samples *vs.* unchallenged samples and *vice versa*.**
- **Active samples *vs.* GFD samples and *vice versa*.**

Observations

- **Co-expression module:** group of co-expressed genes. Clustering is used to group genes with similar expression patterns across multiple samples to produce groups of co-expressed genes rather than pairs. The WGCNA package constructs co-expression modules using hierarchical clustering on a correlation network created from expression data. Hierarchical clustering divides each cluster into sub-clusters to create a tree with branches representing co-expression modules. Modules are then defined by cutting the branches at a certain height. These modules can be interrogated to identify regulators and functional enrichment.
- **Differentially co-expressed genes (DCGs):** genes that show changes in their co-expression relationships with their partners from a co-expression module between two situations. DCGs are highly correlated under one cell state but uncorrelated under another cell state.
- **Module of DCGs:** co-expression modules formed by DCGs. DCGs identified through comparison between two situations were grouped according to the co-expression modules they formed among them.

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

Romero-Garmendia I, Garcia-Etxebarria K, Hernandez-Vargas H, Santin I, Jauregi-Miguel A, Plaza-Izurieta L, Cros MP, Legarda M, Irastorza I, Herceg Z, Fernandez-Jimenez N, Bilbao JR. Transcription factor binding site enrichment analysis in co-expression modules in celiac disease. Genes. 2018 May; 10,9(5).

Romero-Garmendia I*., Jauregi-Miguel A*., Santin I., Bilbao J. R., Castellanos-Rubio A. Subcellular Fractionation from Fresh and Frozen Gastrointestinal Specimens. J. Vis. Exp. 2018 Jul 15;(137).

Fernandez-Jimenez N*, Garcia-Etxebarria K*, Plaza-Izurieta L*, **Romero-Garmendia I**, Jauregi-Miguel A, Legarda M, Ecsedi S, Castellanos-Rubio A, Cahais V, Cuenin C, Degli Esposti D, Irastorza I, Hernandez-Vargas H, Herceg Z, Bilbao JR. The methylome of the celiac intestinal epithelium harbours genotype-independent alterations in the HLA region. Sci Rep. 2018 Under review.

During this thesis I have also participated in the following publications:

Santin I, Jauregi-Miguel A, Velayos T, Castellanos-Rubio A, Garcia-Etxebarria K, **Romero-Garmendia I**, Fernandez-Jimenez N, Irastorza I, Castaño L, Bilbao JR. A Celiac Diasease Associated lncRNA Named HCG14 Regulates NOD1 Expression in Intestinal Cells. J Pediatr Gastroenterol Nutr. 2018 Aug;67(2):225-231.

Tentelier C, Barroso-Gomila O, Lepais O, Manicki A, **Romero-Garmendia I**, Jugo BM. Testing mate choice and overdominance at MH in natural families of Atlantic salmon Salmo salar. J Fish Biol. 2017 Apr;90(4):1644-1659.

Garcia-Etxebarria K, Jauregi-Miguel A, **Romero-Garmendia I**, Plaza-Izurieta L, Legarda M, Irastorza I, Bilbao JR. Ancestry-based stratified analysis of Immunochip data identifies novel associations with celiac disease. Eur J Hum Genet. 2016 Dec; 24(12):1831-1834.

Plaza-Izurieta L, Fernandez-Jimenez N, Irastorza I, Jauregi-Miguel A, **Romero-Garmendia I**, Vitoria JC, Bilbao JR. Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease. Eur J Hum Genet. 2015 Aug;23(8):1100-5.

PROJECT JUSTIFICATION AND SCOPE

Celiac disease (CD) is a chronic immune-mediated gastrointestinal 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 tissue transglutaminase (tTG), reflecting an ongoing immune mediated tissue-destruction process that initiates only among genetically predisposed individuals.

Therefore, it becomes essential to define which genes, pathways and regulatory mechanisms are involved in disease susceptibility. In this way pathogenic mechanisms underlying CD development will be better understood, and will provide genetic and epigenetic markers capable of discriminating individuals at risk of this disease. Additionally, the identification of master regulators of complex pathways and groups of genes will help to identify novel targets for intervention. Although gluten-free diet will probably remain the treatment of choice for celiac individuals, works like the present thesis will open the door to novel therapeutical alternatives for accidental transgressions to the diet or other acute intoxications with gluten. Moreover, the new findings could be extrapolated to other complex and autoimmune diseases and postulate new points of view to the understanding of genetic diseases.

In order to dissect the genetics and epigenetics of CD, the current project has focused on the search of functional determinants using novel bioinformatics approaches and on the functional validation of those candidates. More specifically, in the present doctoral thesis we have scrutinized four mechanisms of gene regulation: transcription factors (TF), microRNAs (miRNAs), chromatin organization and DNA methylation.

Introduction

1. Celiac disease

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

1.1. Clinical features and diagnosis

The adverse effects of ingested gluten in CD were not recognized until 1950 (Dicke, 1950) although the clinical picture of the disease had been first described by Samuel Gee more than 60 years before (Gee, 1888). If untreated, classical CD presents a wide 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) (Feighery, 1999). Other atypical symptoms are also associated with CD, and include neurological events, dental enamel defects, infertility, osteoporosis, joint symptoms and elevated liver enzyme concentrations (Mäki and Collin, 1997).

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 crypt hyperplasia. The degree and severity of gluten-induced mucosal alterations are classified in the Marsh-Oberhuber classification (Marsh, 1992) with Marsh 0 reflecting healthy, normal intestinal *villi*, and Marsh 4 indicating hypoplasia of small intestine architecture (Figure 1). Although this classification system has undergone some modifications, it is still used to identify and treat patients at an early stage of CD. In the first stage of the disease, there is an infiltration by intraepithelial

lymphocytes (IELs) of the villous epithelium (Marsh 1), which is followed by hyperplasia of the crypts (Marsh 2), while the *villi* are not shortened. In the more advanced stage (Marsh 3), 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 4), 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 (Stenman et al., 2008).

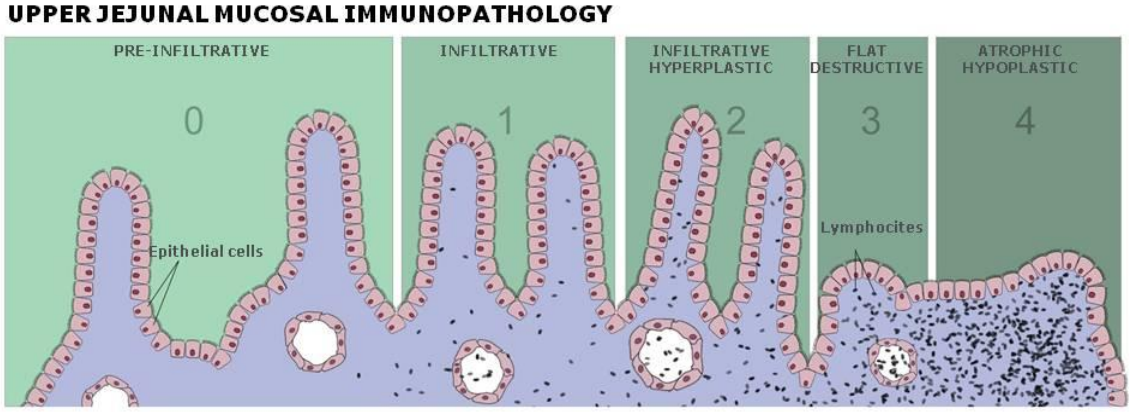


Figure 1. Gluten-induced mucosal changes in different stages of CD according to the Marsh classification. Image from www.theglutensyndrome.net.

Taking into account these pathological features, diagnostic criteria for pediatric CD are established by the European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN). Previous criteria were based on the presence of characteristic histological injuries in a biopsy of the small intestine and by positive serologic results, although the latter were not essential. Additionally, follow-up control biopsies were performed in CD patients after diagnosis (Mäki, 1995). In the diagnostic guidelines published in 2012, duodenal biopsies can be excluded in symptomatic children with IgA class TGA titers above 10 times the upper limit of normal levels and positive EMA in an independent sample. Additionally, Human Leucocyte Antigen (HLA) genotyping

is valuable since CD is very unlikely if risk haplotypes are absent (Husby et al., 2012).

1.2. Epidemiology

Until the end of the last century, CD was considered a comparatively uncommon disorder affecting almost exclusively people of European origin, with prevalence rates of 1/1000 (Feighery, 1999). However, subsequent population studies have shown that the prevalence of CD is around 1% in Western Europe, although there are differences among populations (Dubé et al., 2005).

Classically, CD has been regarded as a disorder affecting almost exclusively people of European origin, but prevalence studies have shown a frequency ranging from 1:100 to 1:200 in unselected populations of North America and Australia (Cataldo and Montalto, 2007; de Kauwe et al., 2009). CD was also believed to be rare in Latin America (Galvao et al., 1992; Rabassa et al., 1981; Sagaro and Jimenez, 1981), North Africa (Al-Tawaty and Elbargathy, 1998; Suliman, 1978), and the Middle East (Al-hassany, 1975; Khuffash et al., 1987) where there were only limited cases and occasional observations of the disease. Additionally, CD has also been historically considered absent in the Far East (China, Japan, Korea...) (Fasano and Catassi, 2001), but recent screening studies performed in these areas have demonstrated that the prevalence of CD has been underestimated and that it is similar to that of the so-called Western countries. 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 (de Kauwe et al., 2009).

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 (Ivarsson et al., 2000). Recently available data suggest that CD incidence is truly increasing and that CD is more common in some areas than previously appreciated (Catassi et al., 2014). On the other hand, the diagnosis of adult CD has also risen dramatically in most areas of the world where data are available (Bodé and Gudmand-Høyer, 1996; Collin et al., 1997; Murray et al., 2003). Environmental risk factors with seasonal patterns, and certain viral infections (Plot and Amital, 2009) as well as changes in gluten consumption and infant feeding (Catassi et al., 2014) have been proposed as risk factors for CD.

To investigate the possible primary prevention of CD, the PreventCD project (www.preventcd.com), a European multicenter study, was initiated (Hogen Esch et al., 2010). The hypothesis of this study was that tolerance to gluten could be induced by introducing small quantities of gluten to genetically predisposed, at risk infants (preferably while they were still being breast-fed). The results indicated that neither introduction of small quantities of gluten nor breast-feeding were able to reduce the risk of celiac disease at 3 years of age (Vriezinga et al., 2014).

Regarding gender, females are more commonly affected than males. Among patients presenting the disease a female to male ratio of almost 3 to 1 has been observed during their fertile years (Feighery et al., 1998).

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 (Feighery, 1999; Di Sabatino and Corazza, 2009). A dramatic reduction in symptoms occurs within days or weeks, and often precedes normalization of serological markers and of duodenal villous atrophy (Murray et al., 2004). Inadequately 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 (Lerner, 2010).

Despite treatment effectiveness, complying with gluten free diet (GFD) is difficult for many people, and solutions to improve quality of life in CD patients are needed (Samasca et al., 2014). Additionally, a small subgroup of patients may show non-responsive or refractory CD (RCD), having persistent or recurrent symptoms, inflammation of the intestine and villous atrophy despite strict adherence to a GFD (Rubio-Tapia and Murray, 2010). Therefore, the development of a safe, effective and affordable alternative therapy is necessary.

Together with the knowledge of the pathophysiological mechanisms of CD, some novel non-dietary measures have been developed as an addition or substitution to GFD (Freeman, 2013). Many of the non-dietary approaches considered have already been studied in several clinical trials, such as larazotide acetate, a synthetic peptide that blocks paracellular permeability in cell monolayers and prevents the passage of gliadin peptides through the epithelial barrier (Paterson et al., 2007). In a placebo-controlled study, they assessed larazotide acetate 0.5, 1, or 2 mg three times daily in 342 CD patients that had been on a GFD for more than a year. They found out that larazotide acetate 0.5 mg reduced signs and symptoms in CD patients on a GFD better than a GFD alone. This therapeutic agent targeting Tight Junction regulation in patients with CD who are symptomatic despite a GFD, could represent an important therapeutic option for CD patients with persistent symptoms, although further examination is needed (Leffler et al., 2015). Recently, a therapeutic vaccine, Nexvax2, including epitopes for gluten-specific CD4+ T cells to inactivate the immune process when gluten is ingested (Goel et al., 2017), has been tested showing a good profile in a phase I trial, even though further studies are needed in order to demonstrate its efficacy to cure CD.

2. Pathogenesis of celiac disease

Advances in our understanding of the mechanisms involved 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.

It has been known for a long time that CD is a T cell-mediated disease. In summary, due to an altered transport, gluten peptides cross the epithelium into the *lamina propria* and 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- and Th17-mediated response that leads to the infiltration of the epithelial *lamina propria* by inflammatory cells, together with crypt hyperplasia, and villous atrophy (Castellanos-Rubio et al., 2009; Schuppan et al., 2009).

Many studies 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 (Hüe et al., 2004; Maiuri et al., 2003) (Figure 2). The major mediator of the innate immune response to gliadin peptides is the cytokine interleukin-15 (IL-15) which can modulate intraepithelial lymphocyte (IEL) function (Escudero-Hernández et al., 2017).

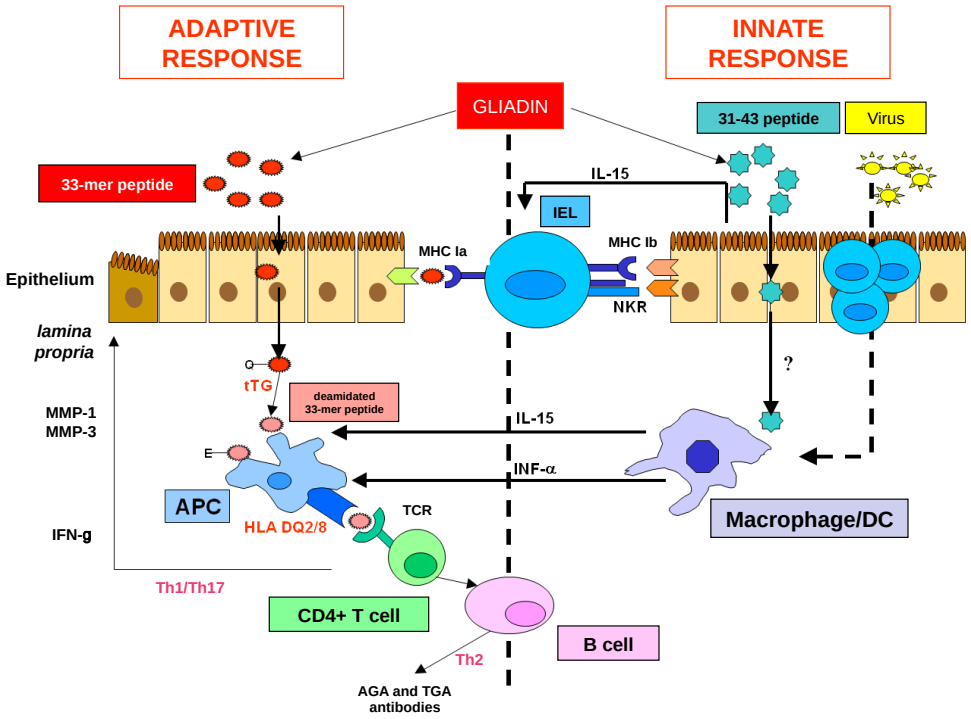


Figure 2. Pathogenic mechanisms of celiac disease: adaptive and innate immune branches. APC, antigen presenting cells; MMP, Matrix metalloproteinase; IFN, interferon; Th, T helper cell; HLA-DQ2/8, human leukocyte antigen DQ2/8; AGA, anti-gliadin antibodies; TGA, anti-transglutaminase antibodies; IEL, intraepithelial lymphocytes; IL-15, interleukin 15; NKR, Natural Killer (NK) cell receptors; DC, dendritic cell.

Many aspects of the pathogenesis still need to be clarified, especially those regarding the interaction between gluten and epithelial cells, the passage of gluten peptides into the *lamina propria*, tissue transglutaminase type 2 (TG2) activation, mechanisms that regulate IL-15 expression, and autoantibody production.

2.1. 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. The high proline content of gluten proteins makes them very resistant to degradation by

intestinal proteases so that long fragments (10-50 residues) are present in the gut lumen (Shan et al., 2002). These fragments are good substrates for the TG2 enzyme, which can deamidate gluten peptides converting certain glutamine residues to glutamate, increasing their affinity to HLA-DQ2 or HLA-DQ8 molecules and leading to a gluten-specific CD4+ Th1 cell activation that results in inflammation of the intestinal mucosa, malabsorption and other CD symptoms. In addition, gluten can also trigger CD8+ T cell mediated responses in the *lamina propria* and may expand IEL population independently of HLA presentation (Jabri and Sollid, 2009; Schuppan et al., 2009).

Different *in vitro* studies support the idea that gluten affects the innate immune system. The most studied gluten-derived peptide for its innate immune properties is a peptide spanning aminoacids 31-43 of the alpha gliadin molecule. This peptide does not induce T cell-specific responses, but it induces an innate response in tissues from patients with CD and a substantial increase in epithelial apoptosis (Maiuri et al., 2003). More recently, an 8-mer gliadin peptide that once deamidated is able to induce a humoral immune response *in vivo* in CD patients has been identified, and is an antigen for specific IgA antibodies in CD patients (Vallejo-Diez et al., 2013).

2.2. Transglutaminase

TG2 is a ubiquitously expressed multifunctional protein which is usually active in the extracellular space and catalyzes the covalent and irreversible cross-linking of a protein with a glutamine residue to a second protein with a lysine residue (Folk and Cole, 1966; Folk and Chung, 1985). It plays a central role in the activation of the adaptive immune response since it induces the enzymatic modification of gliadin peptides and increases their affinity to HLA-DQ2 and HLA-DQ8 molecules (Arentz-Hansen et al., 2000). Gluten is rich in prolines and glutamines, and contains very few negative residues necessary to bind to the groove of HLA-DQ2 or HLA-DQ8. TG2 deamidation of specific gliadin peptides

provokes an increase of negative charges, favoring their binding to HLA molecules HLA-DQ2 and HLA-DQ8 and triggering the presentation of these peptides to CD4+ T cells (Molberg et al., 1998; Ráki et al., 2007; Wal et al., 1998). Hence, TG2 transforms gliadin from a non-stimulatory molecule to an efficient T-cell antigen capable of evoking a massive secretion of local cytokines, leading to alterations in enterocyte differentiation and proliferation.

Additionally, TG2-crosslinking between gliadin peptides and the enzyme leads to the formation of TG2-gliadin complexes that trigger the production of IgA-class autoantibodies against TG2, or TGA (Caputo et al., 2009). Many studies have highlighted a possible pathogenetic role of TGA, since effects on cell cycle, apoptosis, angiogenesis and intestinal permeability have been reported, suggesting that they could be pathologically relevant (Caputo et al., 2009; Lindfors et al., 2009). Many studies have pointed out a possible pathogenic role of anti-TG2 antibodies, since they modulate TG2 activity. Autoantibodies also alter the uptake of the alpha-gliadin peptide 31-43, responsible of the innate immune response in CD, protecting cells from p31-43 damaging effects in an intestinal cell line. In a work carried out recently, they investigated whether anti-TG2 antibodies protect cells from p31-43-induced damage in a CD model consisting of primary dermal fibroblasts. They found out that the antibodies reduced the uptake of p31-43 by fibroblasts derived from healthy individuals but not in those derived from CD patients, pointing out to a loose of a protective role in the disease (Paolella et al., 2017).

2.3. Adaptive immunity

Adaptive immunity includes T cell-mediated and humoral immunity, and both of them 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 (Lundin et al., 1990; Mazzarella et al., 2003). Gliadin-specific T lymphocytes from celiac mucosa are mainly of the Th1 phenotype and release prevalently proinflammatory cytokines, dominated by IFN-γ (Nilsen et al., 1995; Troncone et al., 1998). In addition to IFN-γ, other Th1-inducing cytokines such as interleukin 18 and IFN-α are also increased (León et al., 2006; Monteleone et al., 2001; Steinman, 2007). A different linage of CD4+ T-helper cells that differentiate in the presence of IL-6 and TGFβ, and produce interleukin 17 cytokine-family members (Th17 lymphocytes) is responsible for pathogenic effects previously attributed to the IL-12/INFγ network (Castellanos-Rubio et al., 2009). Both Th1 and Th17 responses are present in the active CD lesion, a phenomenon that has also been described in other immune-mediated conditions (Harris et al., 2010; Monteleone et al., 2010; Sjöström et al., 1998).

2.4. 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 HLA-DQ8.

Several *in vivo* challenge studies have demonstrated that peptide 31-43 from α-gliadin is capable of inducing disease symptoms, and several changes characteristic of CD have been observed in biopsy cultures (Maiuri et al., 1996; Picarelli et al., 1999; Sturgess et al., 1994). This peptide does not appear to stimulate a T cell-mediated response (Anderson et al., 2000; Arentz-Hansen et al., 2000), so it is likely that the toxicity of peptide 31-43 is based on its capacity of activating the innate immune response. Recently, it has been observed that this peptide activates IFN-α in the intestine of CD patients and in the Caco-2 cell line in cooperation with loxoribine (LOX) (a Toll-like receptor 7, TLR7,-specific

viral ligand) by activating the TLR7 pathway and interfering with endocytic trafficking, suggesting that viral infections and alimentary proteins are able to potentiate the innate immune response (Nanayakkara et al., 2018).

Several studies have implicated MyD88, the major signal transducer of TLR4 on monocytes, macrophages and dendritic cells, as well as TLR4 itself as the primary receptor for innate responses to cereal proteins (Schuppan et al., 2009). Innate immune activation of IELs by gluten induces expression of *MICA* on the intestinal epithelium, a ligand for the NKG2D receptor on natural killer, γδT cells and subsets of CD4+ and CD8+ T cells. Epithelial MICA, together with upregulated IL-15, leads to the activation of NKG2D on IELs, triggering antigen-specific, lymphocyte-mediated cytotoxicity. Subsequently, innate cytotoxic and cytokine production responses are activated in the initial stages of the disease, linking innate and adaptive immunity. Finally, IL-21 is an additional driving force of innate immunity that often acts in concert with IL-15 (Fina et al., 2008; Hüe et al., 2004; Sarra et al., 2013).

2.5. Other biological pathways

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

In the last years, several whole genome expression analyses have been performed in different tissues and include microarrays and RNA sequencing (RNA-seq) approaches. These studies have been performed with the aim of elucidating CD-

related alterations in the expression patterns of single genes, groups of coregulated genes and pathways (Castellanos-Rubio and Bilbao, 2017).

The first expression microarray study in CD (Juuti-Uusitalo et al., 2004) in jejunal biopsies from active CD patients, patients on GFD, and non-celiac controls identified genes related to T-cell activation, B-cell maturation and epithelial cell differentiations. Genes with an altered expression in both active CD and GFD patients (constitutively altered genes) were classified as potentially important in the etiology of CD. More recently, it has been shown that constitutively altered genes normally belong to the core of biological pathways, while those that are altered only in the active disease are located more peripherally (Fernandez-Jimenez et al., 2014).

In a second microarray study (Diosdado et al., 2004) gene expression was studied in duodenal biopsies from 15 CD patients (active CD and GFD CD) and 7 non-celiac controls. On the one hand, they studied intestinal damage regardless of gluten ingestion, and found an increment of T cells and macrophages, a more intense Th1 response and an enhanced cell proliferation in CD. On the other hand, they studied the effect of gluten, and genes related to cell cycle and cell division were detected.

A third microarray study was carried out in intestinal biopsies of CD patients (Castellanos-Rubio et al., 2010). In this third experiment, the acute and long-term effects of gliadin exposure on duodenal mucosa were studied. The analyses found out that cell cycle, apoptosis, extracellular matrix and cellular communication pathways are altered in both types of responses to gluten. Dysregulation of more complex signaling pathways related to TGF-β, Jak-Stat, and NFκB was mainly observed in the long-term response experiment, revealing their involvement in the perpetuation of the disease process rather than in its origin.

In order to find genes that participate in epithelial proliferation and differentiation and could be related to CD, microarray analyses were carried out in a TGF-β1-induced T84 epithelial cell differentiation model (Juuti-Uusitalo et al., 2007). Several genes that are members of the epidermal growth factor (EGF)-mediated signaling pathway were identified. Another microarray study was performed in intestinal Caco-2 cells exposed to gliadin in order to identify gliadin target genes (Parmar et al., 2013). Even though many genes were identified, they were not able to conclude that the observed effects were gliadin-specific. Additionally, *in vitro* models must always be handled with caution. Another study on epithelial cells isolated from intestinal biopsies of CD patients and controls detected gene expression changes in the proliferation, cell death, and differentiation pathways, among others (Bracken et al., 2008).

RNA-seq is a powerful alternative to microarrays, with lower background signals and higher sensitivity. In addition, RNA-seq allows the identification of novel transcribed regions that could be important in disease pathogenesis (Wang et al., 2009). To date, one RNA-seq study has been published in CD (Quinn et al., 2015). In this study, the transcriptome of CD4+ T lymphocytes derived from peripheral blood mononuclear cells of CD patients and controls was studied in basal conditions and in response to anti-CD3/CD28 antibody and phorbol 12-myristate 13-acetate (PMA) stimulations. The authors found no altered pathway in the unstimulated cell group, while in the antibody-stimulated group the cytokine–cytokine receptor interaction pathway was enriched. The PMA-stimulated samples showed enrichment of immune-related genes. Altered genes from the antibody-stimulation were grouped into co-expression modules. They showed an overrepresentation of immune-related pathways, namely the TGF-β receptor pathway, as well as enrichment on genes that had previously been associated with genetic risk. In particular, the candidate gene *BACH2* (Dubois et al., 2010) was downregulated under all the conditions tested, supporting previous evidences of its important role in the regulation of T cell differentiation and autoimmune disease prevention.

In summary, although the results from the different experiments are difficult to reconcile, the main findings point to an intensification of the immune response, as well as dysregulation of signaling and cell cycle pathways. These studies have provided numerous altered genes and pathways, so that we are beginning to understand the complexity of the interactions among the environmental trigger, the genetic polymorphisms, and the gene expression. The identification of the key genomic players in those events could contribute to identify potential therapeutic targets of this and other autoimmune disorders.

3. Genetics of celiac disease

Although the precise inheritance model of CD is still not completely understood, 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, especially from 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 between mono- and di-zygotic twins has allowed scientists to estimate the genetic component of CD, which is higher than what has been discovered for other immunological complex diseases, such as type 1 diabetes (T1D) (around 30% concordance in monozygotic and 6% in dizygotic twins) (Sollid and Thorsby, 1993). 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. In summary, accumulated evidence

suggests that CD has a very strong genetic component and it has been calculated that the heritability of this disease (proportion of the risk of suffering from CD attributable to genetic factors, compared to environmental determinants) is around 87% (Dieli-Crimi et al., 2015; Greco et al., 2002; Nistico et al., 2006).

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 extended MHC region contribute an additional 18% of genetic variance (Gutierrez-Achury et al., 2015), and altogether explain approximately 40% of CD risk. Consequently, additional small effects conferred by non-HLA susceptibility *loci* must exist.

In the last years a great effort has been done in order to identify additional genetic susceptibility determinants in CD. Approaches including genetic linkage studies, candidate gene association studies and, in the last decade, genome-wide association (GWA) and follow-up studies have been performed. 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.

3.1. Contribution of HLA region

HLA is the name for the MHC in humans; it is a super *locus* located on chromosomal region 6p21, and contains a large number of genes involved in the immune response. This region 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 in distinguishing 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 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 (Horton et al., 2004).

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 (Ludwig et al., 1973). Subsequent molecular studies have revealed that HLA class II genes encoding both HLA-DQ2 and HLA-DQ8 molecules are the directly implicated factors (Sollid, 1989; Spurkland et al., 1992) (Figure 3). HLA-DQ2 and HLA-DQ8 variants are in LD with HLA-DR3 and HLA-DR4, respectively. Thus, we often refer to these risk variants as HLA DR3-DQ2 and HLA DR4-DQ8 haplotypes (Sollid, 1989).

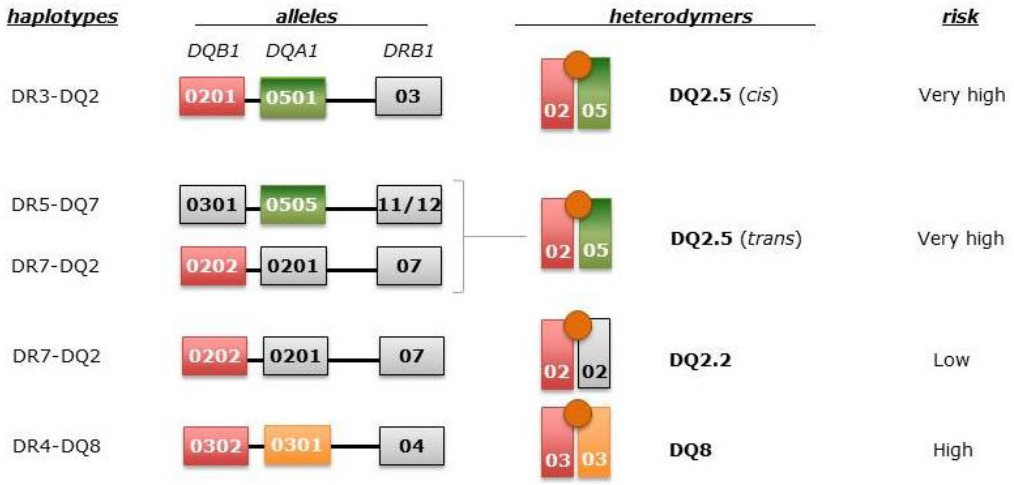


Figure 3. HLA associations in CD. HLA-DQ2 molecule is the major factor conferring risk to CD. The majority of CD patients express the heterodimer HLA-DQ2.5, encoded by alleles HLA-DQA1*05 (α chain) and HLA-DQB1*02 (β chain). These two alleles can be present in *cis* on the DR3-DQ2 haplotype or in *trans* on DR5-DQ7 and DR7-DQ2.2 heterozygous individuals. The HLA-DQ2.2 dimer, a variant of HLA-DQ2 encoded by alleles HLA-DQA1*02:01 and HLA-DQB1*02:02, confer a low risk to develop the disease. Most of DQ2-negative patients express HLA-DQ8, encoded by the DR4-DQ8 haplotype. Adapted from Abadie et al., 2011.

The primary function of DQ heterodimers is to present exogenous peptide antigens 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. As mentioned before, HLA-DQ2 and HLA-DQ8 molecules are the primary genetic factors associated with CD, however, HLA-DQ2 is more strongly associated with CD than HLA-DQ8 (Louka and Sollid, 2003).

The HLA-DQ2.5 heterodimer, formed by the combination of the products of DQA1*05 and DQB1*02 alleles, is present in 90% of celiac patients. The α and β chains of the heterodimer can be encoded in *cis* but they may also be encoded in *trans*. The differences between these two types of HLA-DQ2.5 dimers affect a single amino acid of the DQα chain (DQA1*05:01 *vs*. DQA1*05:05) and another residue of the membrane region of the DQβ chain (DQB1*02:01 *vs.* DQB1*02:02), although they seem not to have any functional consequences and they are associated with a similar risk effect. However, the risk conferred by another HLA-DQ2 variant, the HLA-DQ2.2 dimer, is very low (Sollid, 2002; Sollid and Thorsby, 1993).

Most of the patients that are 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 (Mäki and Collin, 1997). A very small portion of the patients are negative for both DQ2 and DQ8, but it has been observed that these

individuals present at least one of the two alleles encoding the DQ2 molecule (DQA1*05 or DQB1*02) (Karell et al., 2003; Spurkland et al., 1992).

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, harbor the maximum genetic risk to develop CD (van Belzen et al., 2004; Lundin et al., 1993; Ploski et al., 1993). In this sense, it has to be mentioned that patients with RCD (those that do not respond to GFD) present a higher proportion 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% of the heritability of CD (Gutierrez-Achury et al., 2015).

Although HLA genes contribute greatly to the genetic susceptibility to CD, 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 (Sollid, 2002).

3.2. Contribution of Non-HLA susceptibility regions

Given the fact that HLA alone can only explain around 40% of the genetic component of CD, several studies have been performed to localize and identify

non-HLA susceptibility genes that could clarify the complex genetics of this disorder. Two have been the major strategies used for this aim: on the one hand, linkage studies in affected families, and on the other hand, association studies based on population screening.

Linkage studies in families allow the identification of chromosomal regions which are repeatedly and consistently inherited by the affected members of a family through several generations. Thus, regions potentially relevant to the development of the disease can be selected and fenced in. Genes localized in these regions are considered positional candidates, due to the fact that it is their position in the genome that is conferring them the candidate identity. 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: CELIAC2 located on chromosome 5q31-33 (Greco et al., 1998); CELIAC3 located in chromosome 2q32 and containing T lymphocyte regulatory genes *CD28*, *CTLA4* and *ICOS* (Holopainen et al., 2004); and CELIAC4 located in 19p13 containing myosin IXB gene *MYO9B* (Van Belzen et al., 2003).

More recently, high-throughput genotyping platforms have enabled GWA studies, and have changed 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 GWA studies performed in CD analyzed approximately 5,000 patients and 10,000 controls, and revealed a total of 26 non-HLA associated regions (Dubois et al., 2010; van Heel et al., 2007). One year later, roughly 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 (Trynka et al., 2011). Recently, a reanalysis of the Immunochip has identified five additional genomic regions involved in CD (Garcia-Etxebarria et al., 2016). Overall, there is a total of 44 non-HLA regions associated with CD, containing 62 independent association signals that together contribute 5-7% to the genetic risk (Figure 4). Nineteen of these regions pinpoint to a single candidate gene, but only 3 of the associated SNPs are linked to protein-altering variants located in exons, whereas some potentially causative genes have been proposed due to the existence of signals near 5' or 3' regulatory regions.

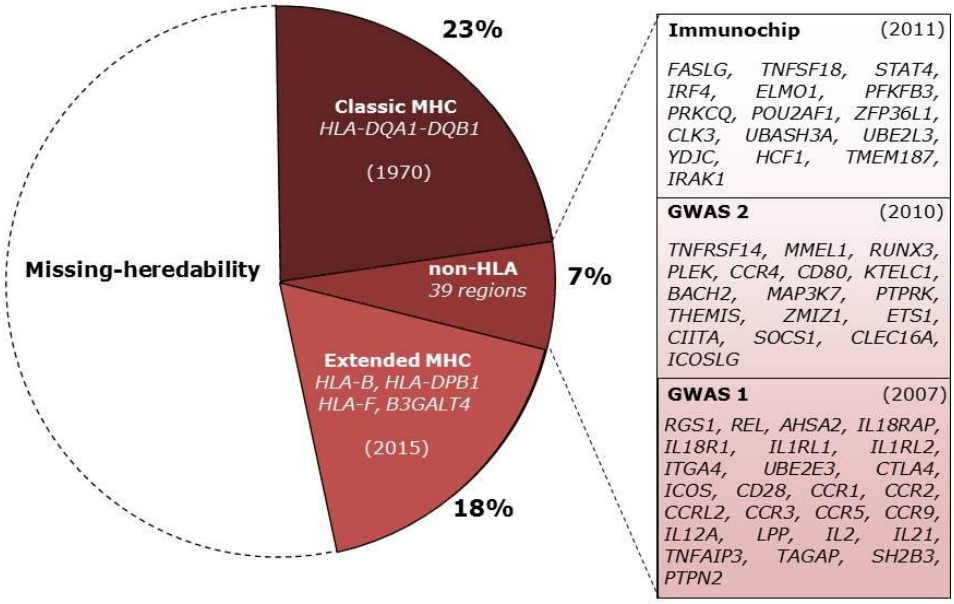


Figure 4. Progress in the Genetics of CD. After the Immunochip study 39 non-HLA *loci* have been found to contribute to the genetic risk to develop the disease. Later, five additional non-HLA *loci* with modest contribution have been linked to CD (Garcia-Etxebarria et al., 2016).

On the other hand, the implication of rare variants with 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 (Hunt et al., 2013).

Despite the success of GWA and follow-up studies in discovering CD susceptible *loci*, the variants identified only explain a small proportion of the genetic contribution to disease and many more remain to be found. Moreover, most of the GWA studies SNPs are probably not the causal variants and it is thought that they are merely pointing to associated regions due to LD, and the regions 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. An example of the regulatory nature of the associated SNPs is a long non-coding RNA (lncRNA) that is close *IL18RAP* and harbors the CD-associated SNP rs917997. This lncRNA has been shown to be a key regulator of genes in the NFκB pathway. Lnc13 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 risk allele binds the protein complex less efficiently causing an increase in the expression of inflammatory genes, which in turn will predispose to disease development (Castellanos-Rubio et al., 2016).

Most of the approaches made to understand the functional effects underlying the association signals have been limited to the search of possible alterations in the expression levels of nearby genes, or *cis*-expression quantitative trait *loci* (*cis*-eQTLs). Results have been limited suggesting a more complex scenario in CD pathogenesis where the coordinated response of groups of genes or pathways is an additional functional layer that could be altered (Plaza-Izurieta et al., 2015).

In this context, our group has shown that the expression of *PTPRK* and *THEMIS*, two immune-related genes located on the GWA study peak on hg19 chr6:128.31-128.70 Mb, was positively correlated in CD patients but not in controls (Bondar et al., 2014). In another experiment, a group of 93 genes related to the NFκB pathway showed a strong correlation mRNA levels in the intestinal mucosa of

non-celiac controls that was completely disrupted in active CD (Fernandez-Jimenez et al., 2014). These findings indicate that co-expression, correlation of expression of two or more genes, is an affected layer in the disease, and regulatory elements such as transcription factors (TFs), non-coding RNAs (ncRNAs), chromatin structure, DNA methylation, etc. could control the expression of groups of genes rather than isolated genes.

In brief, the genomics of CD remains elusive, with known genetic factors accounting for approximately half of its heritability. 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 the deregulation is a very complicated task, which is still far from being fully resolved.

4. Gene regulation

Gene regulation is a complex process that includes all those elements that affect the production level of specific gene products (mRNA or protein), namely enhancers and promoters, TFs and their binding sites, regulatory RNAs such as lncRNAs or microRNAs (miRNAs), methylation of genomic DNA, histone modifications, etc. Alterations in regulation can cause a broad range of diseases such as cancer, autoimmunity, neurological disorders, developmental syndromes, diabetes, cardiovascular disease or obesity (Lee and Young, 2013).

Moreover, it has been observed that individual genes do not work alone, but interact with each other creating groups, gene-networks or pathways; and those interactions could affect human health. Each gene is estimated to interact with an average of 4-8 other genes (Arnone and Davidson, 1997) and to be involved in 10 biological functions (Miklos and Rubin, 1996). Hence, coordinated response of groups or pathways of genes could be an additional functional layer in

complex diseases, and regulation of groups of genes should be also studied (Li et al., 2018).

4.1. Transcription factors

TFs are regulatory proteins that control gene expression catalyzed by RNA polymerase II (Fulton et al., 2009; Vaquerizas et al., 2009) through their binding to a specific DNA sequence: their activity determines cell function and response to environment. One TF is able to regulate different genes in different cell types (Geertz et al., 2012), and they allow unique expression of a gene in different cell types.

TFs have DNA-binding domains that allow them bind to specific DNA sequences and activate or inhibit transcription. Apart from the DNA binding domain, they also have an activation/repression domain that interacts with cofactors, protein complexes that contribute to activation (coactivators) and repression (corepressors). TF cofactors include protein complexes that can either activate RNA polymerase II or modify local chromatin structure, controlling transcription rate (Bhagwat and Vakoc, 2015). TFs have traditionally been classified as activators or repressors, but many of them are able to recruit both coactivators and corepressors, questioning this classification (Frietze and Farnham, 2011; Rosenfeld et al., 2006; Schmitges et al., 2016).

Each TF recognizes a collection of similar DNA sequences, which can be represented as binding site motifs. Motifs are short sequences recognized by a given TF, which can be used to investigate longer sequences to finally identify potential binding sites. Over the last decade, knowledge in motifs and genomic binding sites has improved, leading to TF-DNA interactions data. In the last years motif collections such as TRANSFAC (Matys et al., 2006), JASPAR (Mathelier et al., 2016), HT-SELEX (Jolma et al., 2013, 2015; Yin et al., 2017), UniPROBE (Hume et al., 2015), and CisBP (Weirauch et al., 2014), along with

previous catalogs of human TFs (Fulton et al., 2009; Vaquerizas et al., 2009; Wingender et al., 2015) have been created. Sequence context, including flanking sequences and DNA shape, as well as interacting cofactors and TFs, can alter sequence recognition (Siggers and Gordân, 2014). These features together with differential TF expression and chromatin accessibility determine condition-specific TF binding and ultimately gene regulation (Wang et al., 2012).

Some TFs bind to DNA promoter sequences in order to help the formation of the transcription initiation complex, while others bind to regulatory sequences as enhancers and stimulate or repress transcription. Regulatory sequences can be thousands of base pairs upstream or downstream from the gene which is being transcribed.

TFs represent 8% of all human genes, and they have been associated with a wide range of diseases and phenotypes. Mutations in TF genes are often highly deleterious, which could explain the high conservation observed in the *loci* encoding TFs (Bejerano et al., 2004). Different studies have identified them as drivers of many disease processes such as breast cancer where *AGTR2*, *ZNF132*, *TFDP3* and others have been identified as regulators (Tovar et al., 2015), or human periodontitis where 41 master regulators have been found (Sawle et al., 2016). A number of TFs have been implicated in CD; including the master regulator of T cell differentiation *BACH2*, which is downregulated in CD, while 98 of its targets have altered expression (Quinn et al., 2015). However, the contribution of other TFs to CD pathogenesis is still unknown.

4.2. microRNAs

miRNAs are short (18-22 nucleotides long) non-coding RNAs (ncRNAs) that regulate gene expression at the post-transcriptional level targeting mRNAs. These small ncRNAs are predicted to regulate the translation of up to 60% of

protein-coding genes (Esteller, 2011) and they are able to regulate TF activity, affecting several intracellular pathways.

The genes codifying for miRNAs are transcribed as long primary transcripts (pri-miRNAs). pri-miRNA transcripts are encoded on exons (13-20%) and introns (80-87%) of protein-coding and non-coding genes (Rodriguez et al., 2004). pri-miRNAs are transcribed by RNA polymerase II and III (Borchert et al., 2006) and are 70 to 100 nucleotides long. After the pri-miRNA has been generated in the nucleus, it is cleaved into one or several precursor-miRNAs (pre-miRNAs) (around 70 nucleotides long) by the enzymes Drosha and DGCR8. Exportin-5 binds the pre-miRNA and translocates it from the nucleus to the cytoplasm (Sun et al., 2010). Once the pre-miRNA has been exported to the cytoplasm, the RNase III enzyme Dicer cleaves the pre-miRNA into a ~20 bp miRNA/miRNA* duplex. The miRNA duplex could give rise to two different mature miRNAs, but usually only one strand is incorporated into the RNA-induced silencing complex (RISC) and serves as a functional, mature miRNA that binds to complementary sequences in the 3' UTR of target mRNAs (Figure 5). If a perfect pairing occurs, degradation of the target mRNA will happen, whereas if there is a mismatch, translational inhibition without the reduction of the mRNA level will occur (Dalmay, 2008). Consequently, changes in miRNA levels will result in downregulation or upregulation of their targets.

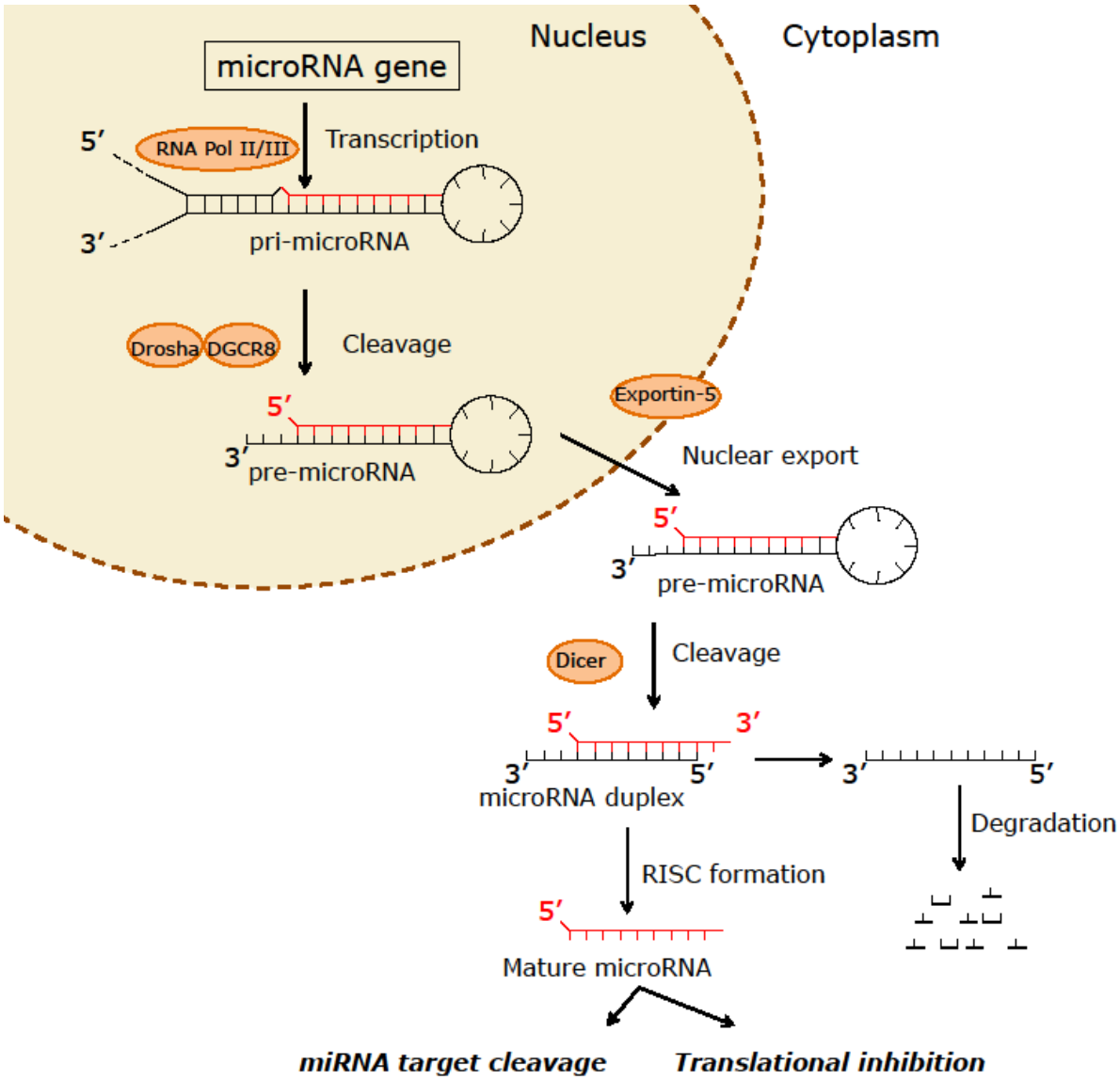


Figure 5. Canonical pathway of miRNA processing. Adapted from Winter et al., 2009.

miRNAs have been shown to be involved in several diseases such as cancer, neurological disorders, cardiovascular disorders or inflammatory diseases (Esteller, 2011). Alterations in miRNA regulation have also been related to the development of immune dysfunctions and autoimmunity. Several studies have focused on the role of miRNAs in autoimmune diseases and different expression profiles have been identified as biomarkers of certain autoimmune conditions such as systemic lupus erythematosus, rheumatoid arthritis and Sjögren's syndrome (Chen et al., 2016). In CD, only a few studies have been performed, and have shown that the expression of specific miRNAs is altered in the disease (Buoli Comani et al., 2015; Capuano et al., 2011; Magni et al., 2014; Vaira et al.,

2014), suggesting a role in disease pathogenesis and a potential use for the diagnosis of CD (Felli et al., 2017).

4.3. Chromatin structure

To understand the functioning of the human genome, it is essential to understand the three-dimensional (3D) folding and spatial organization of chromosomes in the cell nucleus, or the 3D genome, since regulation of gene expression is closely related to organization of chromosomes (Lupiáñez et al., 2016; Sexton et al., 2007).

At the nuclear level, heterochromatin (inactive chromatin) tends to be near the periphery, while euchromatin (active chromatin) is localized more internally (Bickmore, 2013). Regarding chromosomes, these are divided into large compartments containing active and open or inactive and closed chromatin (A/B compartments) (Simonis et al., 2006). Active and open compartments interact with other active and open compartments, as closed compartments do so with closed compartments. These compartments change depending on cell type, since different cells express different gene sets. However, the understanding of chromosome organization at a sub-megabase level has been limited.

Recently, Topologically Associating Domains (TADs) have been described. The A/B compartments mentioned above are composed of TADs, basic structural and functional units of chromatin (Dekker and Heard, 2015; Dixon et al., 2016). TADs represent spatial domains of high-frequency chromatin interactions (Dixon et al., 2012). These chromatin domains are approximately 1 Mb long, and are largely stable across cell types and conserved among species. TADs are enriched for chromatin-chromatin interactions, while interactions between neighboring domains are low.

Interactions between chromatin regions are thought to be important in gene regulation. In particular, TADs comprise the majority of characterized enhancer-promoter pairs (Rao et al., 2014). Those enhancers, located in large regulatory domains, exert their effects and lead to gene coregulation inside TAD domains (Le Dily et al., 2014).

Three-dimensional chromatin structure is maintained by multiple factors (Phillips-Cremins et al., 2013), but CCCTC-binding factor protein (CTCF) and the cohesin complex are the main ones. CTCF is the main insulator protein described in vertebrates, and defines chromatin boundaries and contributes to chromatin looping (Ong and Corces, 2014). However, only 15% of the CTCF binding sites in the genome are present at TAD borders, while 85% are inside TADs (Dixon et al., 2012). The function of these CTCF may be to direct enhancers within the TAD to the appropriate gene promoter (Figure 6).

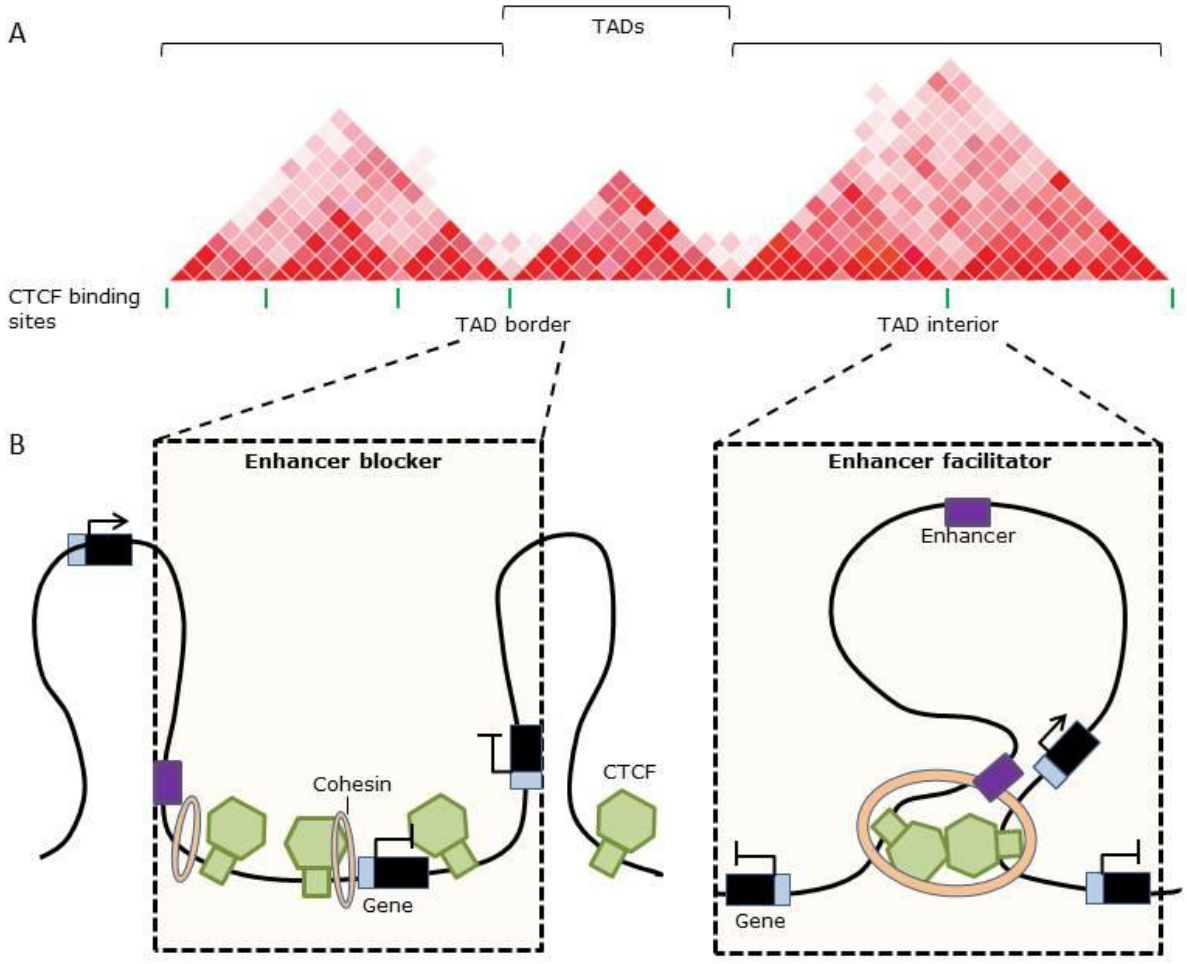


Figure 6. Characterization of CTCF-mediated loops. A) Schematic data of TADs and their borders. B) Representation of how CTCF binding sites at TAD borders contribute to the establishment of the border, and CTCF proteins act as enhancer blockers. On the other hand, CTCF binding sites within TADs facilitate enhancer-promoter looping. The blue box indicates the promoter of the gene. Adapted from Ong and Corces, 2014.

Recently, a link between the disruption of the 3D genome and disease has been observed. In human diseases, these derangements are caused by mutations, SNPs, small insertions or deletions (indels), and chromosomal abnormalities affecting 3D genome organizers. In most cases, it is hard to distinguish whether a disease-associated 3D genome derangement is the cause or consequence of disease development, but aberrant genome architecture could cause the dysregulation of genome function and lead to disease phenotypes (Krumm and Duan, 2018).

At least two mechanisms are known by which alterations of the expression of disease-related genes occur through the disruption of the 3D genome, and consequently lead to disease phenotypes. In the first mechanism, disease-associated regulatory SNPs alter promoter-enhancer interactions leading to aberrant gene expression and disease. Many studies have demonstrated that this mechanism is involved in several human diseases, such as coronary artery disease (Mumbach et al., 2017), inflammatory bowel disease (Meddens et al., 2016), and autoimmune diseases like rheumatoid arthritis, type 1 diabetes, psoriatic arthritis and juvenile idiopathic arthritis (Martin et al., 2015). These studies have demonstrated that disease-associated regulatory SNPs can alter target gene expression by regulating chromatin interactions between genes and regulatory elements.

In the second mechanism, disruption of TAD boundaries can lead to the fusion of two flanking TADs, affecting the communication between enhancers and genes that are usually insulated from each other. This mechanism, called enhancer adoption or hijacking, allows distal enhancers to ectopically activate disease-relevant genes. Indeed, a variety of pathogenic events leading to enhancer

adoption through TAD disruption have been observed (Lupiáñez et al., 2016; Valton and Dekker, 2016). Genetic disruption or epigenetic inactivation of the architectural protein binding sites (i.e., CTCF binding sites) in a TAD boundary can cause the fusion of two flanking TADs (Flavahan et al., 2016; Lupiáñez et al., 2015). Deletion of an entire TAD boundary will also affect TAD organization (Giorgio et al., 2014). Genomic rearrangements such as inversions, deletions, duplications and translocations are able to break TADs and cause the fusion or formation of new TADs (Lupiáñez et al., 2015; Weischenfeldt et al., 2017).

Given the role of TADs in gene regulation, the study of the formation of TADs and how their boundaries are involved in looping interactions could be useful for the understanding of how genes are regulated in physiological situations, and how they are dysregulated in disease.

4.4. DNA methylation

DNA methylation is a heritable epigenetic mark in which a methyl group is transferred by DNA methyltransferases (DNMTs) to the C-5 position of the cytosine ring of DNA (Robertson, 2005). More than 98% of DNA methylation occurs in CpG dinucleotides in somatic cells (Lister et al., 2009).

DNA methylation is known to be associated with gene transcription repression by interfering with DNA-binding proteins, such as TFs (Bird, 2002). However, the analysis of genome-wide methylation patterns has revealed that DNA methylation is not always negatively correlated with gene expression. Challenging the traditional view that DNA methylation represses gene expression, it has been revealed that DNA methylation sites can also be positively correlated with gene expression (Irizarry et al., 2009).

DNA methylation is crucial for normal development, and it plays a very important role in many key processes such as genomic imprinting, X-chromosome inactivation, and suppression of repetitive element transcription and transposition. When this epigenetic information is not properly established it can lead to diseases like cancer (Delpu et al., 2013; Robertson, 2005), in which loss of methylation is a frequent event, and correlates with disease severity and metastatic potential in many tumor types (Widschwendter et al., 2004; Zhang et al., 2015).

Methylation is also known to be critical in the correct function of immune cells. Recent studies suggest that aberrant DNA methylation levels are related to autoimmunity. Environmental factors and genetic polymorphisms can cause aberrant methylation patterns, which may affect transcription levels of certain immune-related genes (Sun et al., 2016).

Regarding CD, a candidate-gene methylation analysis in duodenal biopsies of active, treated patients and non-celiac controls was able to detect some changes in promoters of several NFκB-related genes (Fernandez-Jimenez et al., 2014), suggesting that methylation could participate in CD pathogenesis. In that study, only a partial reversion of the trend seen in the active disease status was observed in GFD patients, indicating an altered methylation reversion upon diet changes.

Aims of the study

The present work has four main objectives that aim to contribute to our understanding of the implication of genetic and epigenetic regulators in CD pathogenesis:

1. To determine whether the acute and long-term exposures to gliadin can modify genome-wide co-expression patterns in the small intestine of CD patients.

2. To identify the regulatory elements underlying the differential gene co-expression observed in CD, particularly TFs and miRNAs, and to perform biological validation experiments of a subset of the candidates.

    a. To analyze the expression of candidate TFs and miRNAs and their targets in duodenal biopsies of CD patients and controls.

    b. To study the nuclear translocation and promoter binding of candidate TFs in response to gliadin in a cellular model.

3. To search for potential TAD boundary alterations that could affect CD pathogenesis using *in silico* approaches and publicly available genomic datasets, and to construct stable cellular models recapitulating those TAD merges and disruptions, in order to confirm whether they are involved in co-expression changes.

4. To detect changes in DNA methylation in candidate genomic regions in the acute response to gliadin and to confirm the involvement of the methylation of *HLA-B* and *TAP1 loci* in CD.

Material and methods

1. Material

1.1. Subjects

1.1.1. Ethical approval

All the studies performed are part of Research Projects PI13/01201 and PI16/00258 (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, and were approved by the Cruces University Hospital and Basque Autonomous Clinical Trials and Ethics Committees (codes CEIC-E13/20, CEIC-E16/46 and PI2013072, respectively). All samples were collected during routine diagnosis endoscopy, after informed consent 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 ESPGHAN criteria in force at the time of recruitment (Husby et al., 2012), including determination of antibodies against gliadin and EMA or TGA as well as a confirmatory small bowel biopsy. For the present study, a total of 92 duodenal biopsies from celiac and non-celiac patients were collected by endoscopy (Table 1).

Patients can be classified into 3 groups, as follows:

- Active CD group (29 individuals): 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 (37 individuals): normalized CD patients following a strict GFD for more than two years (asymptomatic, antibody-negative and with a recovered intestinal epithelium).
- Control group (26 individuals): non-celiac individuals without intestinal inflammation at the time of endoscopy.

Table 1. Clinical, immunological and HLA information of the celiac patients included in the study. Dx, Diagnosis; TGA, anti-tissue transglutaminase; HLA, human leukocyte antigen; AGA, anti-gliadin antibodies;

Characteristics	**Values**
Gender	**Female:** 73%; **Male:** 27%
Age at Dx (months)	31.7 ± 22.9
TGA at Dx	91.66%*
MARSH score at Dx	**3c:** 81.25%; **3b:** 18.75%
HLA	**DQ2:** 82.6%; **DQ8:** 4,4%; **DQ2/DQ8:** 13%

*One patient < 1 year positive for AGA
*Two patients positive for tTG IgG

1.2. Cell lines and cell culture

Human epithelial cell lines were obtained from the American Type Culture Collection (ATTC, Manassas, VA, USA), namely colon-derived C2BBe1 [sub-clone of Caco-2] (cat. no. CRL-2102), HCT116 (cat. no. CCL-247), and HCT15 (cat. no. CCL-225) cell lines, and embryonic kidney cell line HEK293FT (cat. no. CRL-1573).

Human epithelial cells were cultured at 37°C in a 5% CO_2 humidified atmosphere according to standard mammalian tissue culture protocols and using the recommended culture medium for each cell line (Table 2). DMEM and DMEM F-12 were purchased from Lonza (Lonza group, Basel, Switzerland; cat. nos. BE12-604F and BE12-719F respectively) and RPMI Medium 1640 (1X) from Gibco (Gibco, Carlsbad, CA, USA; cat. no. 21875-034). Media were supplemented with penicillin-streptomycin (Lonza group; cat. no. DE17-602E),

non-essential amino acids (NEAA) (Lonza group; cat. no. BE13-114E) and heat-inactivated fetal bovine serum (FBS) (Biochrom, Cambridge, UK; cat. no. S 0415).

Table 2. Complete media for each cell line. DMEM, Dulbecco's Modified Eagle's Medium; RPMI, Roswell Park Memorial Institute; 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
HCT116	DMEM F-12	10%	1%	-
HCT15	RPMI Medium 1640	10%	1%	-
HEK293FT	DMEM F-12	10%	1%	-

Culture medium was changed every 2-3 days and cells were detached using 0.25% trypsin-EDTA (Gibco; cat. no. 25200056) and sub-cultured after reaching 70–80% confluence. For cryopreservation, cells were resuspended in growth medium containing 5% DMSO (Sigma-Aldrich, St. Louis, MO, USA; cat. no. 67-68-5), frozen at -1°C/minute in an isopropanol-filled freezing container placed in a -80°C freezer for 24 hours and stored in liquid nitrogen afterwards. To thaw the frozen cells, vials were placed in a water bath at 37°C before resuspending them in growth medium and transferring them into culture flasks.

1.3. Stimulation of cell culture and biopsies

For *in vitro* stimulation of cultured cells and biopsies, pepsin-trypsin digest of gliadin (PT-G) was prepared as described previously (Bondar et al., 2014). Briefly, 500 mg of gliadin (Sigma-Aldrich; cat. no. G3375) were dissolved in 5 ml 0.2N HCl and incubated with 5 mg of pepsin (Sigma-Aldrich; cat. no. P6887) with continuous agitation at 37°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; cat. no. T9935) at 37°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 2,000 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°C. An enzymatic digest of BSA (pepsin-trypsin digested bovine serum albumin; PT-BSA) (Thermo Fisher Scientific; cat. no. SH30574.03) was prepared in the same way and used as a control.

C2BBe1 cells were incubated at 37°C in 5% of CO_2, and challenged with medium containing either 1mg/ml PT-G or PT-BSA for 4 h. For the stimulation of biopsies, four tissue portions were taken from each of the 7 patients on GFD and 8 non-CD controls studied. Two portions from each individual were incubated in 150 μl RPMI-1640 10X medium at 37°C and 5% CO_2 during 4 h, with (challenged) or without (unchallenged) 250 μg/ml PT-G.

1.4. DNA and RNA isolation

Total DNA and RNA were isolated using the NucleoSpin Blood kit and NuceloSpin RNA kit, respectively (Macherey-Nagel, Düren, Germany; cat. no. 740955.250 and 740951.250) according to the protocol supplied by the manufacturer. Briefly, frozen biopsy samples were disrupted with disposable plastic pellet pestles. In the case of cultured cells, these were collected by centrifugation.

For DNA isolation, 25 μl of proteinase K and 200 μl of Buffer B3 were added to the samples, and these were incubated at 70ºC for 20-30 min. Ethanol (210 μl) was added to each sample and these were loaded to NucleoSpin Blood Columns. The samples were centrifuged at 11,000 g for 1 min, and washing steps were performed. DNA was eluted in H_2O.

For RNA isolation, 350 μl Buffer RA1 and 3.5 μl β-mercaptoethanol were added to the cell pellet or to the ground tissue, and the samples were loaded to

NucleoSpin Filter. Ethanol (350 µl) was added to the homogenized lysate, and this was loaded to NucleoSpin RNA Column. The samples were centrifuged at 11,000 g for 30 s and 350 µl MDB was added. The samples were centrifuged at 11,000 g for 30 s and 95 µl DNase reaction mixture was applied onto the center of the silica membrane column. After three washing steps RNA was eluted in RNase-free H_2O.

The concentration and purity of the DNA and RNA samples were determined by UV spectrophotometry on a NanoDrop 1000 apparatus (Thermo Fisher Scientific, Boston, MA, USA) and samples were stored at -80°C until use.

1.5. Data sets

For co-expression analyses, microarray data from previous Human U133 Plus 2.0 Array experiments (Affymetrix, Santa Clara, CA, USA) (Castellanos-Rubio et al., 2008) was downloaded from the EBI Array Express database (http://www.ebi.ac.uk/microarray-as/ae/):

- Long-term experiment: data from the E-MEXP-1828 experiment, consisting of nine active CD patients at diagnosis who were on a gluten-containing diet at that time, with 9 CD patients on GFD for more than 2 years were used to study the long-term exposure to gliadin.
- Acute experiment: data from the E-MEXP-1823 experiment, consisting of portions of 10 GFD-treated CD patients that had been incubated separately in 1 ml of RPMI medium, with and without the addition of 10 µg/ml gliadin for 4 h were used to study acute exposure to gliadin.

On the other hand, RNA-seq data with experiment number SRP077708 was retrieved from NCBI's Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra/). RNA-seq was performed by paired-end sequencing of 75 nucleotides was carried out in a HiScanSQ platform (Illumina, San Diego, CA, USA).

- Epithelial cell enriched fraction: epithelial cells express the epithelial cell adhesion molecule EpCAM (CD326) on their surface. Data from 10 active CD patients and 12 non-celiac controls were available.
- Immune cell enriched fraction: leukocytes harbor the CD45 antigen encoded by the *PTPRC* gene. Data from 7 active CD patients and 5 non-celiac controls were available.

For the study of DNA methylation in CD, data from a previous Illumina Infinium HumanMethylation450 microarray were retrieved from NCBI's Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/); experiment GSE84745 consited of methylation data from the epithelial and immune fractions separated from duodenal biopsies of 10 CD patients and 10 controls.

The thesis manuscript as well as the appendixes of this thesis (Appendix 1 to 9) and their legends are deposited in https://labur.eus/SvteM .

2. Methods

2.1. Whole genome co-expression in CD

2.1.1. Identification of TFs and miRNAs in co-expression changes

2.1.1.1. Co-expression analysis

Co-expression analyses were performed using microarray data. Probes from the expression arrays were translated to gene names using the manifest of the array or using BLAST (Altschul et al., 1997) of the probe sequences against the human reference genome hg19. If a gene was represented by more than one probe, the median of all probe intensities was taken as the expression level of the gene.

To prioritize the most informative genes, only those genes whose expression was more variable than the median gene were selected for co-expression analyses,

using the "varianceBasedfilter" function in the Differential Co-expression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data (DCGL) R package (Yang et al., 2013). Co-expression modules were defined using the Weighted Correlation Network Analysis (WGCNA) R package (Langfelder and Horvath, 2008). The "power" parameter was set using the best value in the "Scale Free Topology Model Fit" analysis. In addition, gliadin-induced changes in co-expression were also analyzed, considering differentially co-expressed genes (DCGs) as those genes that change significantly their relationships with the rest of the genes in their original module, when comparing active and GFD samples in the case of the long-term experiment, and gliadin-challenged and unchallenged samples in the case of the acute experiment, and *viceversa*. DCGs were identified using the DCGL R package, and those with a nominal p-value < 0.05 were considered. Finally, we also identified the co-expression modules within the DCGs themselves, using the WGCNA package. The workflow of the study is summarized in Figure 7.

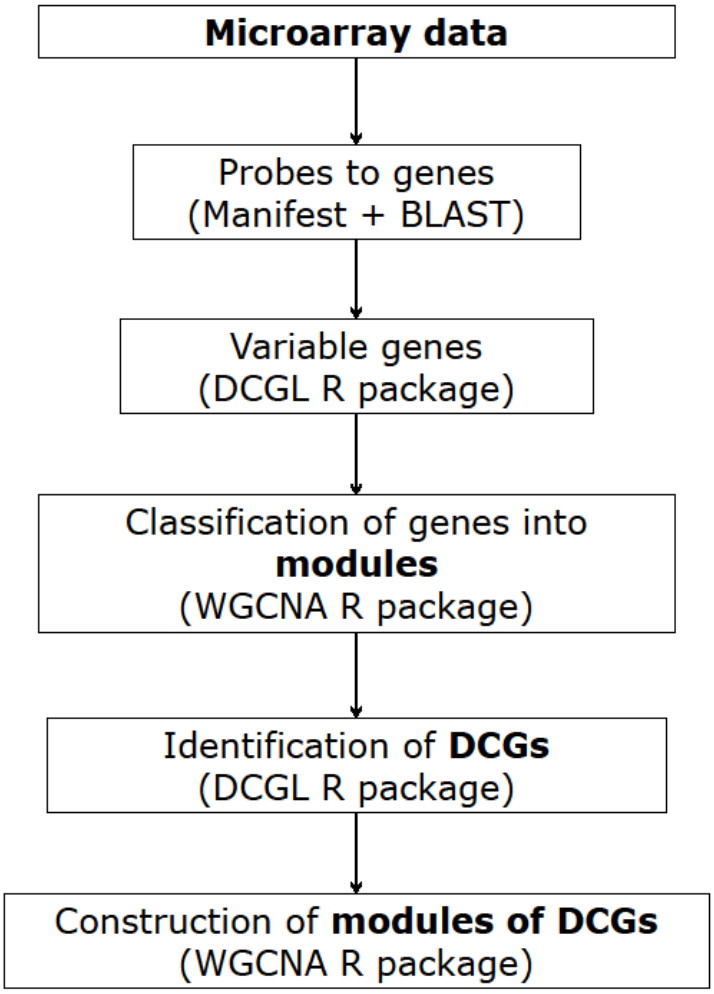


Figure 7. Flowchart of the microarray data analysis.

2.1.1.2. Selection of regulatory candidates

DCGs within each co-expression module were compared to the rest of the genome in order to detect overrepresented miRNA and transcription factor binding sites (TFBS), using the FatiGO tool available in the Babelomics v4.3 suite (http://v4.babelomics.org) (Medina et al., 2010). In general, default parameters were used ($p < 0.05$, adjusted by FDR). Particularly, in the case of miRNAs, miRBase Target database (Griffiths-Jones et al., 2006) was used to identify miRNA-target relationships. In the case of TFs, the promoter regions of the selected genes were interrogated for TF-specific motifs, according to the TRANSFAC curated TFBS database (Wingender, 2000). Groups of DCGs with

enrichment for a particular TFBS as well as their complete, original modules were compared to the rest of the genome in order to detect significantly overrepresented Gene Ontology (GO) terms (levels from 4 to 7) using the FatiGO tool.

2.1.2. Experimental confirmation of candidates

2.1.2.1. Gene expression analyses

mRNA transcription in duodenal biopsies was measured by quantitative real-time polymerase chain reaction (qRT-PCR).

2.1.2.1.1. Candidate Genes and Assays

We selected regulatory TFs and miRNAs out of the significant hits in the enrichment analysis for experimental validation, using the following criteria:

1) Repetition of the terms in at least three different modules.
2) Previous literature.

Application of these criteria resulted in 5 TFs: ELK1, NFKB1, HOXA5, CREB1 and IRF1.

In the case of miRNAs, application of the same criteria resulted in 9 miRNAs. pri-miRNAs of the candidate miRNAs were selected as a surrogate of their expression at the genic level (Table 3). Mature miRNAs of two of them were also studied.

Table 3. Candidate mature miRNAs and their primary miRNAs. Genomic coordinates (GRCh37/hg19) and miRBase accession codes are shown for stem-loop sequences.

primary miRNA	Stem-loop sequence		mature miRNA
	Location (hg19)	miRBase accession code	
hsa-mir-33a	chr22:42296948-42297016	MI0000091	hsa-miR-33a
hsa-mir-92a-1	chr13:92003568-92003645	MI0000093	hsa-miR-92a
hsa-let-7b	chr22:46509566-46509648	MI0000063	hsa-let-7b-3p
hsa-mir-503	chrX:133680358-133680428	MI0003188	hsa-miR-503
hsa-mir-655	chr14:101515887-101515983	MI0003677	hsa-miR-655
hsa-mir-17	chr13:92002859-92002942	MI0000071	hsa-miR-18a-3p
hsa-mir-26b	chr2:219267369-219267445	MI0000084	hsa-miR-26b
hsa-mir-520b	chr19:54204481-54204541	MI0003155	hsa-miR-520b
hsa-mir-520e	chr19:54178965-54179051	MI0003143	hsa-miR-520e

Target genes of the candidate TFs were selected using the following criteria according to ENCODE (https://www.encodeproject.org/) and the University of California Santa Cruz (UCSC) genome browser (https://genome.ucsc.edu/):

1) *In silico* predicted target genes from the DCG sets enriched for a particular TF, prioritizing those that showed:
    a. Target genes with conserved binding sites for the TF at the promoter.
    b. Positive chromatin immunoprecipitation-sequencing (ChIP-seq) data of the TF at the promoter.
    c. H3K27Ac-rich regions as a marker of active chromatin at the promoter.
    d. Literature on the relevance to CD pathogenesis.
2) Target genes that, although not present in the DCGs, are well defined targets of a particular TF. Targets were sorted according to the number of TFs that are controlling them (based on the human/mouse/rat conserved binding sites from the UCSC genome browser *tables* utility). Those that were under the regulation of the fewest TFs were selected in order to find the most specific ones for each candidate TF.

All the resulting genes are listed in Table 4, together with the primer and probe sets used for expression analyses. Primer and probe sets commercially available as TaqMan Assays-on-Demand (Thermo Fisher Scientific) were used for expression studies. Whenever possible, assays with probes that spanned an exon-exon junction were chosen to avoid the detection of genomic DNA.

Table 4. TaqMan assays used for TF, target genes, pri-miRNA and miRNA expression analyses. *TBP* and *RNU48* were used as housekeeping controls for protein-coding and pri-miRNA genes, and for miRNAs, respectively.

TFs	**TaqMan assay Id**	**Location (hg19)**
ELK1	Hs00901847_m1	chrX:47494920-47509887
NFKB1	Hs00765730_m1	chr4:103422486-103538459
HOXA5	Hs00430330_m1	chr7:27180671-27183287
CREB1	Hs00231713_m1	chr2:208394616-208468155
IRF1	Hs00971960_m1	chr5:131817301-131826490
Target genes		
AKTIP	Hs01591423_m1	chr16:53524952-53537163
NAMPT	Hs00237184_m1	chr7:105888731-105925638
TPK1	Hs01558699_m1	chr7:144149034-144533146
ISG15	Hs00192713_m1	chr1:948803-949920
HIST1H4C	Hs00543883_s1*	chr6:26104104-26104538
CRTAM	Hs00219699_m1	chr11:122709208-122743347
PLLP	Hs00762550_s1*	chr16:57290009-57318599
RFX5	Hs00230841_m1	chr1:151313116-151319727
NKG7	Hs01120688_g1*	chr19:51874866-51875969
RAB17	Hs00940833_m1	chr2:238482965-238499736
CISD2	Hs00391903_m1	chr4:103790135-103810399
HDAC4	Hs00195814_m1	chr2:239969864-240322643
WDR43	Hs01064086_m1	chr2:29117509-29171088
CXCL11	Hs04187682_g1*	chr4:76955843-76962568
BATF2	Hs00912737_m1	chr11:64755415-64764517
WNT11	Hs00182986_m1	chr11:75897369-75917576
GSTA4	Hs01119249_m1	chr6:52842751-52860176
TFF1	Hs00907239_m1	chr21:43782391-43786703
PLAUR	Hs00958880_m1	chr19:44152732-44174502
TBP	Hs00427620_m1	chr6:170863421-170881957
pri-miRNAs		
hsa-mir-33a	Hs03293451_pri	chr22:42296948-42297016
hsa-mir-520b	Hs03295424_pri	chr19:54204481-54204541
hsa-mir-26b	Hs03302654_pri	chr2:219267369-219267445
hsa-let-7b	Hs03302548_pri	chr22:46509566-46509648
hsa-mir-655	Hs03304873_pri	chr14:101515887-101515983

hsa-mir-520e	Hs03303928_pri	chr19:54178965-54179051
hsa-mir-92a-1	Hs03302603_pri	chr13:92003568-92003645
hsa-mir-503	Hs03304160_pri	chrX:133680358-133680428
hsa-mir-17	Hs03295901_pri	chr13:92002859-92002942
mature miRNAs		
hsa-miR-26b	000407	chr2:219267369-219267445
hsa-miR-18a-3p	002423	chr13:92002859-92002942
RNU48	001006	chr6:31803040-31803103

*Mapped to single exon.

2.1.2.1.2. cDNA synthesis

Total RNA was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA; cat. no. 1708891). Input RNA (48 ng), 4µl of iScript Reaction Mix and 1 µl of reverse transcriptase were mixed in a total volume of 20 µl and incubated as follows: priming at 25°C for 5 minutes, retrotranscription at 42°C for 30 minutes and inactivation at 85°C for 5 minutes. cDNA samples were stored at -20°C until use.

For mature miRNAs, TaqMan microRNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA; cat. no. 4366596) and TaqMan MicroRNA Assays (miRNA-specific primer) were used to generate complementary DNA for each specific miRNA. Input RNA (300 ng), RT primer pool (6 µl), dNTPs (100 mM) (0.3 µl), MultiScribe Reverse Transcriptase (3 µl), 10X RT Buffer (1.5 µl) and RNase Inhibitor (0.19 µl) were mixed in a total volume of 12 µl, and incubated as follows: at 16°C for 30 minutes, at 42°C for 30 minutes and at 85°C for 5 minutes. cDNA samples were stored at -20°C until use.

2.1.2.1.3. Quantitative PCR

Two different platforms were used for qPCR: the Fluidigm BioMark dynamic array system (Fluidigm Corporation, San Francisco, CA, USA) and the Eco Real-Time PCR System (Illumina). All qPCR analyses were performed in duplicate in

the first case, and in triplicate in the second. The expression of housekeeping genes (*TBP* for TFs, pri-miRNAs and TF target genes; and *RNU48* for mature miRNAs) were simultaneously quantified in each sample and used as endogenous controls of input RNA. The relative expression of each gene was calculated using the accurate Ct method as previously described (Martín-Pagola et al., 2004). The amount of target, normalized to an endogenous reference and relative to a calibrator, is given as:

$$\textbf{Relative expression} = 2^{-\Delta\Delta Ct}$$

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

2.1.2.1.3.1. Fluidigm BioMark dynamic array system

Expression analyses for TFs, pri-miRNAs and TF-target genes were performed at the Gene Expression Unit of the University of the Basque Country (UPV/EHU). Firstly, preamplification of the cDNA from each gene was carried out using the TaqMan PreAmp Master Mix (Applied Biosystems; cat. no. 4391128) following the manufacturer's instructions. Briefly, 2.5 µl of PreAmp Master Mix, 1.25 µl of assay pool (0.2X) and 1.25 µl of cDNA were used in a total volume of 5 µl. Cycling conditions were as follows: 95ºC for 10 minutes followed by 14 cycles of 95ºC for 15 seconds and 60ºC for 4 minutes. Afterwards, preamplified cDNAs were diluted 1:5 with TE buffer. For qPCR analyses, TaqMan Fast Advanced Master Mix (Applied Biosystems; cat. no. 4444557), 2X Assay Loading Reagent (Fluidigm Corporation; cat. no. 85000736) and 20X GE Sample Loading Reagent (Fluidigm Corporation; cat. no. 85000746) were employed. Five microliters of both 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 amplified cDNA) were used per inlet. Cycling conditions were as follows: polymerase activation at 95°C for 1 minute and 35 amplification cycles at 95°C for 5 seconds and at 60°C for 20 seconds.

2.1.2.1.3.2. Eco Real-Time PCR system

For the quantification of mature miRNAs, TaqMan MicroRNA Assays were used. Five microliters of TaqMan Universal PCR master mix (Thermo Fisher Scientific; cat. no. 4440042), 4.5 µl of nuclease-free water, 0.5 µl of TaqMan MicroRNA Assay and 0.7 µl of the RT product were used per inlet. Cycling conditions were as follows: polymerase activation at 95°C for 10 minutes and 40 amplification cycles of 15 seconds at 95°C and 1 minute at 60°C.

2.1.2.2. Cellular localization of TFs

Localizaion of candidate TFs was determined in C2BBe1 cells cultured with PT-G or PT-BSA during 4 h as described in the *Material and Methods* section *1.3*..

2.1.2.2.1. Immunofluorescense assays

C2BBe1 cells were grown on glass coverslips and treated with PT-G or PT-BSA for 4 h. Cells were washed with PBS, fixed with 4% formaldehyde/PBS and then permeabilized with Triton 0.5%, followed by three washes using PBS. Cells were blocked with blocking solution Image-iT FX signal enhancer (Invitrogen, Carlsbad, California, USA; cat. no. A31629) and incubated with primary rabbit IgG antibodies: anti-ELK1 (1:500 dilution; cat. no. ab32106), anti-NFκB

p105/p50 (1:400; cat. no. ab7971), anti-CREB1 (1:1000; cat. no. ab31387) and anti-IRF1 (1:1000; cat. no. ab26109) (ABCAM, Cambridge, UK). Slides were washed with PBS and incubated with the secondary antibody, goat anti-rabbit IgG (Invitrogen, cat. no. A31627).

Slides were washed with PBS and incubated with primary mouse antibody anti-E-Cadherin 1:100 (BD Bioscience, San Jose, CA, USA, cat. no. 610182) (intercellular junction marker), and finally with secondary antibody goat anti–mouse IgG (Invitrogen, cat. no. A31621). Slides were mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, Burlingame, California, USA; cat. no. H-1200) (nuclear staining) and observed under a Nikon Eclipse Ti fluorescence microscope.

2.1.2.2.2. Nuclear and cytoplasmic protein extraction

Nuclear and cytoplasmic protein extracts of C2BBe1 cells incubated with PT-G or PT-BSA for 4 h were isolated using a commercial Nuclear Extract Kit (Active Motif, Carlsbad, CA, USA; cat. no. 40010). Briefly, cells were washed with PBS/Phosphatase Inhibitors and removed from the culture dish by gentle scraping. Cells were collected by centrifugation at 4ºC at 200 g for 5 min. For the collection of the cytoplasmic fraction, cells were resuspended in 500 µl 1X Hypotonic Buffer and incubated 15 min on ice to allow cells to swell. Then, 25 µl of detergent solution was added. Once the cells were lysed, the suspension was centrifuged at 4ºC at 14,000 g for 30 s. The supernatant (cytoplasmic fraction) was stored at -80ºC until used, and the pellet was used for the collection of the nuclear fraction. The pellet was resuspended in 50 µl of Complete Lysis Buffer and incubated on ice on a rocking platform at 150 rpm for 30 min. The sample was vortexed and centrifuged at 4ºC at 14,000 g for 10 min. The supernatant, containing the nuclear fraction, was stored at -80ºC until used.

Protein concentrations were measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific; cat. no. 23227). The working reagent (WR) was prepared mixing 50 parts of BCA Reagent A with 1 part of BCA Reagent B. The contents of one BSA ampule was serially diluted into several vials (total volume of 2.5µl), creating a set of diluted standards as in the manufacturer's protocol. Twenty microliters of WR were added to each standard and sample tube, and they were incubated at 37ºC for 30 min. Once the tubes were cooled to room temperature, absorbance at 562 nm was measured by spectrophotometry and a standard curve was prepared by plotting the average absorbance of each BSA standard *vs.* its concentration in µl/ml. The protein concentration of each sample was determined using the standard curve.

2.1.2.2.3. Immunoblot analysis

Nuclear and cytoplasmic protein extracts were heat-denatured at 95°C for 5 min, loaded into 12% SDS-PAGE gels and transferred to nitrocellulose membranes using the Trans-Blot Turbo Transfer (Bio-Rad). After transfer, membranes were blocked in 5% non-fat milk in Tris-buffered saline with 0.05% Tween (TBST) and incubated overnight at 4° C with the following primary mouse IgG antibodies: anti-HDAC1 (nuclear control) (1:10000 dilution; ABCAM, cat. no. ab7028), and anti-α-tubulin (cytoplasmic control) (1:5000 dilution, Sigma-Aldrich, cat. no. T9026); and the following rabbit IgG antibodies: anti-ELK1 (1:500; cat. no. ab32106), anti-NFκB p105/p50 (1:400; cat. no. ab7971), anti-CREB1 (1:1000; cat. no. ab31387), anti-IRF1 (1:1000; cat. no. ab26109) followed by an incubation of 1 hour with HRP-conjugated anti-mouse IgG (1:1000) or anti-rabbit IgG (1:2000) secondary antibodies (Jackson ImmunoResearch Laboratories, Inc; West Grove, PA, USA, cat no. 115-035-062 and 111-035-003). Proteins were visualized using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific; cat. no. 34094) on a ChemiDoc MP system.

2.1.2.3. Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) experiments were performed in duplicate using Chromatin Shearing Optimization kit-Low SDS (Diagenode, Seraing, Belgium; cat. no. AA-001-0100) according to the manufacturer's instructions. Briefly, C2BBe1 cells were fixed with 1% formaldehyde and then glycine was added to stop fixation. Cells were collected by centrifugation and resuspended in Lysis Buffer iL1 (10 ml buffer iL1 per 10 million cells). Cells were incubated at 4ºC for 10 min and centrifuged to pellet the cells. The supernatant was discarded and Lysis Buffer iL2 was added (10 ml per 10 million cells). Cells were incubated at 4ºC for 10 min and centrifuged to pellet the cells. The supernatant was discarded and finally Complete Shearing Buffer iS1 (1 ml per 10 million cells) was added. Chromatin was sheared by sonication using a Bioruptor apparatus (Diagenode; cat. no. UCD-300 TM) for 5 cycles of 1 min (30 s on and 30 s off) twice at high intensity.

Automated immunoprecipitation was performed using the SX-8G IP-Star Compact (Diagenode; cat. no. UH-002-0001) according to the manufacturer's instructions, with 2.5 µg of the anti-CREB1 and anti-IRF1 antibodies used in the immunofluorescence and the Western blot analyses, together with rabbit IgG were used per sample. Enrichment of ChIP was analyzed by qPCR using Mesa Green MasterMix (Thermo Fisher Scientific) in triplicate. Primer pairs used for the selected targets covered H3K27ac rich promoter regions, and are listed in Table 5.

Table 5. Primer sequences (5' to 3') for ChIP-qPCR analyses.

		Forward primer	**Reverse primer**
IRF1 targets	*CISD2*	GACGAAGTAGAGACAGCAAGAG	GGATACTGTGTGCGATGAGATAA
	HDAC4	CAGCCTTGCGTCACCTC	GACGAGCTCTTCATTAGAAACCA
	WDR43	GGTCACTTACGAGTATGGGAGA	AAGGCACGTACTCCTGGT
	CXCL11	CTCTTTGAGTCATGCACCTTTC	TCACAGTGCTTTCACATTCTTATC
	BATF2	GCCAAGTTTCAGTTTCTCCTAAAG	CGGAAGGCCAGTTCATGTTA
CREB1 targets	*AKTIP*	CGGTCCTGCAAATCAAATCAC	CGATACTTCCATGACTGACAGG
	NAMPT	CGTTGCTTAAGTCACTGCTC	CCCTCTCTCCGTTTCCC
	TPK1	GGCAGCAGTCGCACTTA	GTCGATCGCCGTAGCTC
	ISG15	TCCCTGTCTTTCGGTCATTC	CTTCAGTTTCGGTTTCCCTTTC
	HIST1H4C	GGTCCGCCAAGTTTGTATTTAAG	CGACCAGACATGATTCCTATCG

2.1.2.4. Expression of miRNA target genes in CD

We searched for the target genes of the mature miRNAs associated to each altered pri-miRNA in the miRTarBase database, taking into consideration only data supported by strong evidence, namely reporter assay and Western blotting technologies. The expression of the target genes was interrogated in our RNA-seq dataset of intestinal cell fractions.

2.1.3. Statistical analysis

- We used the corrgram R package to calculate the degree of the correlation of pairwaise gene-expression and display the results graphically (Friendly, 2002).
- Differences in expression levels were analyzed with the Wilcoxon matched-pairs signed rank test (matched active and GFD CD patients) or the Mann Whitney test (unmatched comparisons) using GraphPad Prism version 5.0 software (GraphPad Software Inc, La Jolla, CA).
- For Western blot and ChIP-qPCR experiments, one-tailed Student's t-tests and paired t-tests were used respectively to compare groups using GraphPad Prism version 5.0 software.

2.2. Topologically associating domains in CD

2.2.1. Identification of altered 3D chromatin structures in CD

2.2.1.1. Co-expression analysis

The co-expression analysis was performed using the RNA-seq data retrieved from http://www.ncbi.nlm.nih.gov/sra/, experiment number SRP077708. Particularly, we used the total RNA-seq data from the epithelial cell-enriched fractions from 10 active CD patients and 12 non-celiac controls.

Again, the WGCNA package was used. In this case, due to the characteristics of the data, results were not available for all transcripts. We included those transcripts in which half of the samples had FPKM (Fragments Per Kilobase Million) >1. FPKM values were log10(FPKM+1) transformed, as suggested by the authors of the WGCNA package; and the transcripts with higher variance were kept using the function "varianceBasedfilter" from the DCGL package, and then default procedure was used. In each analysis the "power" parameter was set using the best value in the "Scale Free Topology Model Fit" analysis.

The workflow of the study is summarized in Figure 8.

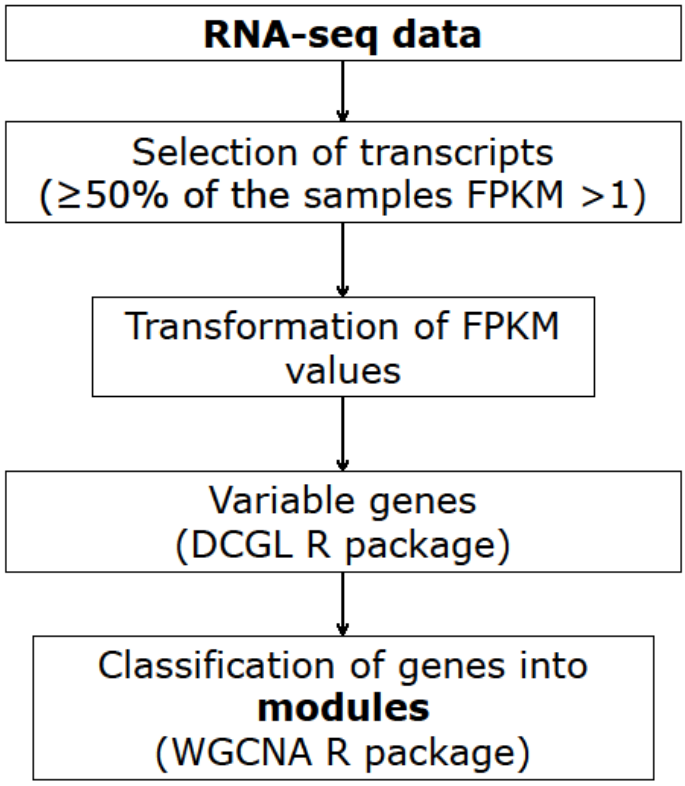


Figure 8. Flowchart of the RNA-seq data analysis.

2.2.1.2. Overlap of selected genomic features

The *Intersect* tool in BedTools (Quinlan and Hall, 2010) was used to find overlaps between genomic coordinates of conserved TADs (Dixon et al., 2012) and genes from the co-expression modules identified in the RNA-seq data.

Genomic regions were selected for further characterization according to the following criteria:

- Potential disruption of TADs: TADs that overlapped with genes that were co-expressed in controls but were not co-expressed in active CD.
- Potential merge of TADs: Regions where nearby genes from different adjacent TADs were not co-expressed in controls, but were so in active CD. The distance between TADs was calculated using BedTools. Two Mb was defined as the maximum distance after removing outliers (Figure 9).

Based on that distance, the *window* method in BedTools was used to find TADs close to each other.

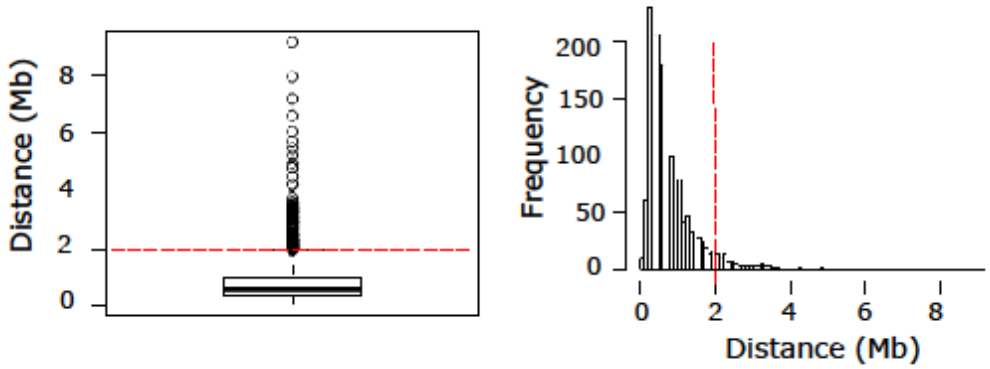


Figure 9. Box-plot and histogram used to calculate the distance between adjacent TADs. The distance between TADs is shown in each of the graphs. The red line shows the maximum distance after removing outliers, 2 Mb.

Finally, genomic coordinates of the selected regions that overlapped with SNPs associated with CD with a nominal p value below 0.05 (Garcia-Etxebarria et al., 2016) were identified using the *intersect* method in BedTools. A candidate was selected for each criterion (disruption and merge) for experimental validation, namely *HSCB-XBP1* region located in hg19 chr22:29040000-29360000, and *PROCR-ROMO1* region located in hg19 chr20:33480000-34320000.

The workflow of the study is summarized in Figure 10.

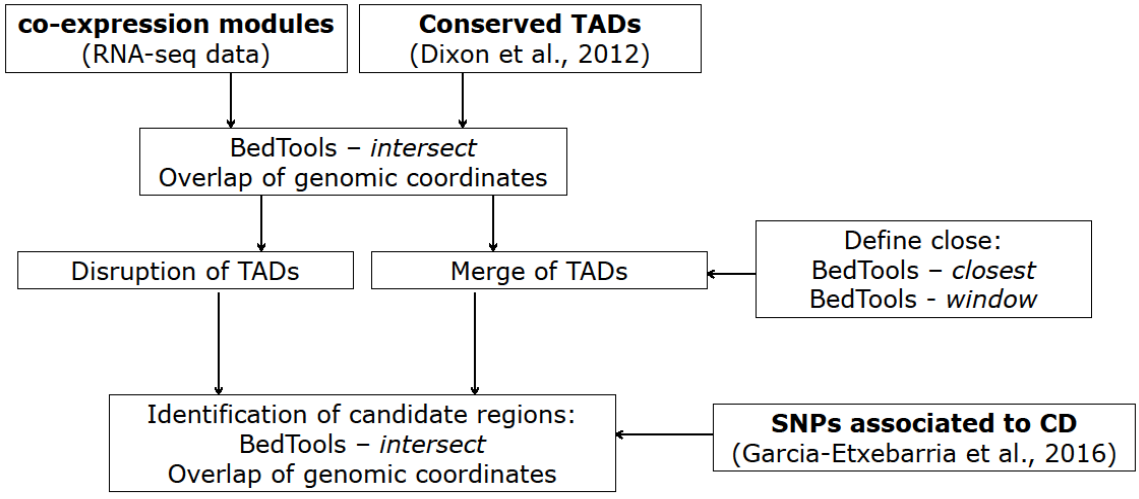


Figure 10. Flowchart of the methodology used for the TADs analysis.

2.2.2. Experimental confirmation of candidates

2.2.2.1. Chromatin accessibility experiment

HCT116, HCT15 and HEK293FT cell cultures were washed with HBSS and cells were harvested by trypsinization. Cells were centrifuged at 1,500 rpm for 5 min and resuspended in 1 ml PBS, 1ml C1 lysis buffer (1.28 M sucrose, 40 mM Tris HCl, 20 mM $MgCl_2$, 4% Triton X-100) and 3 ml H_2O (+ proteinase inhibitor). Cells were incubated on ice for 15 min and nuclei were collected by centrifugation at 4°C and 2,500 rpm for 15 min. Up to 120,000 nuclei were resuspended in a total volume of 100 µl wash buffer (10 mM Tris, pH 7.4, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$ and 300 mM sucrose). DNase (10 µl) (Macherey-Nagel; cat. no. CAS 9003-98-9) and 90 µl of DNase reacting buffer were added to each sample. A control sample was incubated in DNase reacting buffer without DNase. The reactions were incubated at 37°C for 20 min, and stopped by adding stop buffer (nuclei wash buffer supplemented to 50 mM EDTA). DNA was extracted as described in the *Material and Methods* section *1.4.*.

Finally, to confirm whether samples had been successfully digested and to analyze chromatin accessibility in the interrogated regions, qPCR was performed in DNase-treated samples and in undigested controls using primers for constitutively open and closed chromatin regions, and primers for the interrogated regions. Primers were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Table 6).

Table 6. Primer sequences (5' to 3') for chromatin accessibility experiment. *TBP* and *GAPDH* were used as closed and open chromatin control regions, respectively.

	Forward primer	**Reverse primer**
***HSCB-XBP1* region**	TCCCAAAGTGCTGGGATTAC	AATACTGCCACCCAGTGACC
***PROCR-ROMO1* region**	ACTGTCGCCCTTAAGTTCACTC	GCTTTTCCAGCCTCCTGTAG
***TBP* gene body**	TTGGCAGGCCTACAGTTTTC	AAACTGGTCAGCCTTCTTGC
***GAPDH* promoter**	AAGGTGAAGGTCGGAGTCAAC	CCCATACGACTGCAAAGACC

2.2.2.2. Disruption of DNase I hypersensitive sites

CRISPR-Cas9 technique was used to permanently disrupt two DNase I hypersensitive sites (DHSs) harboring CTCF binding sites in different epithelial cell lines by deleting hg19 chr22:29186082-29186523 and hg19 chr20:34026814-34027191 regions within the *HSCB-XBP1* and *PROCR-ROMO1* regions, respectively. A schematic representation of the procedure is shown in Figure 11.

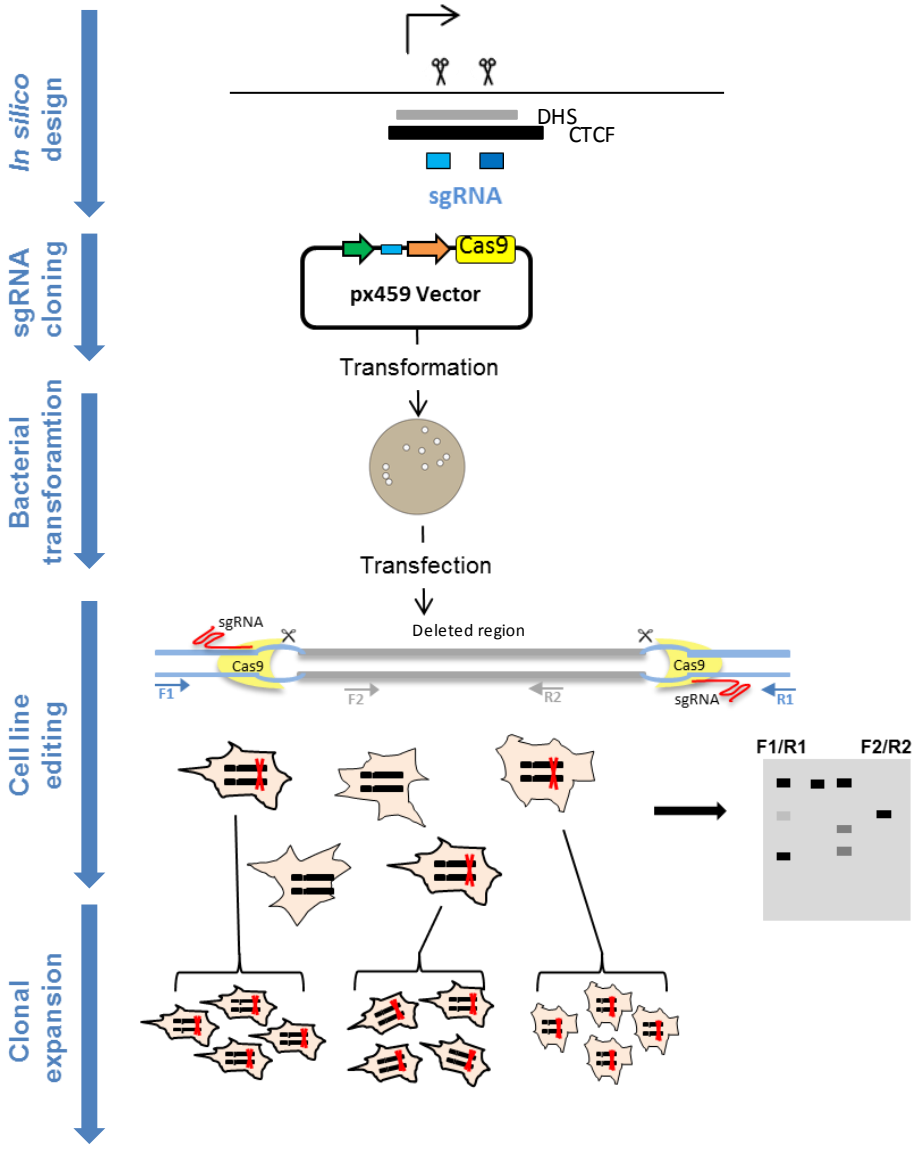


Figure 11. Schematic representation of CRISPR-Cas9 mediated deletion of DNase I hypersensitive sites in epithelial cells. Steps for reagent design, construction, validation and cell line expansion are depicted. Custom sgRNAs (blue bars) and genotyping primers are designed *in silico*. sgRNA guide sequences are cloned into a vector bearing sgRNA scaffold and Cas9. The plasmid is then transfected into cells and the ability to mediate targeted cleavage is checked. Finally, transfected cells are clonally expanded to obtain cell lines with defined mutations. Adapted from Ran et al., 2013 (Ran et al., 2013).

2.2.2.2.1. sgRNA design

The deletion of DHSs was expected to provoke a more drastic effect. Particularly, CTCF binding site enriched regions were chosen in order to alter putative TAD boundaries, as mentioned previously. Flanking regions of up to 200 bp upstream and downstream of each DHS 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 20 bp target options, with different specificity values based on a statistical algorithm of off-target hits. The best outputs for each region were chosen to ensure efficiency and avoid off-target mutations (Table 7).

Table 7. Designed sgRNA oligos for CRISPR-Cas9 mediated DHSs deletion.

		Locus (hg19)	Sequence	PAM*	Antiparallel
***HSCB-XBP1* region**	sgRNA1	chr22:-29186082	CACCGCGGAGTCTCGCTGTTTCGCC	AGG	AAACGGCGAAACAGCGAGACTCCGC
	sgRNA2	chr22:-29186523	CACCGCAGGAGAATCGCCTGAACCC	AGG	AAACGGGTTCAGGCGATTCTCCTGC
***PROCR-ROMO1* region**	sgRNA1	chr20:-34026814	CACCGTCAATGGCATTTGGGGGTA	GGG	AAACTACCCCCAAATGCCATTGAC
	sgRNA2	chr20:-34027191	CACCGACAAACGTAAAAAATCAGCC	AGG	AAACGGCTGATTTTTTACGTTTGTC

*Protospacer adjacent motif: essential sequence for target binding.

2.2.2.2.2. sgRNA cloning

Complementary sgRNA oligos were phosphorylated and annealed by mixing 1 µl (100 µM) of each oligo in 1 µl 10X T4 DNA ligation buffer and 0.5 µl of T4 PNK Polynucleotide kinase (New England Biolabs, Ipswich, MA, USA; cat. nos. B0202S and M0201L) in a final volume of 10 µl under the following conditions: incubation at 37°C for 30 minutes, 95°C at 5 min and ramp down to 25°C at 5°C/min. Then, the annealed sgRNAs were inserted at the sgRNA scaffold site of the px459 plasmid vector (Figure 12) (Addgene, Cambridge, MA, USA). Cloning of the oligonucleotides into the vector consisted of a digestion, followed by a ligation reaction. Two micrograms of vector were digested with 2 µl of fast digest *Bbsl* enzyme and 2 µl of Fast Digest Green Buffer (10X) (Thermo Fisher Scientific; cat. nos. FD1014 and B72) in a final volume of 20 µl at 37ºC for 1 h. The digestion was tested by electrophoresis (150 V – 15 min in 1% gel agarose) visualized on a ChemiDoc MP system, and subsequently purified with Nucleospin Gel and PCR clean up kit (Macherey-Nagel; cat. no. 740609.250) following the manufacturer's instructions. Fifty nanograms of digested vector and 1 µl of annealed oligonucleotides were ligated with T4 DNA ligase (New England Biolabs; cat. no. M0202S) for 1 h.

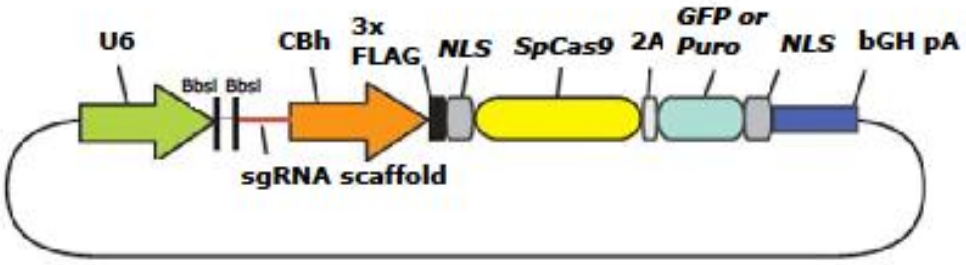


Figure 12. Schematic representation of px459 plasmid containing Cas9 and the sgRNA scaffold. The guide oligos contain overhangs for ligation into the pair of *BbsI* sites in the plasmid. Digestion of the plasmid with *BbsI* allows the insertion of annealed oligos. Adapted from Ran eta al., 2013 (Ran et al., 2013).

2.2.2.2.3. Bacterial transformation and selection

The ligated construct was transformed into Subcloning Efficiency *E. coli* DH5α Competent Cells (Thermo Fisher Scientific; cat. no. 18265-017) by heat shock following the protocol supplied. Bacteria were seeded onto 50 µg/µl ampicillin-containing LB agar plates and incubated overnight at 37°C. Several colonies were picked and incubated overnight at 37°C in LB medium with ampicillin to amplify the recombinant plasmid. Plasmid DNA was extracted using the NucleoSpin Plasmid EasyPure kit (Macherey-Nagel; cat. no. 740727). Digestion of 500 ng of plasmid DNA was performed with 1 µl of *Bbsl* enzyme, 0.5 µl of *Agel* enzyme (Thermo Fisher Scientific; cat. no. ER1461), and 2 µl of Fast Digest Green Buffer (10X) in a final volume of 20 µl and at 37ºC for 1 h. Visualization of electrophoresis bands was used for clone selection.

2.2.2.2.4. Cell line editing

HCT116 and HCT15 cells were reverse transfected with 600 ng of the recombinant plasmid using XTremHP DNA reagent (Invitrogen; cat. no. 06366236001), and HEK293FT cells were reverse transfected with 300 ng of plasmid using Lipofectamine 2000 (Thermo Fisher Scientific; cat. no. 11668-027). Transfection was performed in 24-well plates at a density of 100,000 cells per well, following the manufacturer's protocols. After overnight incubation, puromycin was added to a final concentration of 4 µg/ml, and cells were incubated for another 48 hours to select for transfected cells. For mutational analysis, cells were lysed in 300 µl of Lysis Buffer (50 mM Tris pH8, 50 mM KCl, 2.5 mM EDTA, 0.45% NP-40, 0.45% Tween 20 and 10 µl of 10 µg/µl Proteinase K). Samples were incubated at 60°C for 2 hours and then proteases were inactivated at 95°C for 10 min. PCR was used to amplify the target region using flanking primers, and wild-type and truncated genomic fragments were distinguished by gel electrophoresis.

2.2.2.2.5. Clonal expansion

Single-cell isolation was performed in 96 well plates by seeding cells at low density (1 cell per five wells). Isolated cells were expanded until healthy colonies were formed, and mutational analyses were performed as previously described to select the clones harboring the mutation.

2.2.2.3. Gene expression analysis

For the study of the disruption in the *PROCR-ROMO1* region, expression and co-expression of *PROCR* and *ROMO1* genes was measured in edited and wild-type epithelial cell lines. Expression analyses were performed with SYBR green detection and primers were selected from PrimerBank (https://pga.mgh.harvard.edu/primerbank) (Table 8).

RNA was isolated from edited and wild-type cells as described in the *Material and Methods* section *1.4.*, and cDNA was synthesized as described in the *Material and Methods* section *2.1.2.1.2.*.

qPCR analyses were performed in duplicate and the expression of the housekeeping gene *HPRT* was simultaneously quantified and used as endogenous control of input RNA. The Eco Real-Time PCR system was used and relative expression of each gene was calculated as described in the *Material and Methods* section *2.1.2.1.3.*. Briefly, iTaq universal SYBR Green Supermix (Bio-Rad; cat. no. 172-5121) 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; and 15 seconds at 95°C, followed by 15 seconds at 55°C and 15 seconds at 95°C for the melting curve analysis.

Table 8. Primer sequences (5’ to 3’) for *PROCR* and *ROMO1* genes expression. *HPRT* was used as a housekeeping control.

Gene	PrimerBank ID	Amplicon size (bp)	Forward primer	Reverse primer
PROCR	34335271c1	80	CCTACAACCGCACTCGGTATG	CGCGGAAATATGTTTCTGCACA
ROMO1	115430214c1	107	AAGCTGCTTCGACCGTGTC	CCCGCATTCCGATCCTGAG
HPRT	164518913c2	190	ACCAGTCAACAGGGGACATAA	CTTCGTGGGGTCCTTTTCACC

2.2.2.4. Characterization of cell lines

Genotyping of SNPs rs6060369, rs224371 and rs2104417 was performed in HCT116, HCT15 and HEK293FT cell lines. These SNPs were selected because they are tag SNPs of the associated-SNP haplotypes that are located between the two TADs in the *PROCR-ROMO1* region.

SNPs were genotyped using predesigned rhAmp SNP Assays (Integrated DNA Technologies, Coralville, IA, USA) (Table 9). Bi-allelic discrimination was achieved through the competitive binding of two allele specific forward primers, one labelled with FAM dye and the other with Yakima Yellow (YY) dye.

Table 9. Predesigned Genotyping assays for genotyping experiment.

SNP ID	Assay name	Location (hg19)
rs6060369	Hs.GT.rs6060369.C.1	chr20: 33907160
rs224371	Hs.GT.rs224371.A.1	chr20: 34074830
rs2104417	Hs.GT.rs2104417.A.1	chr20: 34127870

DNA amplification was carried out on an Eco Real-Time PCR System in 10 µl reaction volume, containing 40 ng DNA mixed with rhAmp Genotyping Mix (rhAmp Genotyping Master Mix and rhAmp Reporter Mix w/Reference) (Integrated DNA Technologies; cat. nos. 1076015 and 1076021) and rhAmp SNP Assays (Integrated DNA Technologies). Cycling conditions were 95°C for 10 minutes, followed by 45 cycles of 95°C for 10 seconds, 60°C for 30 seconds and 68°C for 20 seconds.

2.2.3. Statistical analysis

Chromatin accessibility was quantified using the shift in Ct values between digested and undigested chromatin, and candidate regions were compared to the

closed-chromatin control (*TBP*) using one-tailed Student's t-tests. GraphPad Prism version 5.0 software was used to assess differences between groups

Expression and co-expression in cultured cells were assessed by the Mann Whitney test and Spearman's correlation test, respectively. GraphPad Prism version 5.0 software was used to assess differences between groups.

2.3. Acute changes in methylation patterns in CD

DNA methylation was studied in 8 non-celiac control and 7 CD treated patient's biopsy samples incubated with and without gliadin as described in the *Material and Methods* section *1.3*.. DNA was isolated as described in the *Material and Methods* section *1.4*..

2.3.1. Bisulfite conversion

For the conversion of unmethylated cytosines to uracil, EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, CA, USA; cat. no. D5030) was used. Briefly, 130 µl of Lightning Conversion Reagent were added to 20 µl of DNA sample (300 ng), and placed in a thermal cycler at 98ºC for 8 min and 54ºC for 60 min. Samples were loaded into the Zymo-Spin IC Column containing M-Binding Buffer, and after centrifugation, M-Wash Buffer was added to the column. L-Desulphonation Buffer was added and incubated for 15 minutes at room temperature. Converted DNA was eluted in 20 µl of M-Elution buffer.

2.3.2. Amplification, quantification, purification and normalization of selected regions

PCR amplification of bisulfite-converted DNA was performed in relevant regions prior to sequencing. Regions of interest were selected according to the following criteria:

- Differentially methylated regions (DMRs) identified in a previous work from our group (Scientific Reports, under review) that showed expression changes upon 4 h gliadin stimulation in the acute microarray experiment: hg19 chr2:240271171-240271276 region harboring the *HDAC4* gene body.
- Top DMRs identified in the epithelial fraction in the same work: the hg19 chr6:30131458-30132471 region mapping to part of the 5'UTR, the first exon, and the gene body of *TRIM15*; hg19 chr6:32819858-32820249, mapping to the *TAP1* promoter; and hg19 chr6:31322766-31323506, mapping to the *HLA-B* promoter.
- Top expression changes upon 4 h gliadin stimulation in the acute microarray experiment that have differentially methylated positions according to the literature: the hg19 chr17:26732864-26733385 region, mapping to the promoter of *SLC46A1* (Diop-Bove et al., 2009), and the hg19 chr7:24796542-24797487 region, mapping to the promoter of *DFNA5* (Kim et al., 2008).

Primers were designed using the MethPrimer 2.0 online platform (http://www.urogene.org/methprimer2/), which performs *in-silico* bisulfite conversion of the target sequence. In the case of *SLC46A1* and *DFNA5* promoters, primers were designed manually, since the platform was not able to design them appropriately. *GAPDH* was amplified with standard genomic primers and primers for bisulfite-treated samples, in order to evaluate conversion efficiency (Table 10).

Amplifications were performed using the PyroMark PCR Kit (Qiagen, Hilden, Germany; cat. no. 978703 VF 40) in a 40-µl PCR reaction containing 15 ng of bisulfite-converted DNA mixed with 20 µl PyroMark PCR Master Mix 2X, 4 µl CoralLoad Concentrate 10X, 2.4 µl $MgCl_2$ (25 mM), and 0.8 µl of each primer (10 µM). Cycling conditions were 95°C for 15 minutes, followed by 50 cycles of 94°C for 30 seconds, 56°C for 30 seconds and 72°C for 30 seconds, and a final extension at 72°C for 10 minutes.

Amplification products were tested by electrophoresis. The intensity of the bands was measured using Image Lab v5.2.1. PCR products from the same individuals were pooled at equal concentrations to construct equilibrated amplicon pools for each sample. NucleoSpin Gel and PCR Clean-up (Macherey-Nagel; cat. no. 740609.250) was used to purify the amplicon pools of each individual prior to sequencing.

Finally, the DNA concentration in each pool was measured by fluorescence (QuBit) and normalized to 0.2 ng/µl. Libraries were generated with a Nextera XT kit (Illumina) and were sequenced in an Illumina MiSeq system using the Miseq Reagent kit v3 (600 cycles, 25M reads) (Illumina) in the Sequencing and Genotyping Unit of the University of the Basque Country (UPV/EHU).

Table 10. Primer sequences (5’ to 3’) for amplification of selected regions.

	Forward primer	Reverse primer	Amplicon size (bp)
HDAC4	AATTTATTAAAATGATGAAAAGTAGGA	AAAACAAAACCCCTTATACCCAA	293
TRIM15	TGTTAAGAGGAGGAGTAGGTAGAGATTT	CCTATCTACCTTCAATCTAAAAATACC	362
TAP1	TAGGGAATAGATTGAAGGTTTTAGG	CAATCTAACTAAAACTAACCTACTTAAACT	290
HLA-B	AAATTTTTAGTGGGATAAGAAAAT	CCAAAAATAAACAACTATAATAATACCTTC	364
SLC46A1	TTGTAGGATTAAGGTAAAGTTGG	CACTTTACAAATAAAATCATCCC	368
DFNA5	AGGGTGGTTTAGAGAGAAA	CTCTCTAAAACTTCTAAAAAATC	346
GAPDH non-modified	CTCTTGCTACTCTGCTCTGG	GCTAAGTTTAGCCTGCCTGG	189
GAPDH modified	GTATTTGTTGATGGGTTAAGG	ATAAAAACAAATCCCCTACCC	150

2.3.3. Methylation analysis using Next-generation sequencing

Next-generation sequencing (NGS) reads were mapped against the bisulfite converted sequences prepared in Meth Primer 2.0 (http://www.urogene.org/methprimer2/) using Bowtie v2.2.6 (Langmead and Salzberg, 2012). Then, Samtools v1.4 (Li et al., 2009) was used to determine the basic statistics of the mapping success and bam-readcount (https://github.com/genome/bam-readcount) was used to obtain the counts of each nucleotide for each position. Home-made scripts written in Python and R were used to parse the last file to retrieve the C/T proportion for each CpG and non-CpG C position. Only the amplicons with a C proportion smaller than 5% in non-CpG positions were used for downstream analyses.

The workflow of the study is summarized in Figure 13.

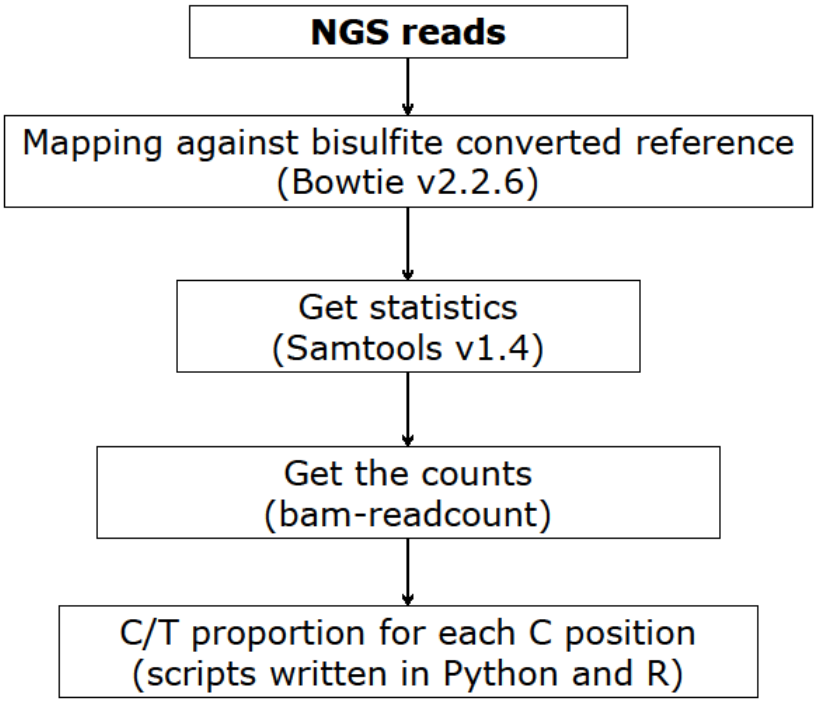


Figure 13. Flowchart of the bisulfite NGS analysis.

2.3.4. Statistical analysis

Methylation differences between GFD without gliadin *vs.* controls without gliadin, as well as controls with gliadin *vs.* without gliadin, and GFD with gliadin *vs.* without gliadin were analyzed with Mann Whitney tests carried out in R.

Results

1. Whole genome co-expression in CD

Using co-expression analyses we aimed to identify groups of co-transcribed genes that might share common regulatory elements.

1.1. Co-expression alterations in CD upon gliadin challenge

Out of the 19,851 protein-coding genes in the expression array, those that were more variable than the median gene of each of the experiments were used to construct co-expression modules. In the acute experiment, 6,863 genes were assigned to 18 modules in unchallenged biopsies (Appendix 1) and to 16 modules in the gliadin-challenged biopsies (Figure 14A) (Appendix 2). In the long-term experiment, 8,342 genes were included in 35 modules in GFD samples (Appendix 3), and in another 35 modules in active samples (Figure 14B) (Appendix 4).

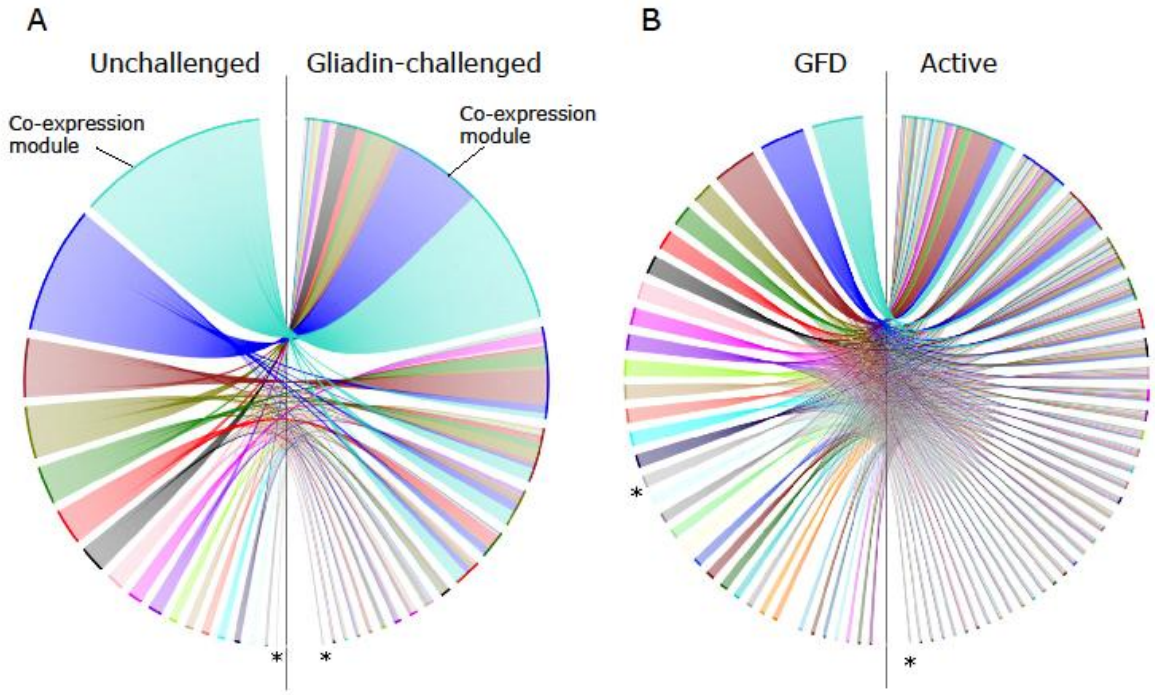


Figure 14. Genome-wide co-expression in the A) acute and B) long-term experiments. Co-expression modules in unchallenged, gliadin-challenged, GFD and active samples are represented. Each color in represents a co-expression module, and the asterisk points to those genes that do not belong to any co-expression module.

We then searched for differentially co-expressed genes or DCGs in each of the two experiments (Figure 15). In the acute experiment, co-expression was disrupted in 21.71% and 9.93% of the genes in unchallenged (Appendix 5) and in gliadin-challenged biopsies (Appendix 6), respectively, when those groups were compared to their counterparts. In the long-term experiment, co-expression was disrupted in 14.68% and 12.60% of the genes in the GFD (Appendix 7) and the active CD patients group (Appendix 8), respectively, when those groups were compared to their counterparts. In addition, DCGs also created co-expression modules (Figure 15, Appendix 5-8).

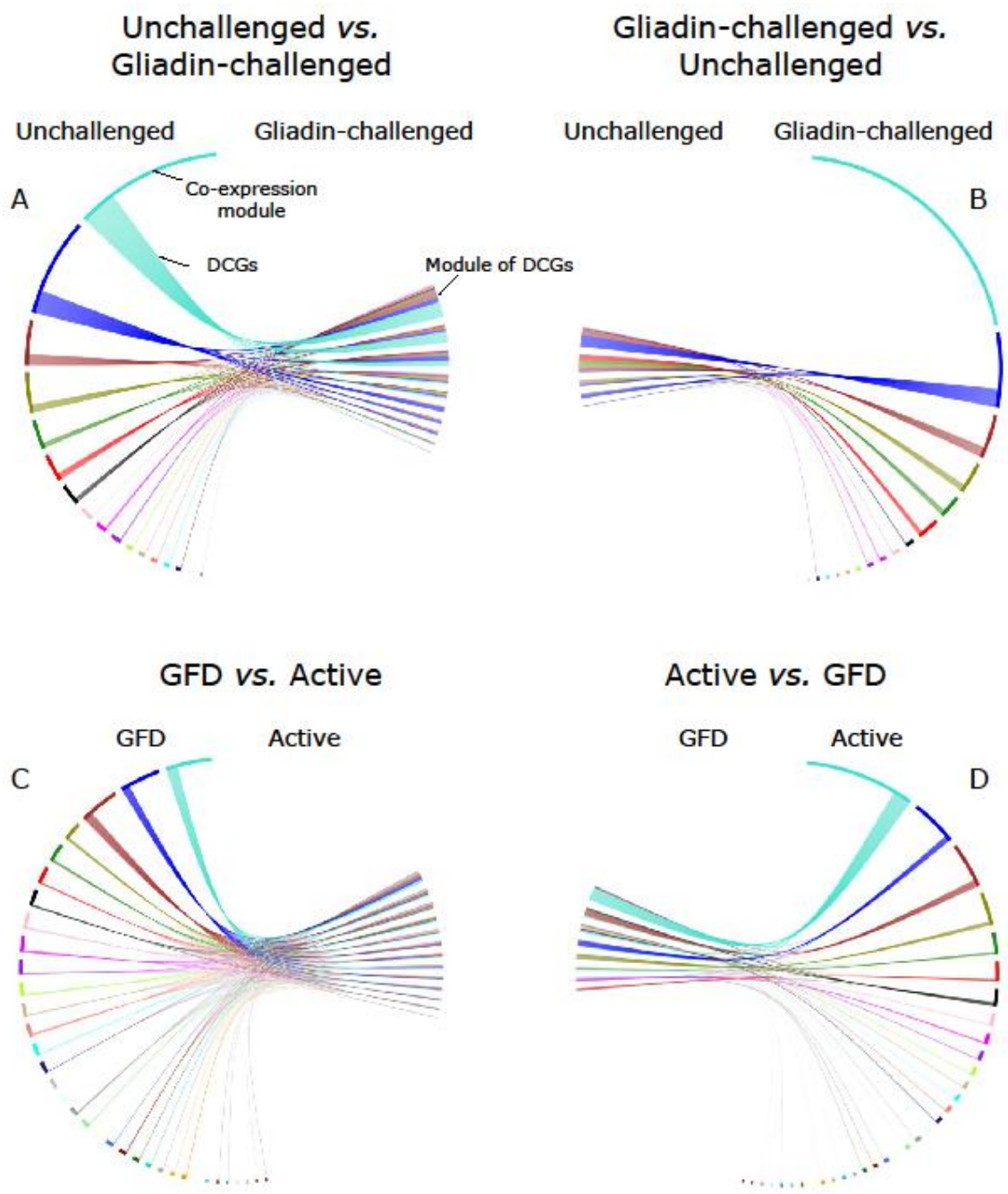


Figure 15. DCGs in the A,B) acute and C,D) long-term experiments. Co-expression modules and DCGs (to the left in A and C, to the right in B and D), and the new co-expression modules formed by DCGs and represented by sections (to the right in A and C, to the left in B and D) are shown for each comparison. An example of each term is indicated in the figure.

1.2. Identification of regulatory elements involved in alterations of co-expression

DCGs and the modules of DCGs were further analyzed to ascertain whether there was any overrepresentation of putative regulatory elements that could be involved in the alterations of co-expression.

Regarding TFs, in the acute experiment three modules showed TFBS enrichment (Table 11): the DCGs of the Magenta module in unchallenged biopsies were enriched for HOXA5 binding sites (Appendix 5); the DCGs of the Purple module in gliadin-challenged biopsies showed increased number of HOXA5, IRF1 and NFKB1 binding sites (Appendix 6); and the Yellow module of DCGs in gliadin-challenged biopsies showed enrichment for HOXA5 binding sites (Appendix 6). In the long-term experiment (Table 11), DCGs of the Brown module in GFD samples showed binding site enrichment for GFI1, FOXI1, ELK1, CREB1, GABPA, HOXA5, MYF and RORA (Appendix 7).

Regarding miRNAs, in the acute experiment, three modules showed miRNA enrichment (Table 12): DCGs of the Turquoise module in unchallenged biopsies (Appendix 5) and the Brown and the Turquoise co-expression modules of DCGs in unchallenged biopsies (Appendix 5). In the long-term experiment, nine modules showed miRNA enrichment (Table 12): DCGs of the Black and the Brown modules in GFD samples (Appendix 7), the Green and the Red modules of DCGs in GFD samples (Appendix 7), DCGs of the Black, the Blue and the Pink modules in active samples (Appendix 8), and the Brown and the Turquoise modules of DCGs in active samples (Appendix 8).

Table 11. Summary of the significant results of the TF enrichment analysis in the diffetent experiments and conditions, in DCGs and modules of DCGs ($P < 0.05$).

Experiment	Comparison	Observation	Module	TF	Function (according to OMIM and/or RefSeq)
Acute	Unchallenged *vs.* Gliadin-challenged	DCGs	Magenta	HOXA5	Tumorigenesis and regulation of p53.
	Gliadin-challenged *vs.* Unchallenged	DCGs	Purple	HOXA5	Tumorigenesis and regulation of p53.
				IRF1	Activation of genes induced by interferons α, β and γ. Apoptosis and tumor-suppression.
				NFKB1	Regulation of proinflammatory genes.
		Modules of DCGs	Yellow	HOXA5	Tumorigenesis and regulation of p53.
Long-term	GFD *vs.* Active	DCGs	Brown	GFI1	Control of histone modifications, CD4 T-cell proliferation.
				FOXI1	Development of the cochlea and vestibulum, and embryogenesis.
				ELK1	Smooth muscle differentiation and proliferation. Nuclear target for the ras-raf-MAPK cascade.
				CREB1	Induction of transcription of genes in response to hormonal stimulation of the cAMP pathway.
				GABPA	Regulation of cytochrome oxidase expression and nuclear control of mitochondrial function.
				HOXA5	Tumorigenesis and regulation of p53.
				MYF	Muscle development.
				RORA	Binding to hormone response elements and gene expression.

Table 12. Summary of the significant results of the miRNA enrichment analysis in the diffetent experiments and conditions, in DCGs and modules of DCGs ($P < 0.05$).

Experiment	Comparison	Observation	Module	miRNA
Acute	Unchallenged *vs.* Gliadin-challenged	DCGs	Turquoise	hsa-miR-369-3p, hsa-miR-937, hsa-miR-655, hsa-miR-324-3p, hsa-miR-891b, hsa-miR-323-5p, hsa-miR-18a-3p
		Modules of DCGs	Brown	hsa-miR-662
			Turquoise	hsa-miR-520h
Long-term	GFD *vs.* Active	DCGs	Black	hsa-miR-33a
			Brown	hsa-miR-92a, hsa-let-7b-3p, hsa-miR-507
		Modules of DCGs	Green	hsa-miR-551b-3p
			Red	hsa-miR-15b, hsa-miR-92a-2-3p, hsa-miR-663, hsa-miR-886-5p, hsa-miR-933, hsa-miR-15a, hsa-miR-16, hsa-miR-195, hsa-miR-615-5p, hsa-miR-650, hsa-miR-922, hsa-miR-503, hsa-miR-518c-3p, hsa-miR-423-3p
	Active *vs.* GFD	DCGs	Black	hsa-miR-26a, hsa-miR-26b
			Blue	hsa-let-7f, hsa-miR-613, hsa-miR-520b, hsa-miR-520c-3p, hsa-miR-518b, hsa-miR-518c, hsa-miR-518d-3p
			Pink	hsa-miR-886-5p, hsa-miR-486-5p
		Modules of DCGs	Brown	hsa-miR-302b, hsa-miR-302c, hsa-miR-518c, hsa-miR-520a-3p, hsa-miR-520b, hsa-miR-520c-3p, hsa-miR-520d-3p, hsa-miR-520e, hsa-miR-134, hsa-miR-933
			Turquoise	hsa-miR-674

1.3. Selection of candidate regulators and TF-target genes for downstream analyses

Selection of candidate regulators for experimental validation was done according to repetition of terms and previous literature as indicated in the *Material and Methods* section.

Several of the selected candidates had been previously related to inflammation, like IRF1 (Guo et al., 2010) and CREB1 (Wen et al., 2010) or to other pathways associated with CD, including the Toll-like receptor pathway, as hsa-miR-92a (Lai et al., 2013) and hsa-let-7b-3p (Teng et al., 2013); the NFκB pathway, as for example hsa-miR-33a (Kuo et al., 2013), hsa-miR-503 (Zhou et al., 2013), hsa-miR-18a-3p (Trenkmann et al., 2013), hsa-miR-26b (Zhao et al., 2014), hsa-miR-520e (Zhang et al., 2012) and NFKB1 (Fernandez-Jimenez et al., 2014); and Tight Junctions, like ELK1 (Al-Sadi et al., 2013). Other candidates, namely hsa-miR-655 and the hsa-miR-520b, target *MAGI2* (Kitamura et al., 2014) and *MICA* (Yadav et al., 2008), respectively, two genes that have been associated with CD risk (Martin-Pagola et al., 2003; Wapenaar et al., 2016). On the other hand, significant enrichment for HOXA5 targets was observed in several modules of both the acute and the chronic response experiments.

Regarding TF-target gene selection, the following target genes were selected according to the criteria defined in the *Material and Methods* section (Table 13):

Table 13. Selected TF-target genes for downstream experiments according to the criteria described previously.

		Target genes from DCG sets				Target genes not present in DCGs sets
		Conserved binding sites	ChIP-seq	H3K27Ac	Literature	Specificity*
IRF1 target	*CISD2*	x	x	x		
	HDAC4	x	x	x		
	WDR43	x	x	x		
	CXCL11					9
	BATF2					11
CREB1 target	*AKTIP*		x	x	Other diseases (Liang et al., 2012)	
	NAMPT		x	x	Inflammation (Garten et al., 2011)	
	TPK1		x	x	Other diseases (Banka et al., 2014)	
	ISG15					4
	HIST1H4					14
ELK1 target	*CRTAM*		x	x	Differentiation of T cells (Takeuchi et al., 2016)	
	PLLP	x	x	x	Differentiation of epithelial cells (Rodríguez-Fraticelli et al., 20	
	RFX5	x	x	x	Activation of MHCII gene expression (Garvie et al., 2007)	
	NKG7					10
	RAB17					11
NFKB1 target	*WNT11*		x			
	GSTA4					7
	TFF1					8
	PLAUR					10

*Number of TFs that are regulating each target gene. Those that were under the regulation of the fewest TFs were regarded more specific, and selected.

1.4. Transcription factors

1.4.1. Biological functions of modules

To ascertain whether the DCGs and modules of DCGs that showed TFBS enrichment shared any kind of functional or biological significance, we performed GO term enrichment analyses. We did not observe any particular enrichment in the DCGs of the Magenta module (Figure 16A) or in the Yellow module of DCGs (Figure 16B), while DCGs from the Purple (Figure 16C) and the Brown (Figure 16D) modules showed several terms related to the disease, including lysosomal transport (Maiuri et al., 2010) (GO:0007041), apoptosis (GO:0006915), lymphocyte activation (GO:0046649) and immune system development (GO:0002520). When the DCGs were removed from their original modules, no significant GO terms were found in the remaining genes.

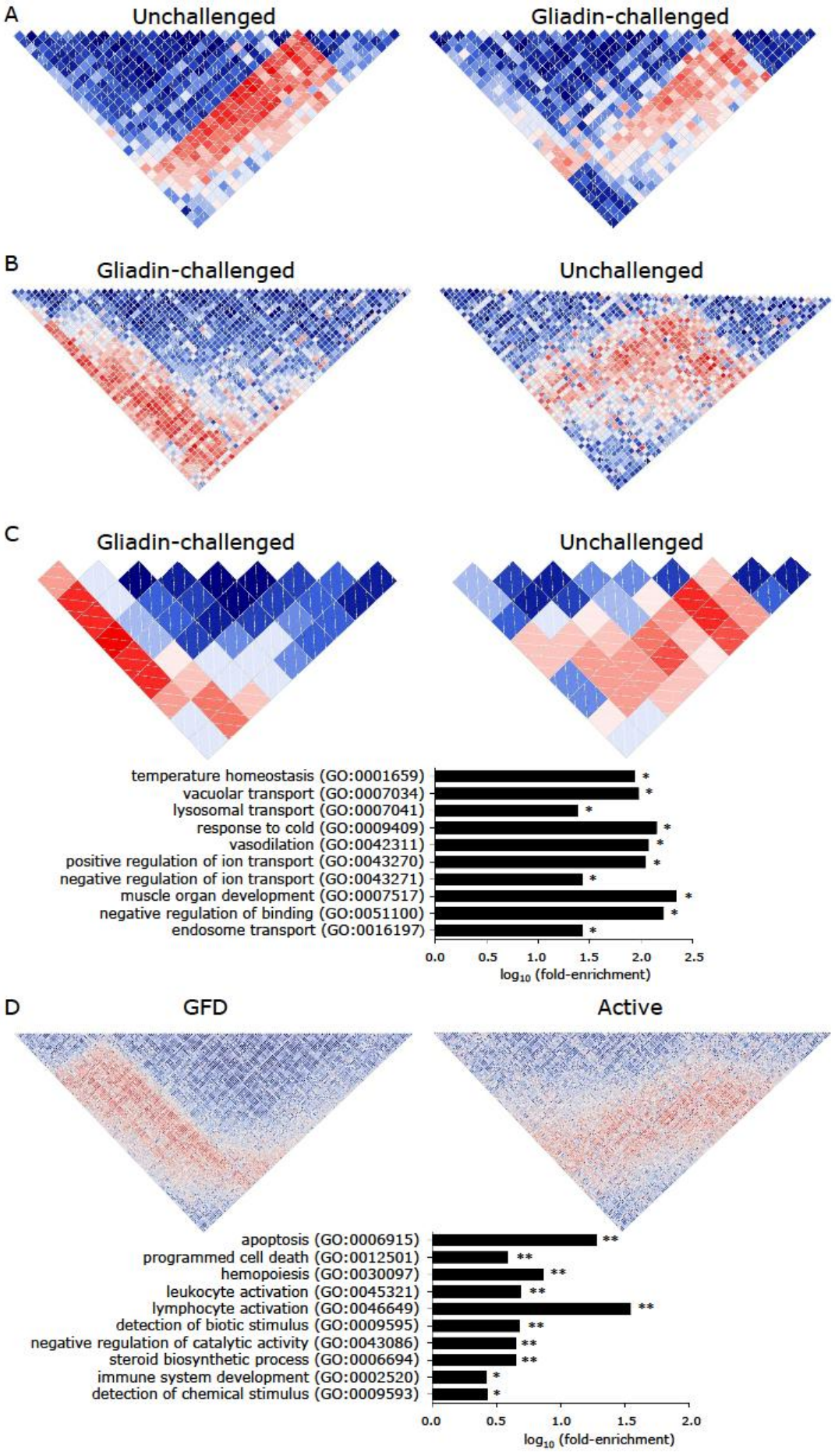

A
Unchallenged
Gliadin-challenged
B
Gliadin-challenged
Unchallenged
C
Gliadin-challenged
Unchallenged
temperature homeostasis (GO:0001659)
vacuolar transport (GO:0007034)
lysosomal transport (GO:0007041)
response to cold (GO:0009409)
vasodilation (GO:0042311)
positive regulation of ion transport (GO:0043270)
negative regulation of ion transport (GO:0043271)
muscle organ development (GO:0007517)
negative regulation of binding (GO:0051100)
endosome transport (GO:0016197)
0.0 0.5 1.0 1.5 2.0 2.5
log10 (fold-enrichment)
D
GFD
Active
apoptosis (GO:0006915)
programmed cell death (GO:0012501)
hemopoiesis (GO:0030097)
leukocyte activation (GO:0045321)
lymphocyte activation (GO:0046649)
detection of biotic stimulus (GO:0009595)
negative regulation of catalytic activity (GO:0043086)
steroid biosynthetic process (GO:0006694)
immune system development (GO:0002520)
detection of chemical stimulus (GO:0009593)
0.0 0.5 1.0 1.5 2.0
log10 (fold-enrichment)

Figure 16. Co-expression and GO annotation studies in DCGs. Co-expression pattern changes of DCGs in A) Magenta, B) Yellow, C) Purple and D) Brown modules. Each small square represents the Spearman correlation coefficient (ρ) between the expression levels of a specific gene pair (the red-blue scale represents positive to negative correlation). Below, the top 10 GO annotations for those DCGs compared to the whole-genome. GO terms are indicated in the Y axis. The X axis shows the log10(fold-enrichment) (ratio between the percentage of genes annotated with the GO term in the test set and the number of genes annotated with such term in the whole-genome (** $P < 0.01$ and * $P < 0.05$)).

1.4.2. Expression of candidate TFs and their target genes in CD

Expression analyses in an independent cohort of CD patients and controls showed that the candidate TFs *IRF1*, *ELK1*, *NFKB1* and *CREB1* were significantly upregulated in active disease when compared to GFD patients and controls. Additionally, out of the 19 target genes selected for those TFs, 17 were differentially expressed ($P < 0.05$) in active CD (Figure 17).

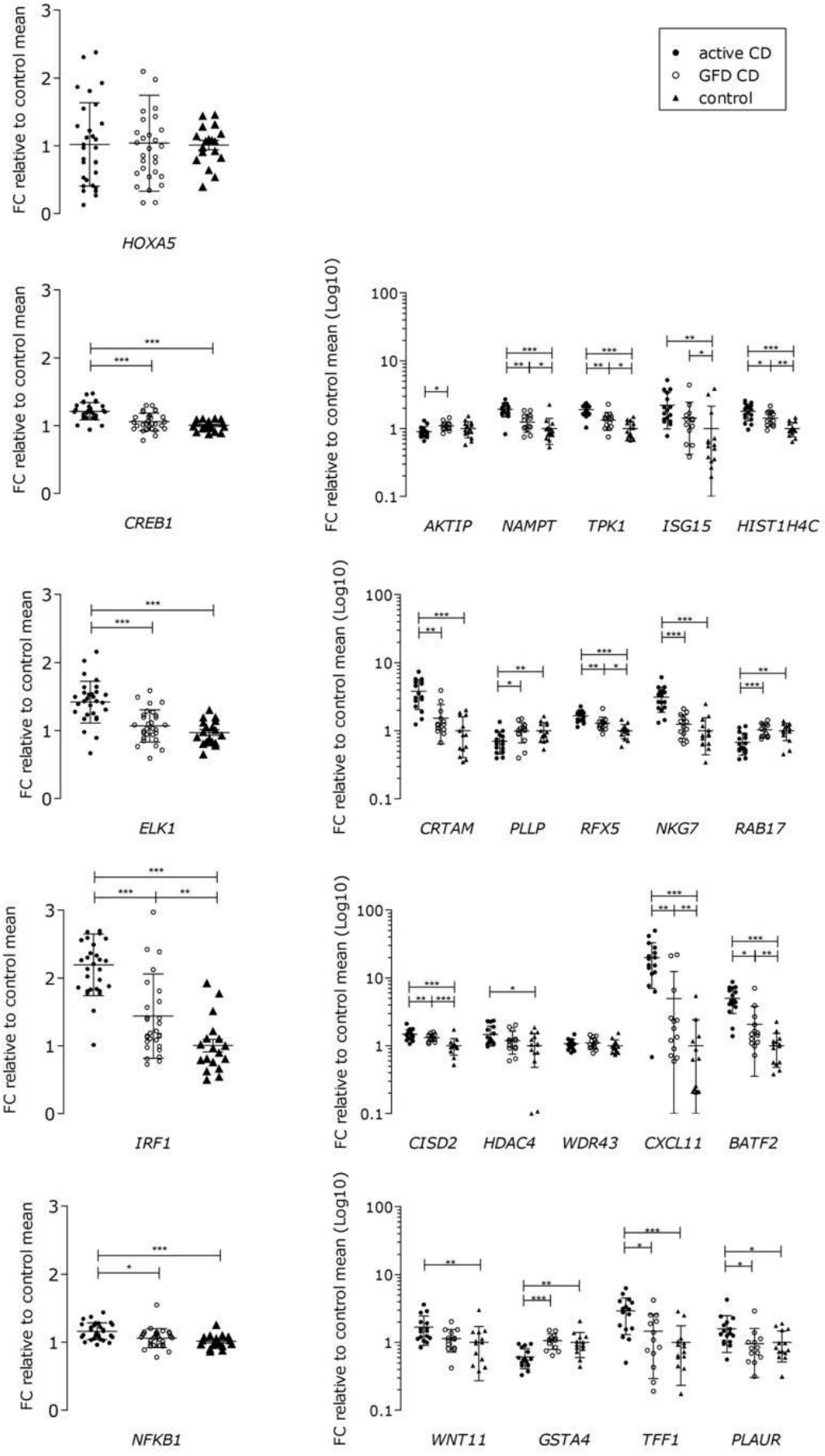
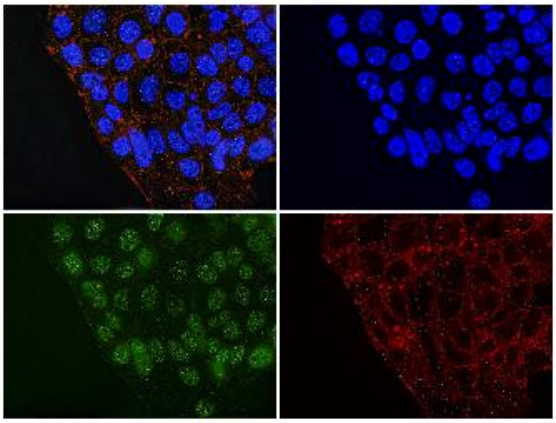

active CD
GFD CD
control
FC relative to control mean
FC relative to control mean (Log10)
HOXA5
CREB1
AKTIP
NAMPT
TPK1
ISG15
HIST1H4C
ELK1
CRTAM
PLLP
RFX5
NKG7
RAB17
IRF1
CISD2
HDAC4
WDR43
CXCL11
BATF2
NFKB1
WNT11
GSTA4
TFF1
PLAUR

Figure 17. Gene expression results of candidate TFs (active CD, n = 30; GFD CD, n = 29; control, n = 18) and their target genes (active CD, n = 16; GFD CD, n = 13; controls, n = 14) in duodenal biopsies. Data are expressed as mean ± SD (standard deviation) (*** $P < 0.001$, ** $P < 0.01$ and * $P < 0.05$).

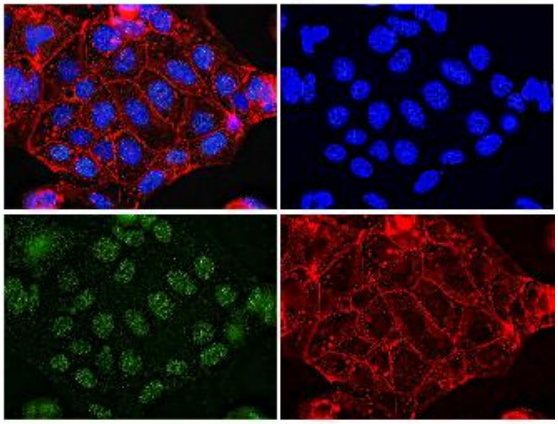
1.4.3. Cellular localization of candidate TFs in model cell line

Since an active TF must enter the nucleus to have a biological effect, we measured the localization of IRF1, NFKB1, ELK1 and CREB1 in C2BBe1 cells cultured with PT-G or PT-BSA for 4h. Immunostaining experiments (Figure 18) showed nuclear localization of the four TFs in both PT-G and PT-BSA cells, but no significant differences were observed between both conditions.

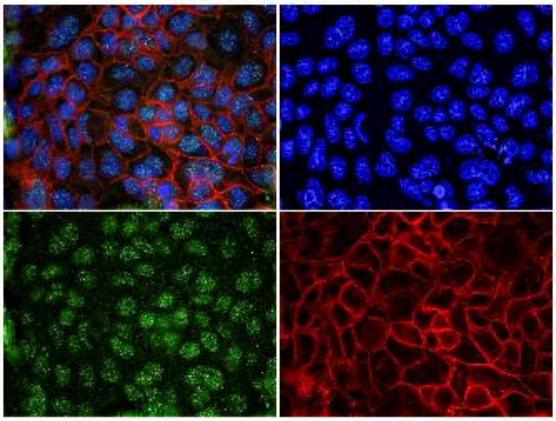
CREB1 PT-BSA exposition



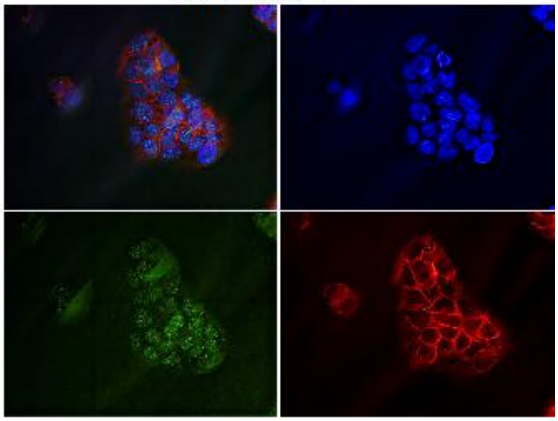
CREB1 PT-G exposition



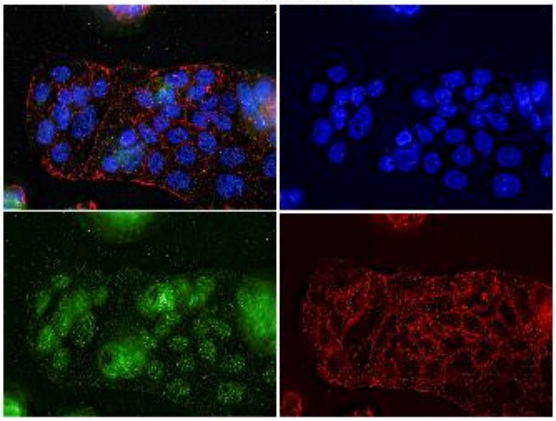
IRF1 PT-BSA exposition



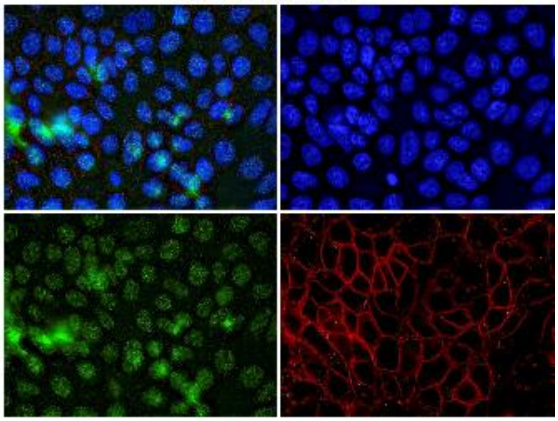
IRF1 PT-G exposition



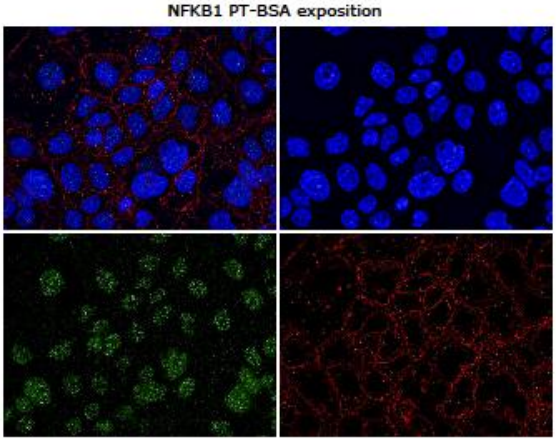
ELK1 PT-BSA exposition



ELK1 PT-G exposition



NFKB1 PT-BSA exposition



NFKB1 PT-G exposition

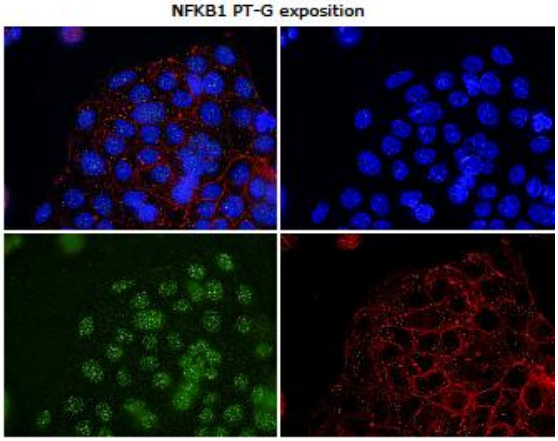


Figure 18. Immunofluorescence staining in C2BBe1 cells upon 4h PT-G/PT-BSA treatment. CREB1, IRF1, ELK1 and NFKB1 (green), DAPI (blue) and E-Cadherin (red).

We then investigated the localization of IRF1, NFKB1, ELK1 and CREB1 by Western blotting of nuclear and cytoplasmic fractions of C2BBe1 cells cultured with PT-G or PT-BSA during 4h, in order to observe whether an increment of nuclear levels of TFs occur upon PT-G stimulation. There were no differences between both conditions in the case of NFKB1 and ELK1, but PT-G induced a slight but consistent increase in the nuclear presence of both CREB1 and IRF1 (Figure 19).

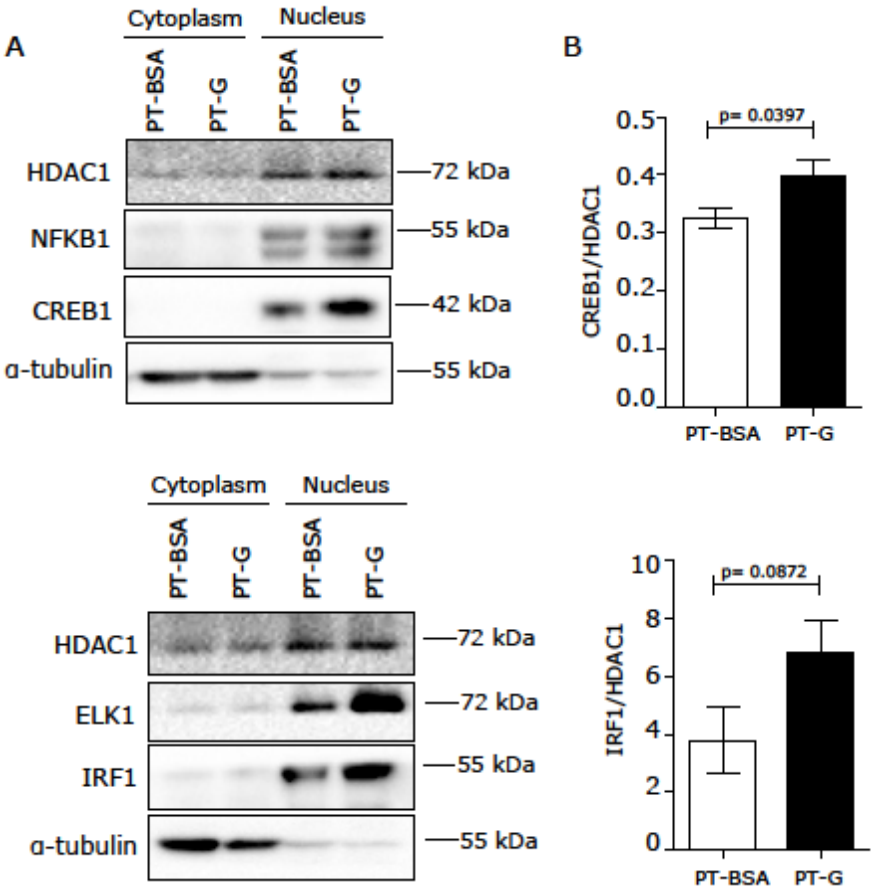


Figure 19. *In vitro* characterization of selected TFs. A) The translocation of NFKB1, CREB1, ELK1 and IRF1 to the nucleus was analyzed by Western blot in 3 independent experiments upon stimulation with pepsin-trypsin digested gliadin (PT-G) or pepsin-trypsin digested BSA (PT-BSA) as control. HDAC1 and α-tubulin were used as nuclear and cytoplamic controls, respectively; B) Band intensities were quantified and normalized to the intensity of the HDAC1 band . Mean values were compared with a t-test. Data are expressed as mean ± SEM.

1.4.4. Binding of candidate TFs to their target genes

We performed ChIP experiments in C2BBe1 cells treated with PT-G or PT-BSA for 4h to determine whether there was any change in binding upon gliadin challenge among IRF1 and CREB1 TFs to their target genes.

We could confirm *in vitro* that *HDAC4* and *WDRA3*, which were picked from a co-expression module, were indeed targets of IRF1, but there were no significant differences between PT-G and PT-BSA conditions. In the case of CREB1, the specific binding to the targets was as low as the control, unspecific binding, so no conclusions could be drawn (Figure 20).

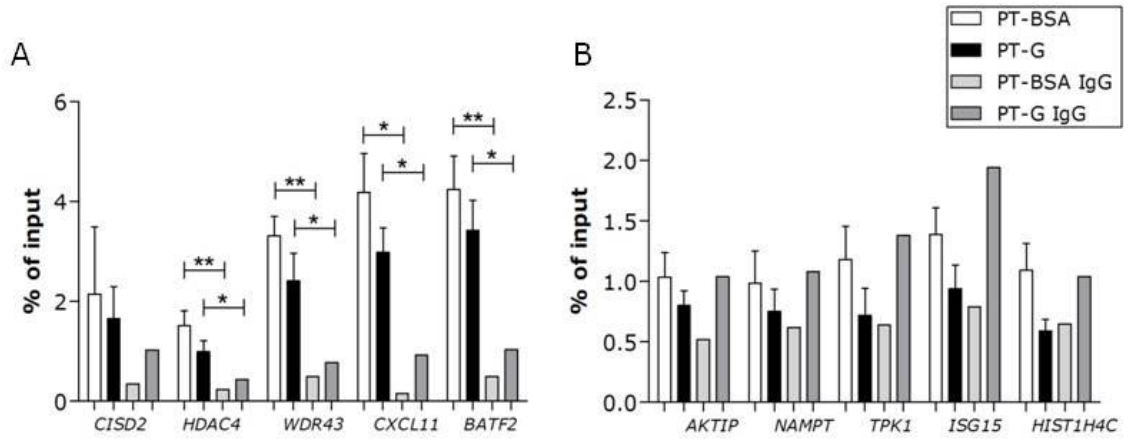


Figure 20. ChIP-qPCR validation for A) IRF1 and B) CREB1 target genes. ChIP assays were performed using anti-TF antibodies in C2BBe1 cells treated with 4h PT-G or PT-BSA. Mean ± SD for the ChIP elution and rabbit IgG (negative control) are shown as percentage of the input. **P < 0.01, *P < 0.05.

1.5. microRNAs

1.5.1. Expression of candidate miRNAs in CD

We quantified the pri-miRNAs of the 9 candidate miRNAs as a measure of their expression at the gene level. Six pri-miRNAs were downregulated in active CD

patients compared to non-celiac controls. These pri-miRNAs remained downregulated in patients on GFD (Figure 21).

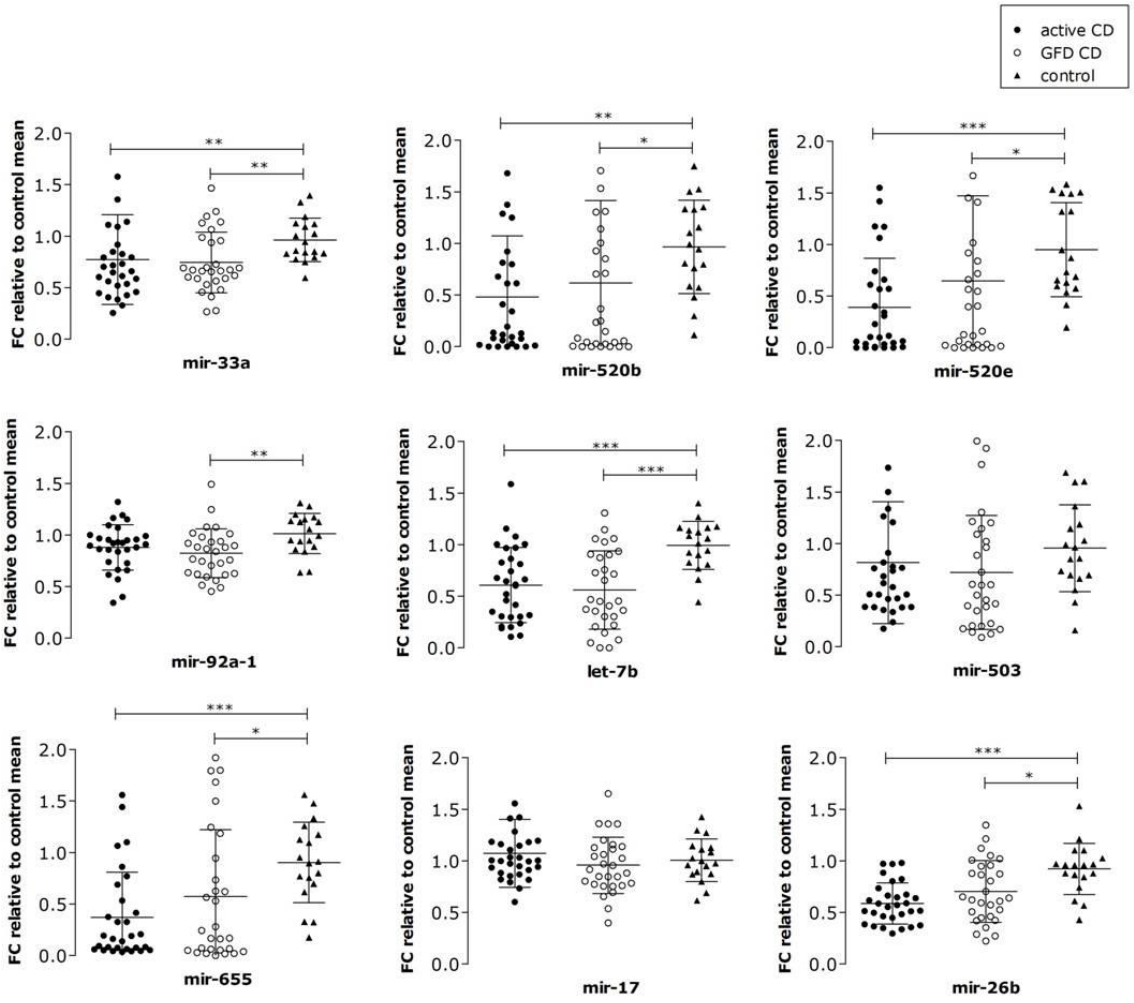


Figure 21. Gene expression analysis of of pri-miRNAs in duodenal biopsies from active (n = 29) and GFD treated (n = 29) CD patients and non-celiac controls (n = 18). Data are mean ± SD. (***P < 0.001, **P < 0.01 and *P < 0.05).

We also analyzed the expression of two mature miRNAs to check for concordance with the pri-miRNA results. The mature miRNA has-miR-18-3p, whose pri-miRNA (has-mir-17) was unaltered in CD, did not show any change. On the other hand, has-miR-26b, whose primary miRNA was significantly downregulated in active CD compared to both controls (P = 0.001) and GFD (P = 0.0138), was also less abundant in active disease compared to non-celiac controls (P = 0.0025) and GFD (P = 0.0125) (Figure 22).

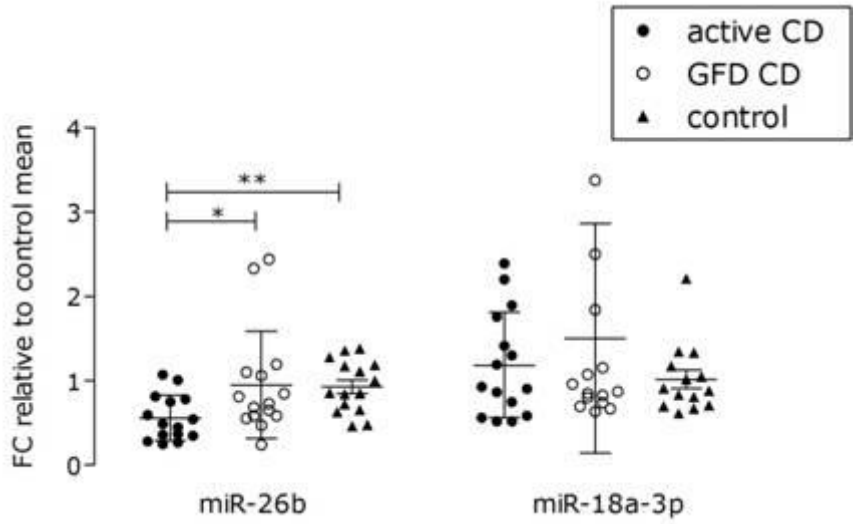


Figure 22. Gene expression analysis of two mature miRNAs in duodenal biopsies from active (n = 15) and GFD treated (n= 15) CD patients and non-celiac controls (n = 15). Data are mean ± SD. (**P <0.01 and *P <0.05).

1.5.2. Expression of miRNA target genes in CD

Expression of miRNA target genes was interrogated in the total RNA-seq data from the epithelial and immune cell-enriched fractions from active CD patients and non-celiac controls.

In the case of hsa-miR-26b, seven of the 35 target genes identified in the MirTarBase database were differentially expressed ($P < 0.05$) in the epithelial fraction (Figure 23A), while in the immune fraction five target genes showed changes in expression (Figure 23B).

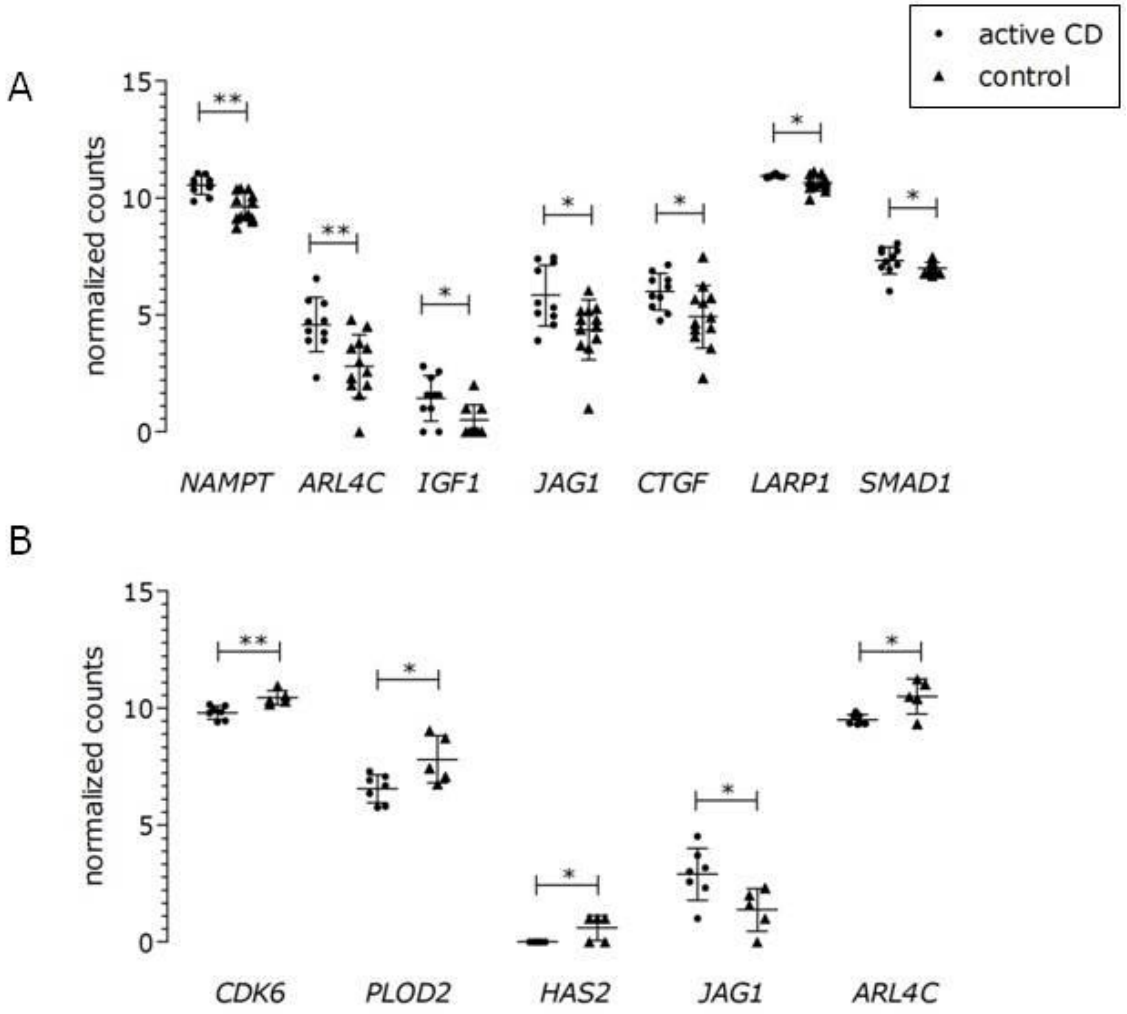


Figure 23. Expression analysis of target genes for hsa-miR-26b in biopsies from A) epithelial and B) immune fractions of active CD patients and non-celiac controls. Data are mean ± SD. (**P < 0.01 and *P < 0.05).

Regarding has-miR-33a, two of the 34 target genes identified in the database were differentially expressed (P < 0.05) in active CD in both epithelial and immune fractions (Figure 24).

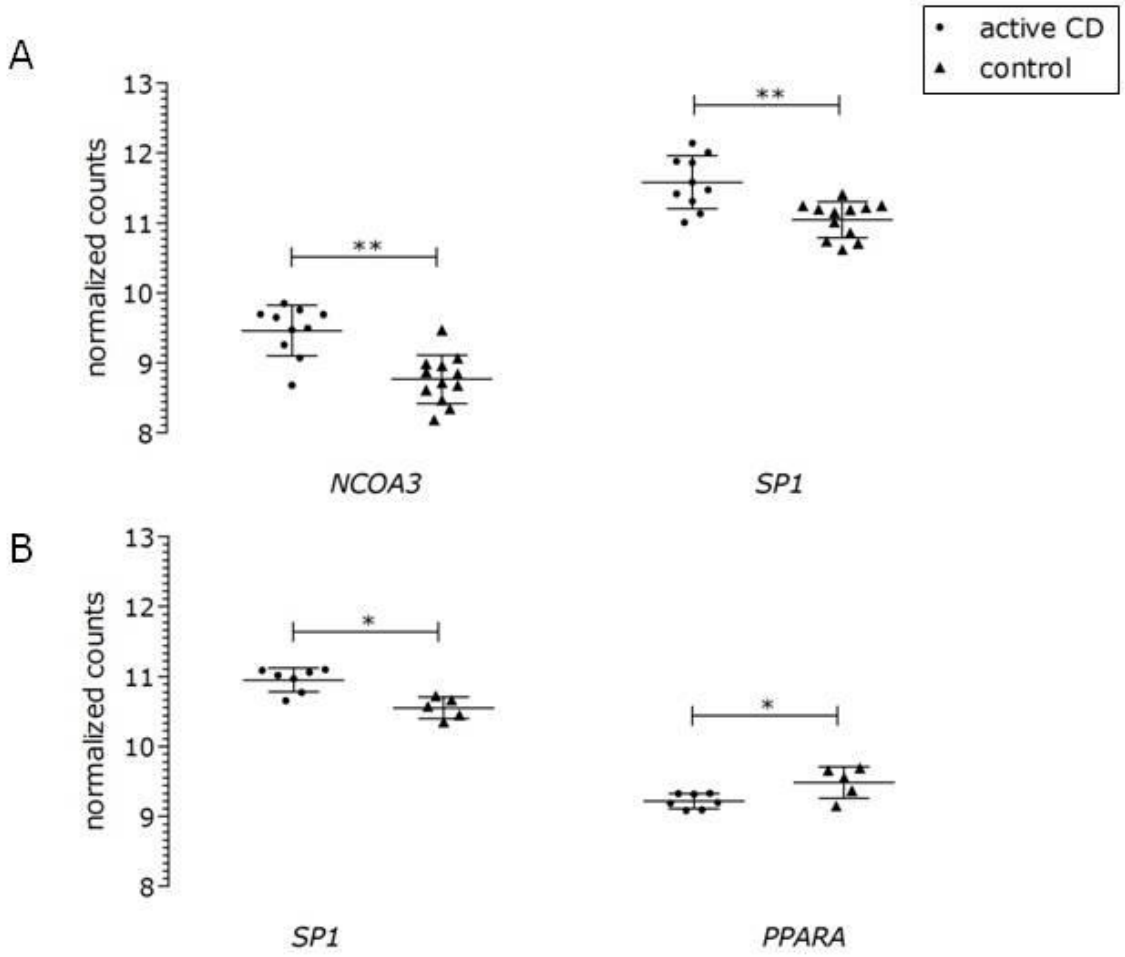


Figure 24. Expression analysis of target genes for hsa-miR-33a in biopsies from A) epithelial and B) immune fractions of active CD patients and non-celiac controls. Data are mean ± SD. (**P < 0.01 and *P < 0.05).

Regarding has-let-7b-3p, four of the 36 target genes identified in the database were differentially expressed (P < 0.05) in active CD epithelial fraction (Figure 25A), while in the immune fraction 2 were differentially expressed (Figure 25B).

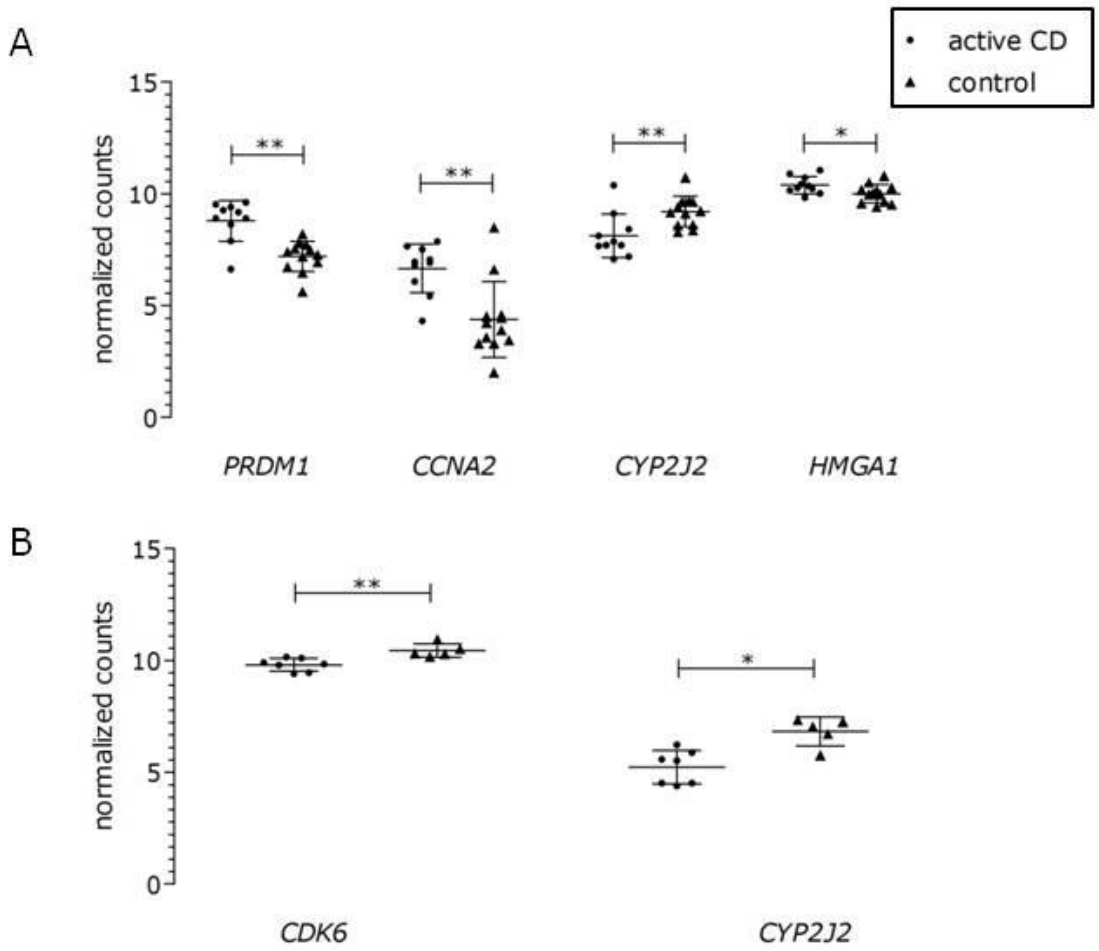


Figure 25. Expression analysis of target genes for hsa-let-7b-3p in biopsies from A) epithelial and B) immune fractions of active CD patients and non-celiac controls. Data are mean ± SD. (**P < 0.01 and *P < 0.05).

Expression of one of the 9 targets for hsa-miR-520b was altered in epithelial fractions of active CD patients, as well as two targets out of the six identified in the case of hsa-miR-520e, and 2 out of the 6 identified in the case of has-miR-655 (Figure 26).

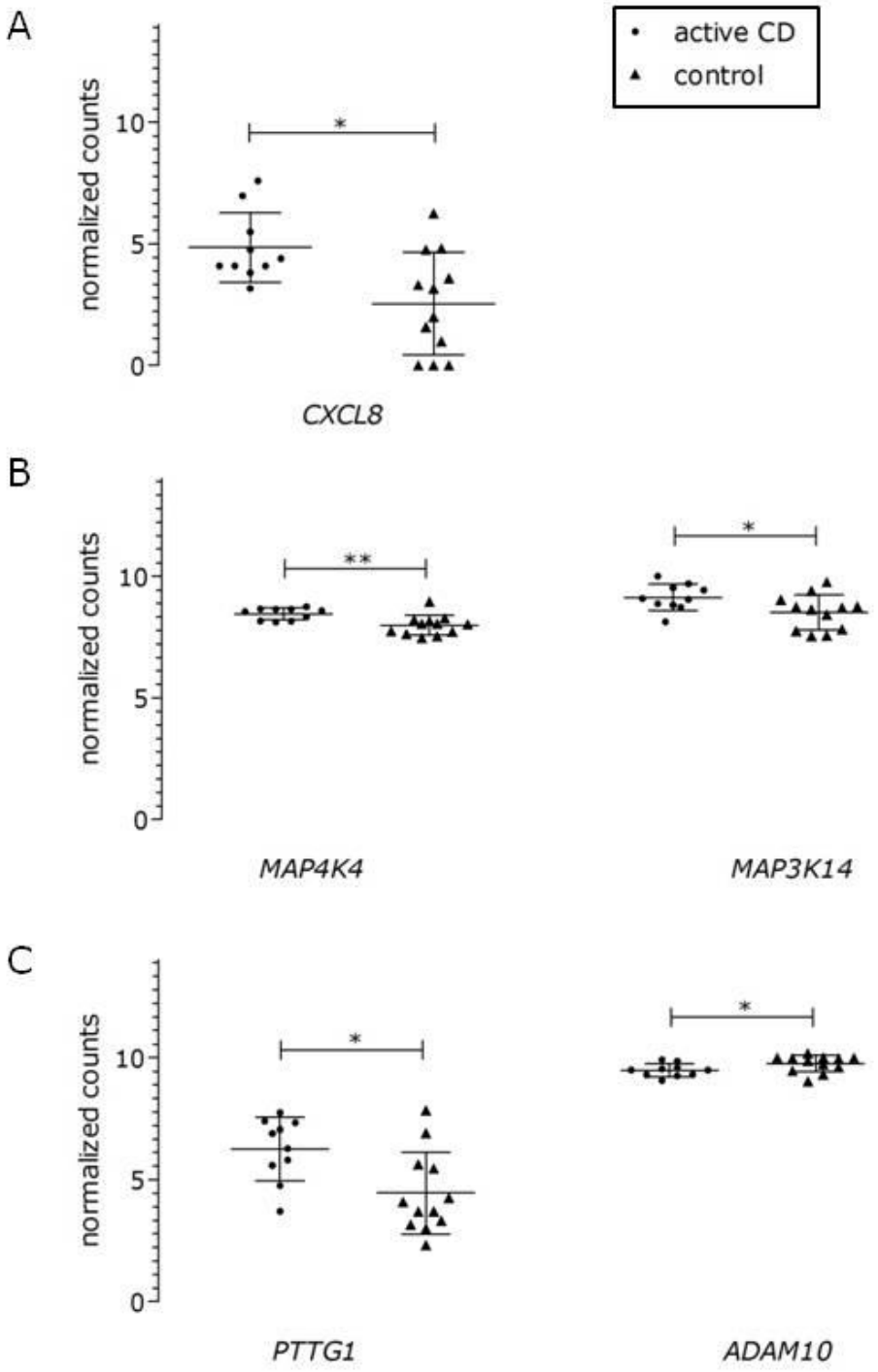


Figure 26. Expression analysis of target genes for A) hsa-miR-520b, B) hsa-miR-520e and C) has-miR-655 in biopsies from epithelial fractions of active CD patients and non-celiac controls. Data are mean ± SD. (**P < 0.01 and *P < 0.05).

2. Topologically associating domains in CD

CD is a complex disease showing numerous alterations, including changes in expression and co-expression patterns, but the underlying mechanisms remain unclear. TADs are functional domains of gene expression coordination and could be involved in those changes.

2.1. Identification and characterization of candidate regions and genes

We hypothesized that co-expression changes among neighboring genes could be associated with alterations in TAD organization (Figure 27). Thus, we used RNA-seq data from the epithelial-cell fraction to build co-expression patterns, and checked whether co-expressed genes overlapped with conserved TADs. In healthy individuals 2,826 genes from 11 co-expression modules were interrogated, and 739 genes were located in 486 TADs. In the case of CD active patients, 3,992 genes from 18 co-expression modules were studied; and 613 of those genes were located in 430 TADs.

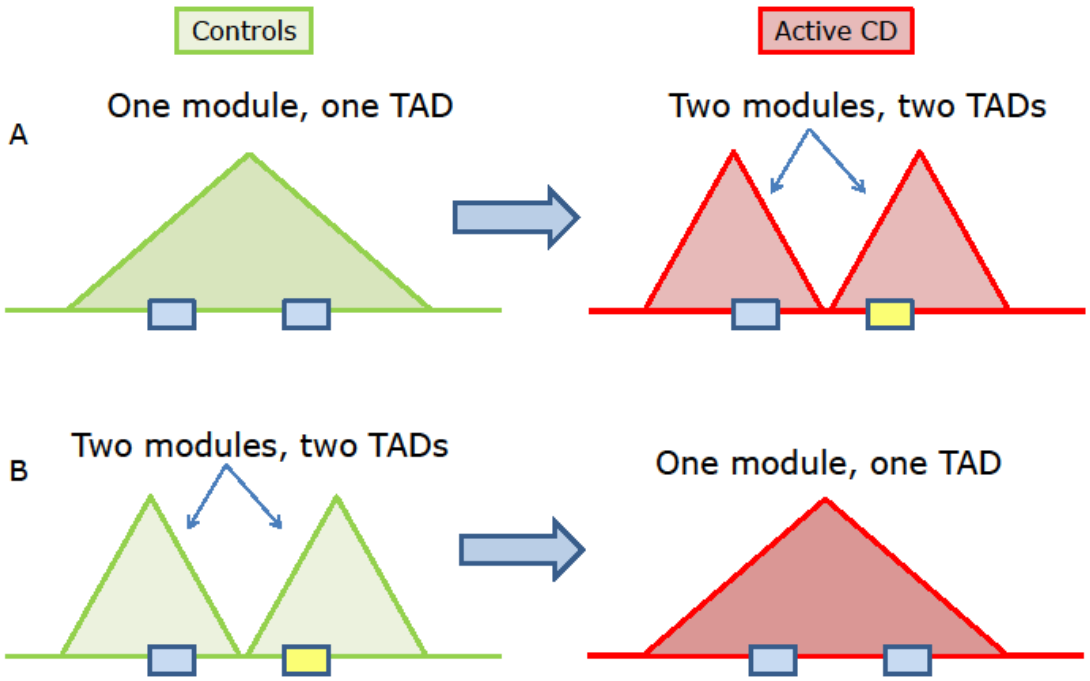


Figure 27. Hypothetical models of TAD alterations in CD. Possible cases of A) disruption of TADs and B) merge of TADs are shown. In the first case two neighboring genes are located in a

TAD and are co-expressed in controls, while they are not co-expressed in CD, suggesting TAD disruption. In the second, two neighboring genes are located in different TADs and are not co-expressed in controls, while they are in CD, suggesting a merge of TADs.

On the one hand, we identified 16 TADs in which changes in co-expression could indicate disruption of TADs. In these cases, co-expressed genes were located in the same TAD in controls, but divided into two different co-expression modules in active disease, suggesting a TAD disruption (Figure 27A) (Table 14). On the other hand, we also identified 30 putative TADs in which changes in co-expression could indicate a merge of TADs and consequent new co-expression. Those TADs are close to each other (less than 2 Mb) and harbor genes that are not co-expressed in controls, but are so in active CD (Figure 27B) (Table 15).

Table 14. Disruption of TADs and co-expression. Each row representas a TAD and shows the genes co-expressed in controls but not in CD, and the CD-associated SNPs located inside. Only the SNP with the lowest p values are listed, and the number of SNPs in each TAD is given in brackets.

TAD coordinates (hg19)	genes	top SNP	nominal p value
chr1:178680000-178760000	*RALGPS2; RP11-428K3.1*		
chr6:26040000-26160000	*HIST1H2BB; XLOC_059355*	rs3857546 (22)	4.35e-09
chr6:52600000-53120000	*GSTA2; GSTA1*		
chr7:16560000-16960000	*AC073333.1; AGR2*		
chr8:71720000-71800000	*XLOC_065435; XLOC_065445*		
chr8:86040000-86440000	*E2F5; CA3; CA2*		
chr9:131400000-131480000	*SET; XLOC_070707*		
chr10:81920000-82080000	*XLOC_009508; XLOC_009510*	rs12219032 (4)	0.005309
chr11:62320000-62480000	*UQCC3; EEF1G*		
chr12:13240000-13360000	*XLOC_014371; XLOC_014372; XLOC_014373*		
chr15:40520000-40800000	*DISP2, PHGR1; IVD*		
chr16:67880000-67960000	*XLOC_025619; XLOC_027035*		
chr2:130840000-131240000	*MZT2B; CCDC115*		
chr22:29040000-29360000	*HSCB; XBP1*	rs2097461 (3)	0.02041
chr22:30000000-30240000	*NF2; UQCR10*	imm_22_28471822 (15)	0.0004322
chr22:37360000-37520000	*MPST; TST*	rs5995385 (1)	0.03635

**XLOC*: unannotated expressed signals.

Table 15. Merge of TADs and co-expression. Each row represents a putative merge of two TADs in disease, and shows the genes co-expressed in CD patients but not in controls, and the CD-associated SNPs located inside. Only the SNPs with the lowest p values are listed, and the number of SNPs in each TAD is given in brackets.

TAD A + B (hg19)	TAD A (hg19)	gene A	TAD B (hg19)	gene B	top SNP	nominal p value
chr1:178400000-178760000	chr1:178400000-178520000	*XLOC_006089*	chr1:178680000-178760000	*RALGPS2*		
chr2:134280000-135000000	chr2:134280000-134400000	*NCKAP5*	chr2:134880000-135000000	*XLOC_037431*	rs10207341 (9)	0.0006984
chr3:23960000-25680000	chr3:23960000-24040000	*RPL15*	chr3:25560000-25680000	*RARB*		
chr3:32680000-33960000	chr3:32680000-33160000	*CRTAP*	chr3:33800000-33960000	*RP11-10C24.1*	rs9867080 (1)	0.03304
chr6:144360000-146360000	chr6:144360000-144440000	*SF3B5*	chr6:146040000-146360000	*RP11-545I5.3*		
chr6:99760000-101480000	chr6:99760000-100040000	*COQ3*	chr6:101000000-101480000	*ASCC3*	rs924974 (2)	0.0136
chr8:41480000-42400000	chr8:41480000-41640000	*NKX6-3*	chr8:42320000-42400000	*RP11-503E24.2*		
chr11:33680000-34520000	chr11:33680000-33840000	*RP4-541C22.5 CD59 C11orf91*	chr11:34320000-34520000	*ABTB2*		
chr11:34320000-34920000	chr11:34320000-34520000	*ABTB2*	chr11:34840000-34920000	*APIP*		
chr11:34320000-36840000	chr11:34320000-34520000	*ABTB2*	chr11:36440000-36840000	*C11orf74*	rs4755450 (10)	0.000522
chr14:74600000-76440000	chr14:74600000-74680000	*LIN52*	chr14:75920000-76440000	*C14orf1*	rs1569328 (4)	0.0003735
chr15:63320000-65040000	chr15:63320000-63440000	*RP11-244F12.3*	chr15:64920000-65040000	*OAZ2*	rs4411464 (2)	0.008319
chr15:71960000-72480000	chr15:71960000-72080000	*THSD4 RP11-592N21.2*	chr15:72400000-72480000	*GRAMD2*		
chr17:18840000-20400000	chr17:18840000-19120000	*SNORD3A*	chr17:20160000-20400000	*LGALS9B*	rs2703817 (1)	0.03975
chr17:18840000-20400000	chr17:18840000-19120000	*SNORD3C*	chr17:20160000-20400000	*LGALS9B*	rs2703817 (1)	0.03975
chr17:3480000-5080000	chr17:3480000-3720000	*RP11-235E17.4*	chr17:4680000-5080000	*RP5-1050D4.4*	rs9906760 (8)	0.001804
chr17:40520000-42160000	chr17:40520000-40680000	*STAT3*	chr17:41720000-42160000	*TMEM101*	imm_17_38056077 (77)	7,51E-02
chr17:40920000-42160000	chr17:40920000-41080000	*G6PC*	chr17:41720000-42160000	*HDAC5*	rs382571 (4)	0.005572
chr17:41720000-43320000	chr17:41720000-42160000	*HDAC5*	chr17:43040000-43320000	*PLCD3 MIR6784*	rs9903582 (1)	0.0114
chr17:9360000-10800000	chr17:9360000-9800000	*USP43*	chr17:10440000-10800000	*XLOC_027658*	rs9903582 (1)	0.0114
chr17:9360000-10800000	chr17:9360000-9800000	*USP43*	chr17:10440000-10800000	*TMEM220-AS1*		
chr19:12880000-14160000	chr19:12880000-13000000	*RNASEH2A*	chr19:14040000-14160000	*CTB-55O6.13*	rs416162 (1)	0.04632
chr19:13800000-14520000	chr19:13800000-13880000	*XLOC_032850*	chr19:14400000-14520000	*ADGRE5*	rs416162 (1)	0.04632
chr19:51840000-53160000	chr19:51840000-51960000	*XLOC_033809*	chr19:53080000-53160000	*ZNF83*	rs16982743 (3)	0.03813
chr20:32800000-33800000	chr20:32800000-32920000	*RP4-785G19.5 AHCY CTD-3216D2.5*	chr20:33480000-33800000	*PROCR*	rs6059916 (1)	0.03293
chr20:33480000-34320000	chr20:33480000-33800000	*PROCR*	chr20:34160000-34320000	*ROMO1*	rs224436 (11)	0.003579
chr22:41200000-43640000	chr22:41200000-41400000	*RBX1*	chr22:43400000-43640000	*TSPO*	rs5758209 (2)	0.004118
chr22:41200000-43640000	chr22:41200000-41400000	*RBX1*	chr22:43400000-43640000	*MCAT*	rs5758209 (2)	0.004118
chr22:41800000-43640000	chr22:41800000-41960000	*XLOC_045800*	chr22:43400000-43640000	*BIK*		
chr22:41800000-43640000	chr22:41800000-41960000	*XLOC_045800*	chr22:43400000-43640000	*XLOC_045853*		

**XLOC*: unannotated expressed signals.

We then searched for SNPs associated with CD with a nominal p value below 0.05 that could affect TAD structures and explain changes in co-expression, and thus we crossed TAD and CD-associated SNP coordinates. We localized 434 Immunochip SNPs grouped into 13 regions overlapping a TAD, out of which 45 were significantly associated with CD, suggesting that they participated in the disruption of TAD organization (Table 14). We localized 916 SNPs into 25 regions overlapping a putative merge of adjacent TADs, out of which 142 were significantly associated (Table 15).

Two of these regions were selected for further characterization because they could be implicated in co-expression changes of annotated genes, rather than unannotated XLOC signals, and a set of SNPs was identified within them, rather than an isolated SNP. On the one hand, the *HSCB-XBP1* region was located in hg19 chr22:29040000-29360000 and could be implicated in the disruption of a TAD, provoking co-expression alterations between the two genes. On the other hand, the *PROCR-ROMO1* region was located in hg19 chr20:33480000-34320000 and could be implicated in the merge of two adjacent TADs, altering co-expression between *PROCR* and *ROMO1*.

2.2. Chromatin accessibility of the identified regions

We next sought to identify DNase I hypersensitive sites harboring CTCF binding motifs, in order to determine active chromatin regions that could be TAD boundaries. In partcular, hg19 chr22:29186201-29186490 inside the *HSCB-XBP1* region, and hg19 chr20:34026861-34027090 within the *PROCR-ROMO1* region were identified as DNase I hypersensitive sites (Figures 28, 29).

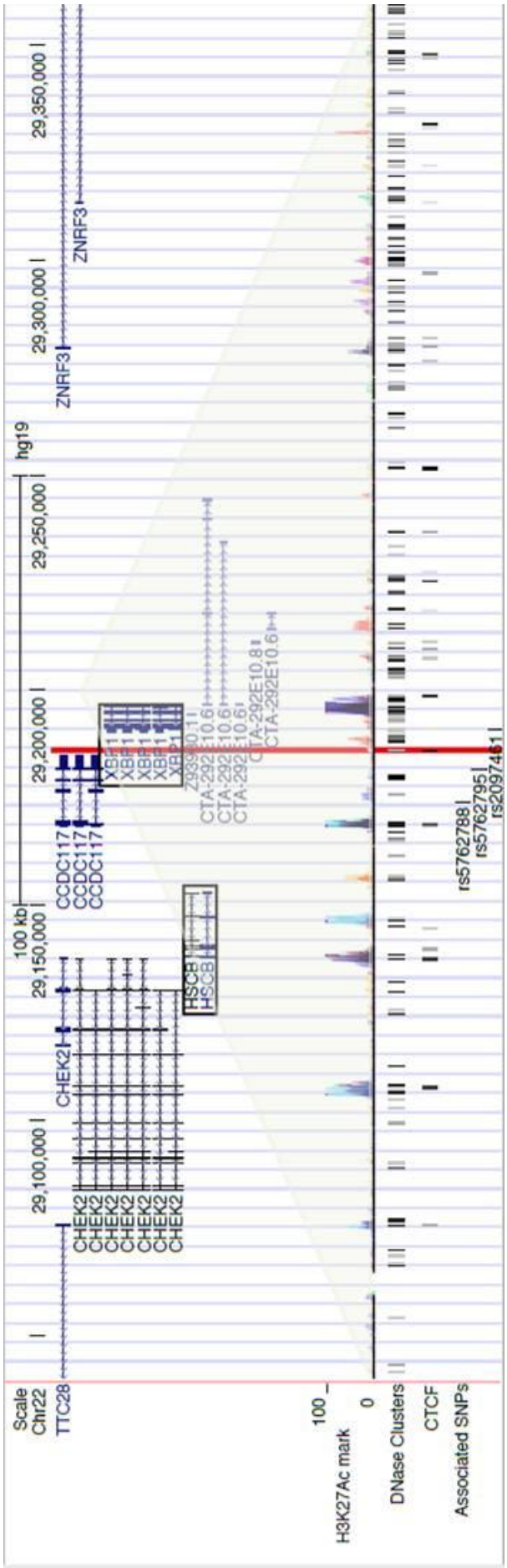


Figure 28. Candidate region *HSCB-XBP1* (a disrupted TAD). The black rectangles surround genes that change co-expression in CD; SNPs associated to CD that could alter TAD formation are shown. The red line indicates the DNase I hypersensitive site harboring CTCF binding sites. The light green triangle spans the TAD in controls.

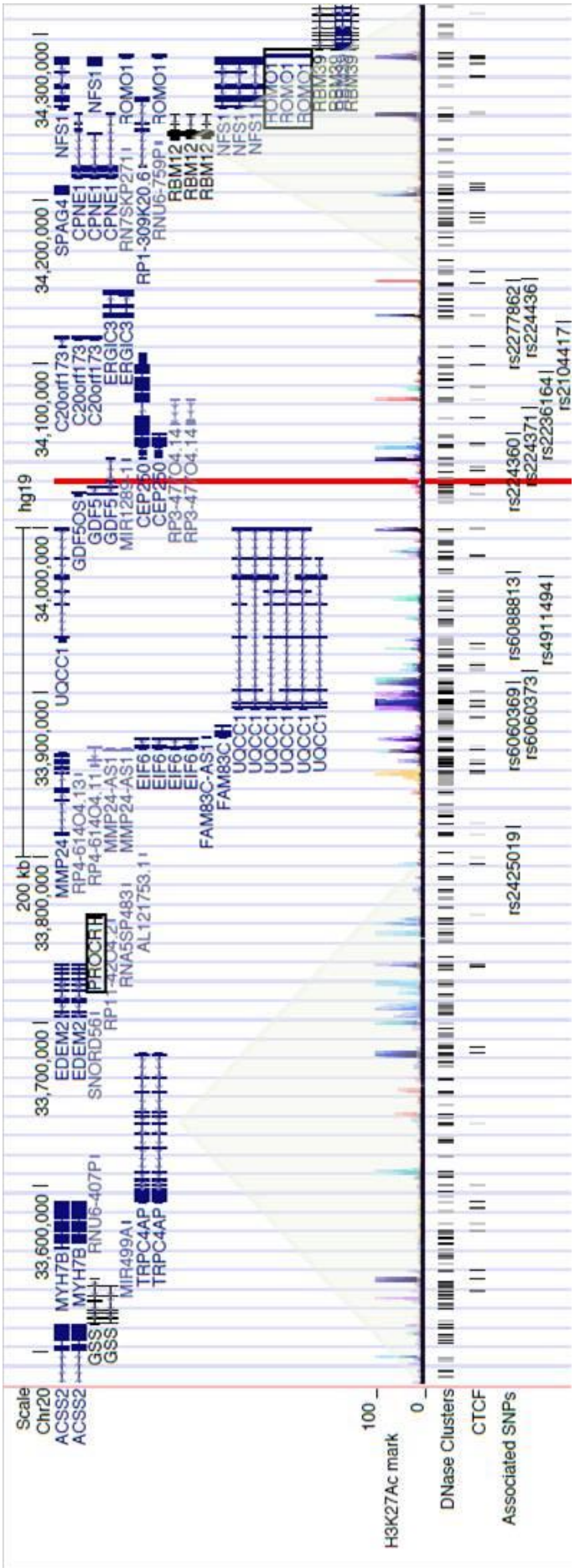


Figure 29. Candidate region *PROCR-ROMO1* (a merge of two TADs). The black rectangles surround genes that change co-expression in CD; SNPs associated to CD that could alter TAD formation are shown. The red line indicates the DNase I hypersensitive site harboring CTCF binding sites. The light green triangles span the two adjacent TADs in controls.

Quantitative PCR of DNase digested chromatin was performed in HCT116, HCT15 and HEK293FT cell lines to determine whether the identified DNase I hypersensitive sites were accessible to DNase in those cells. Both sites showed high DNase accessibility in HCT116 and HEK293FT cell lines, indicative of euchromatin structure; whereas the DNase I hypersensitive site inside the *PROCR-ROMO1* region but not the one inside the *HSCB-XBP1* region showed high DNase accessibility in HCT15 cells (Figure 30).

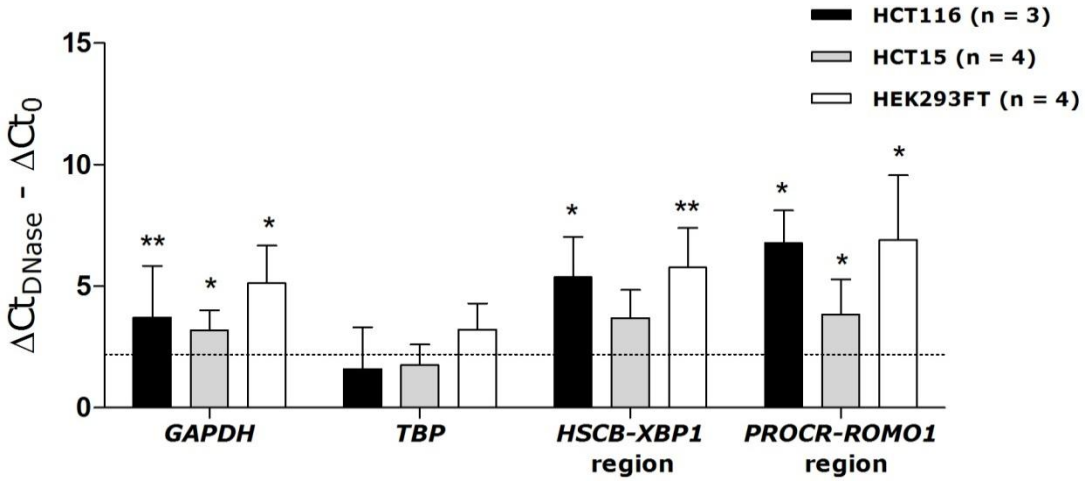


Figure 30. DNase accessibility. Cultured HCT116, HCT15 and HEK293FT cells were subjected to chromatin digestion by DNase, followed by quantitative PCR. DNase accessibility (chromatin accessibility) was quantified using the shift in Ct values between digested and undigested chromatin. Candidate regions were compared to closed (*TBP*) chromatin control in each cell line. Data are mean ± SD (**P < 0.01 and *P < 0.05). Mean *TBP* value is represented with the dotted line.

2.3. Gene editing of selected regions

To elucidate the function of the regions identified, CRISPR-Cas9 was used to permanently disrupt the two DNase I hypersensitive sites harboring CTCF binding sites in the different epithelial cell models. We performed a 441 bp deletion inside the *HSCB-XBP1* region and a 377 bp deletion in the *PROCR-ROMO1* region.

2.3.1. Confirmation of the deletion

Appropriate sgRNAs were cloned in a Cas9 vector and recombinant plasmids were transfected into epithelial cell lines (HCT116, HCT15 and HEK293FT). Potential deletions were assessed by PCR using primers flanking the target region. The deletion in the *PROCR-ROMO1* region was observed in HCT15 and HEK293FT cells (Figure 31).

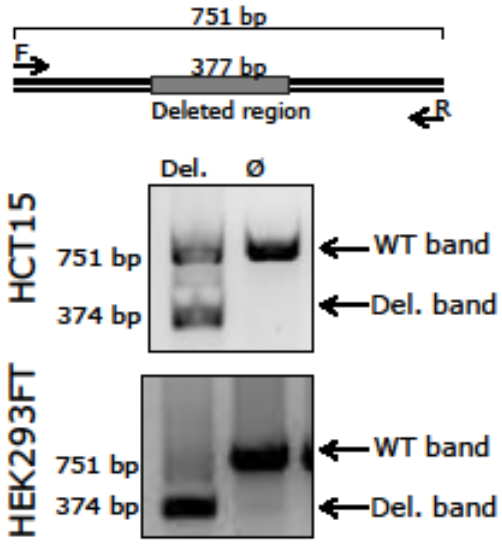


Figure 31. Deletion of DNase I hypersensitive sites harboring CTCF binding site within the *PROCR-ROMO1* region in human cells using CRISPR-Cas9 gene editing technology. Primers (F and R) flanking the target region were used to confirm the deletion. Del. band = deletion band; WT band = Wild type band; Ø = non-edited lines; Del. = edited lines.

As genome editing was not 100% efficient (WT band is present in all the cases), a mixed (heterogeneous) population of cells was obtained. We isolated single cells from this population for clonal expansion. PCR was performed to determine the presence of the deletion in the clonal cell lines (Figure 32). We were able to subclone the HEK293FT cells and obtain 6 homozygous clones for the deletion inside the *PROCR-ROMO1* region, but we did not obtain any intestinal clone with the deletion.

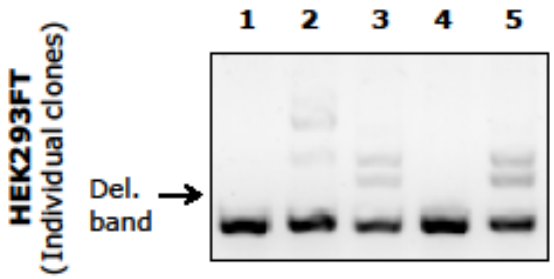


Figure 32. Example of homozygous (1, 4) and heterozygous (2, 3, 5) HEK293FT clones after targeted deletion with CRISPR-Cas9. Clonal lines harboring the targeted deletion were identified by PCR using primers flanking the deletion. Del. band = deletion band.

2.3.2. Expression and co-expression analysis

Expression and co-expression studies were performed to assess whether deletions affected the regulation of nearby genes. Wild-type (WT) and edited HEK293FT clones, as well as WT and mixed HCT15 populations were studied.

Regarding the expression of *PROCR* and *ROMO1*, no significant alterations were observed when HCT15 mixed populations were compared to WT cells (Figure 33A). Nevertheless, significant expression differences were observed in HEK293FT edited cells when compared to WT cells, both in *PROCR* and *ROMO1* (Figure 33B).

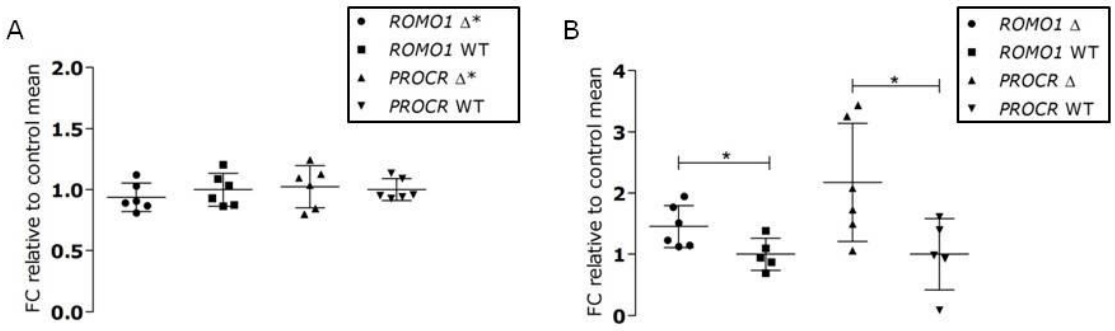


Figure 33. Expression analysis of *PROCR* and *ROMO1* in A) HCT15 and B) HEK293FT cell lines (n = 6 in all conditions). WT = Wild type cells; Δ = edited cells. *mixed population. Data are mean ± SD. (*P < 0.05).

When co-expression was studied, correlation between *PROCR* and *ROMO1* was not observed in WT HCT15 cells (Figure 34A), while the mixed population showed a significant correlation in expression (Figure 34B). In the case of HEK293FT cells, correlation between *PROCR* and *ROMO1* was observed in WT cells (Figure 34C), while co-expression was disrupted in the edited cells (Figure 34D).

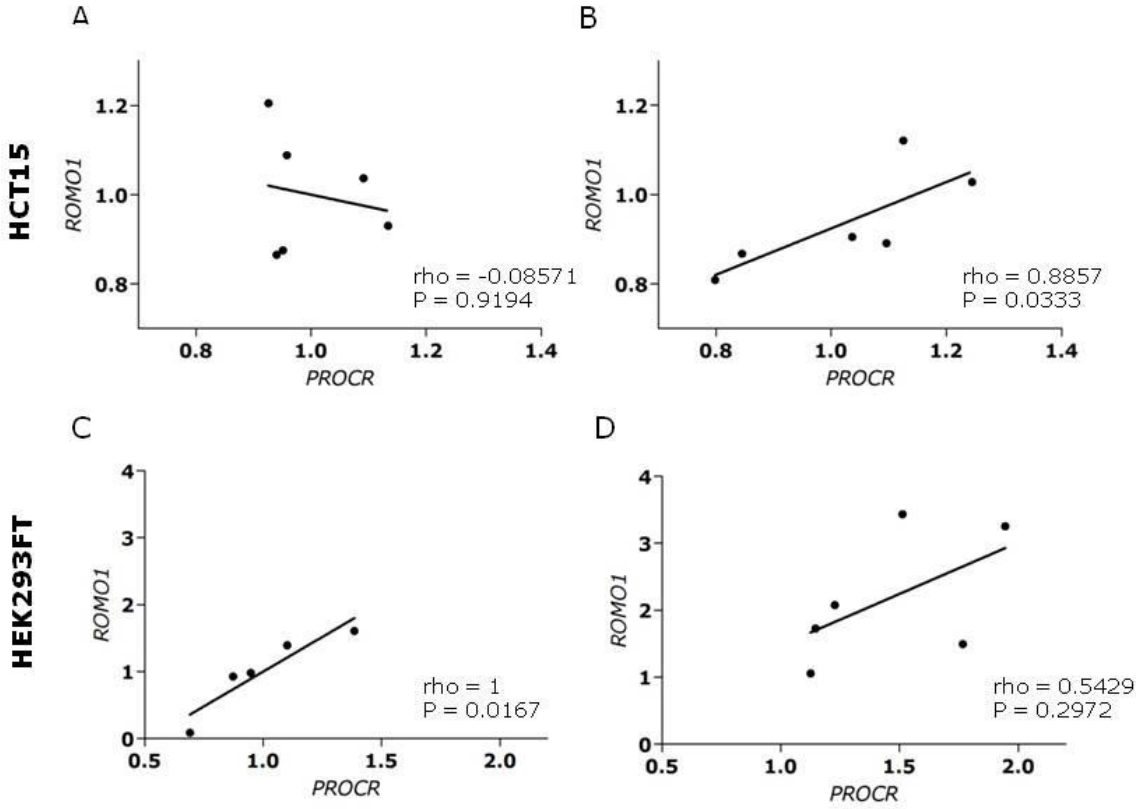


Figure 34. Correlation between *ROMO1* and *PROCR* expression in HCT15 A) WT cell clones and B) mixed population; and in HEK293FT C) WT cell clones and D) edited cell clones (n = 6 in all conditions). Spearman's rho value and p value (two-tailed) are shown.

2.4. Genotyping of cell lines

Three SNPs representing the haplotypes of disease-associated SNPs that were located between the two TADs from the *PROCR-ROMO1* region, were genotyped in HCT15 and HEK293FT cell lines. We wanted to ascertain whether there was a correlation between SNP genotypes and the changes in co-expression observed in the edited cells. HCT15 cells were homozygous for the alternative

allele of both rs6060369 and rs224371 and heterozygous for rs2104417. HEK293FT cells were homozygous for the alternative allele in rs6060369 and heterozygous in the remainig two SNPs (Table 16).

Table 16. Genotypes of HCT15 and HEK293FT cell lines.

	HCT15	**HEK293FT**
rs6060369	GG	AG
rs224371	GG	AG
rs2104417	AG	AG

3. Acute changes in methylation patterns in CD

DNA methylation is an important epigenetic mechanism involved in gene regulation. Aberrant DNA methylation is a feature of a number of human complex diseases, and it could be involved in the response to different stimuli in CD.

3.1. Bisulfite conversion and methylation-specific NGS

Biopsy portions taken from 7 CD patients on GFD as well as from 8 non-CD controls were cultured with (challenged) and without (unchallenged) 250 μg/ml gliadin and genomic DNA was extracted and bisulfite converted. The efficiency of the bisulfite treatment was assessed by PCR amplification using *GAPDH* primers for bisulfite-modified and non-modified DNA (Figure 35).

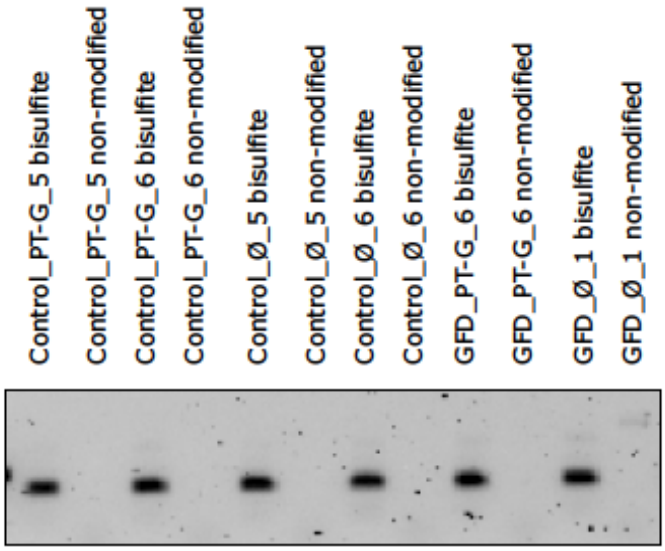


Figure 35. Example of the efficiency of bisulfite conversion. The absence of PCR amplification using non-modified primers indicates efficient conversion. GFD = gluten free treated CD patients; Control = non-celiac controls; PT-G = challenged biopsies; Ø = unchallenged biopsies.

Six genomic regions (mapping to *DFNA5*, *HDAC4*, *HLA-B*, *SLC46A1*, *TRIM15* and *TAP1*) were amplified by methylation-specific PCR in the bisulfite treated DNA. Overall, the six amplicons covered 2,023 bp and included 125 CpG sites. PCR products were checked by gel electrophoresis and band intensities were calculated using Image Lab v5.2.1 and pooled at equal concentrations for each sample.

Amplicon pools were sequenced using pair-ended reads in an Illumina MiSeq machine. In total 23,431,176 reads were obtained and 71.56% of them were properly aligned to a converted reference sequence (Table 17) (Appendix 9).

Table 17. Reads per region and their alignment to a reference sequence.

				Reads per region					
	Total reads	**Mapped %**	**Properly paired %**	***DFNA5***	***HDAC4***	***HLA-B***	***SLC46A1***	***TAP1***	***TRIM15***
Total	23431176			30870	5236231	1322009	155580	2138500	7622379
Average per sample (+/- SD)	781039.2 (+/- 104855.52)	72.12% (+/- 4.67)	71.56% (+/- 4.71)	1029 (+/- 354.89)	174541.03 (+/- 78410.41)	44066967 (+/-26401.09)	5186 (+/- 2931.23)	71283.33 (+/- 30295.26)	254079.3 (+/- 74969.78)

The percentage of cytosines in non-CpG positions was analyzed in order to quantify the efficiency bisulfite treatment. All non-CpG cytosines were converted to thymines and read accordingly in more than 98.12% of the reads in all the samples, pointing the high efficiency of bisulfite conversion (Table 18).

Table 18. Percentage of Cs in non-CpG positions per sample and region. In italics, C proportions in non-CpG positions > %5 (value > 0.05). In those cases, reads were discarded and that particular region was not analyzed in the sample. GFD = gluten free treated CD patients; Control = non-celiac controls; PT-G = challenged biopsies; Ø = unchallenged biopsies.

	C proportion in non-CpGs					
Sample	***DFNA5***	***HDAC4***	***HLA-B***	***SLC46A1***	***TAP1***	***TRIM15***
Control_Ø_1	0.0091	0.0069	0.0115	0.011	0.0031	0.0035
Control_Ø _2	0.0073	*0.1334*	0.0259	0.0093	0.004	0.0046
Control_Ø _3	0.0069	0.0081	0.0484	0.0097	0.0118	0.0068
Control_Ø _4	0.0292	0.0047	0.0041	0.0106	0.0026	0.0053
Control_Ø _5	0.0131	0.0078	0.0071	0.0114	0.0069	0.008
Control_Ø _6	0.0159	*0.1942*	0.0105	0.0092	0.008	0.0062
Control_Ø _7	*0.0741*	0.0078	0.0326	0.0135	0.0068	0.0093
Control_Ø _8	0.0091	0.01	0.0201	0.0098	0.0072	0.0038
Control_PT-G_1	0.0081	0.006	0.0068	0.0115	0.005	0.0053
Control_PT-G_2	*0.1264*	0.005	0.0168	0.0121	0.0062	0.0057
Control_PT-G_3	0.0164	0.0058	0.0117	0.0166	0.0046	0.0041
Control_PT-G_4	0.0358	0.01	0.0103	0.0308	0.0112	0.0181
Control_PT-G_5	0.005	0.0128	0.0227	0.0095	0.0089	0.0086
Control_PT-G_6	0.0077	0.008	0.0069	0.0134	0.0055	0.0062
Control_PT-G_7	0.0426	0.008	0.0186	0.02	0.0066	0.0104
Control_PT-G_8	0.0034	0.0019	0.0049	0.0048	0.0022	0.002
GFD_Ø _1	0.0283	0.0064	0.0142	0.0115	0.0062	0.0062
GFD_Ø _2	*0.1086*	0.0107	0.0228	0.0162	0.0058	0.0056
GFD_Ø _3	0.0108	*0.107*	0.0052	0.0073	0.0442	0.0018
GFD_Ø _4	*0.0756*	0.0101	0.0087	0.0149	0.0071	0.0065
GFD_Ø _5	0.0167	0.0094	0.0029	0.0102	0.0065	0.0054
GFD_Ø _6	0.009	0.0142	0.0162	0.0139	0.0069	0.0054
GFD_Ø _7	0.0087	*0.1653*	0.0083	*0.1422*	0.0042	0.0017
GFD_PT-G_1	*0.1368*	*0.1155*	0.0249	0.009	0.0351	0.0037
GFD_PT-G_2	0.0073	0.0496	0.0221	0.0113	0.0252	0.0037
GFD_PT-G_3	0.0054	0.0066	0.0317	0.014	*0.0511*	0.0022
GFD_PT-G_4	0.0236	0.0067	0.0175	0.0101	0.0095	0.0062
GFD_PT-G_5	0.0162	*0.0671*	0.0129	0.0308	0.0059	0.0051
GFD_PT-G_6	0.0116	0.0036	0.0232	0.0115	0.0035	0.0054
GFD_PT-G_7	0.0077	0.0238	0.0025	0.0086	0.0094	0.0062
Average (+/- SD)	**0.0142 (+/-0.0102)**	**0.0102 (+/- 0.0094)**	**0.0157 (+/- 0.0104)**	**0.0128 (+/- 0.0058)**	**0.0093 (+/- 0.0095)**	**0.0058 (+/-0.0031)**

3.2. Methylation alterations in CD

We observed significant hypomethylation in *HDAC4*, upon gliadin challenge in control biopsies (Figure 36) but not in samples from GFD patients (Figure 37). However, this hypomethylation was present in unchallenged biopsies from CD patients on GFD, when comnpared to unchallenged controls (Figure 38), pointing out that the *HDAC4* region is hypomethylated in GFD patients regardless of acute gluten exposure. Additionaly, these findings suggest that the non-celiac intestine is responsive to gliadin.

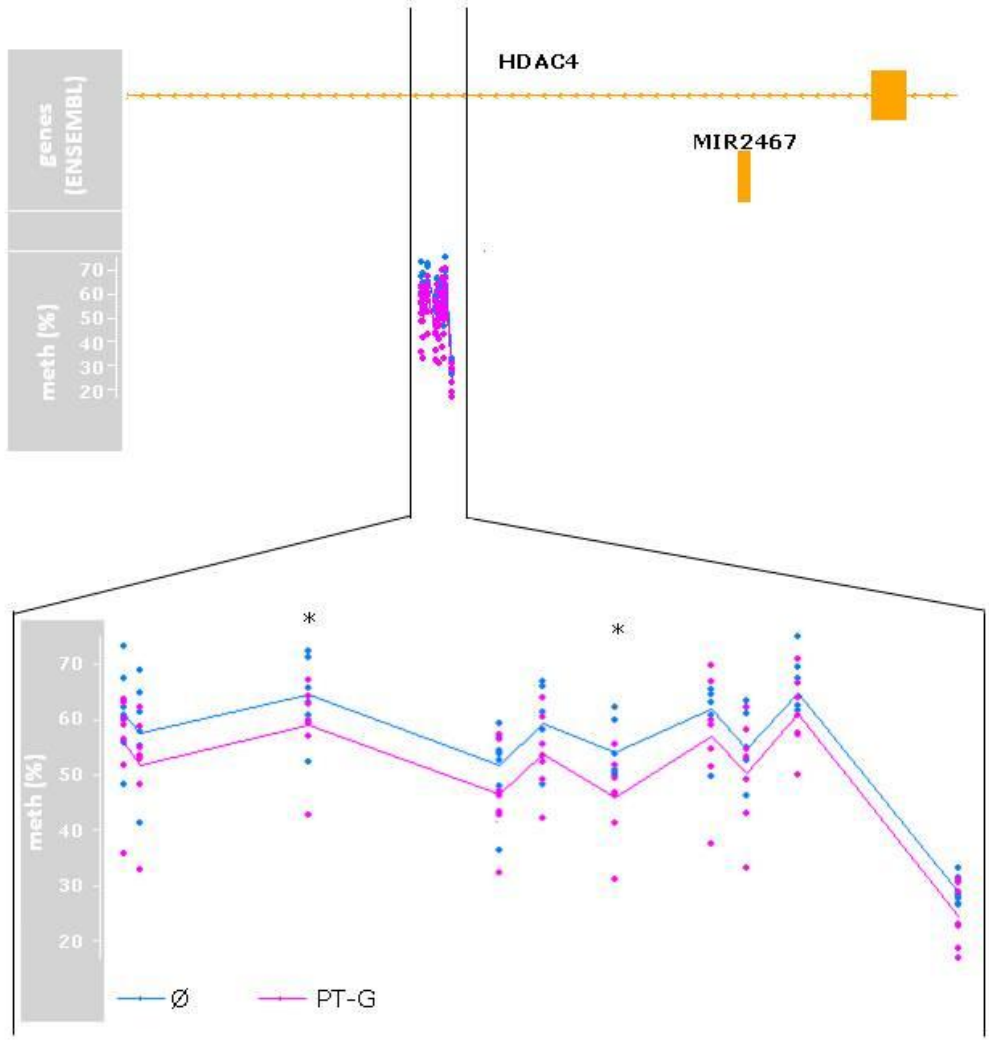


Figure 36. Methylation profiles of the *HDAC4* region in gliadin-challenged and unchallenged biopsies from non-celiac controls. Asterisks represent significant differentially methylated positions (DMPs) (*P < 0.05).

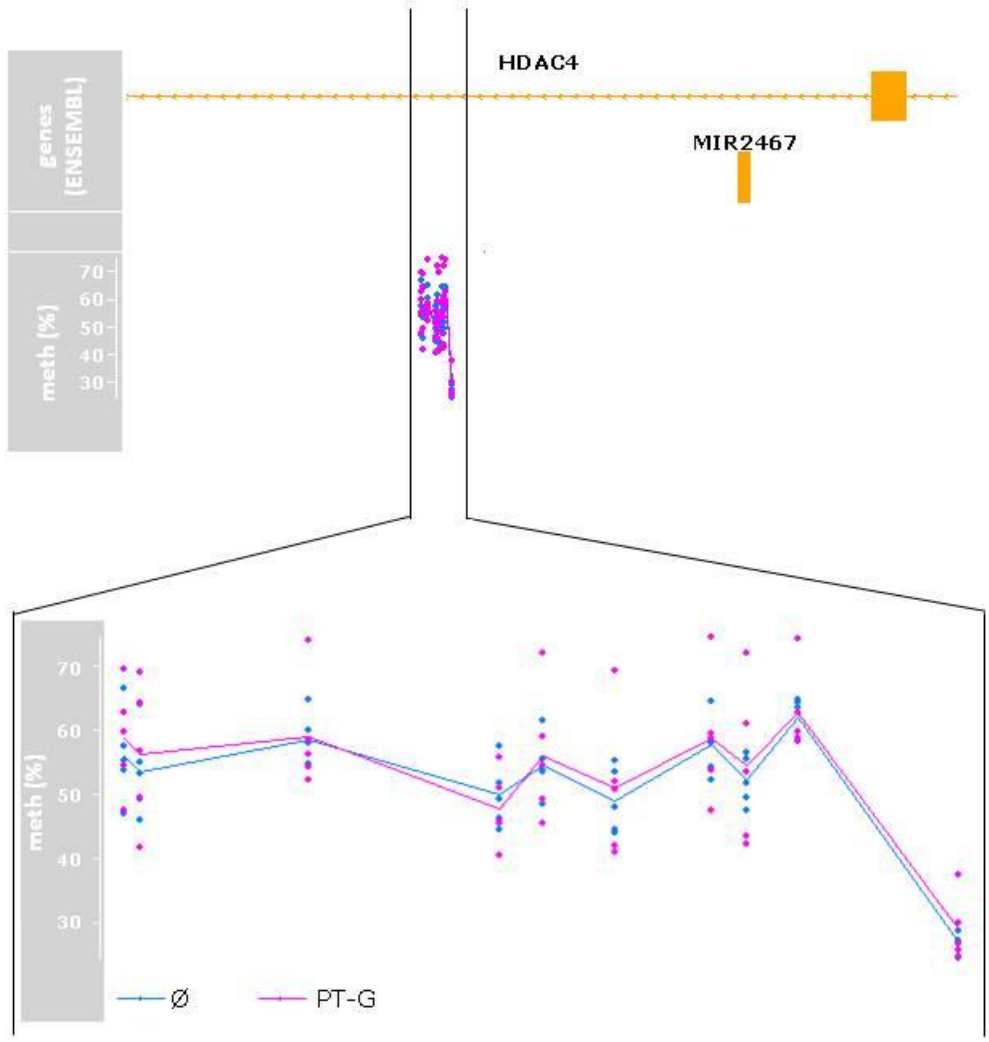


Figure 37. Methylation profiles of the *HDAC4* region in gliadin-challenged and unchallenged biopsies from CD patients on GFD.

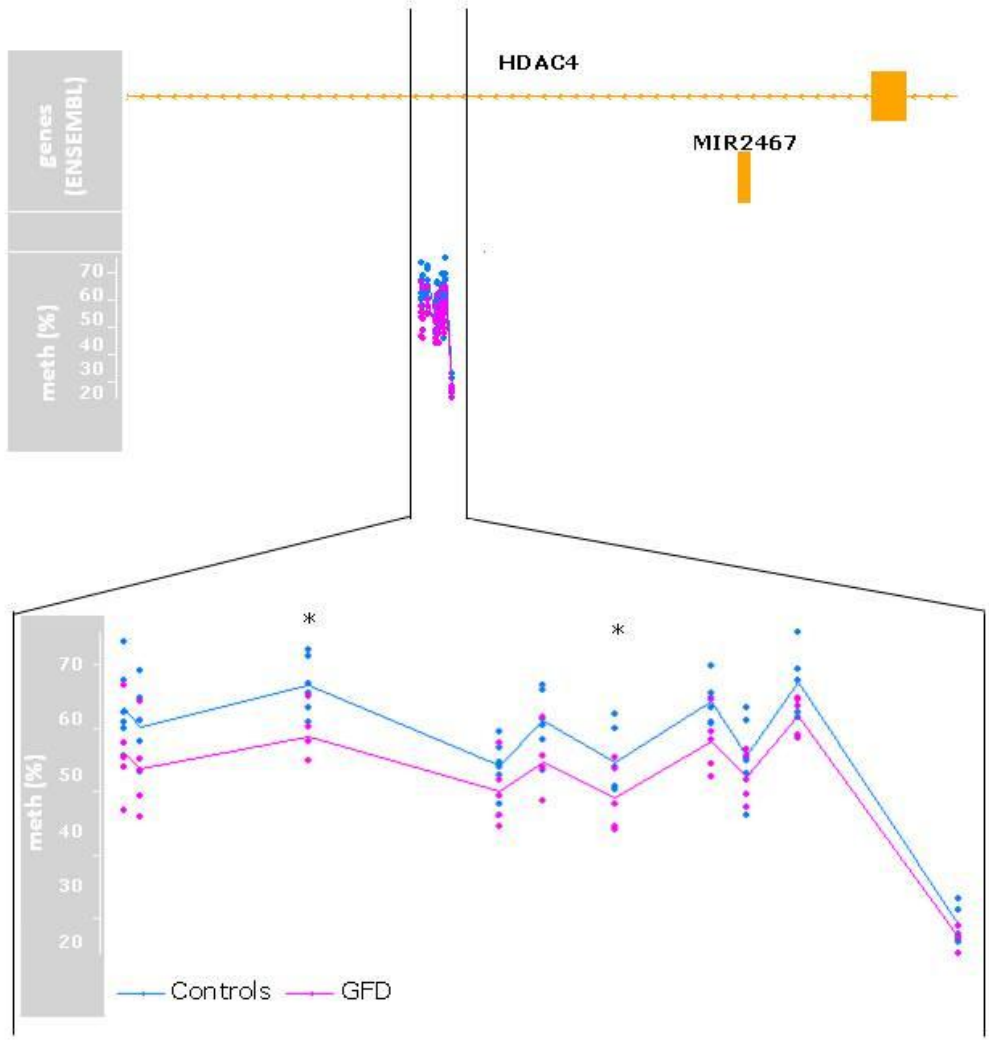


Figure 38. Methylation profiles of the *HDAC4* region comparing biopsies from non-celiac controls and CD patients on GFD. Asterisks represent significant DMPs (*P < 0.05).

On the other hand, results from the present study were used to validate a previous work carried out by our group (Scientific Reports, under review), where methylation signatures of CD were studied in the epithelial and immune cell populations from the intestine. In that work, a hypomethylated differentially methylated region (DMR) was observed in *TAP1* in both the epithelial and the immune cells from active CD patients compared to controls. Another DMR was found in the epithelial fraction in *HLA-B*. In this thesis, 6 CpG sites per *locus* were studied in an independent cohort of 7 GFD CD patients and 8 controls, and the results were replicated (Figures 39 and 40).

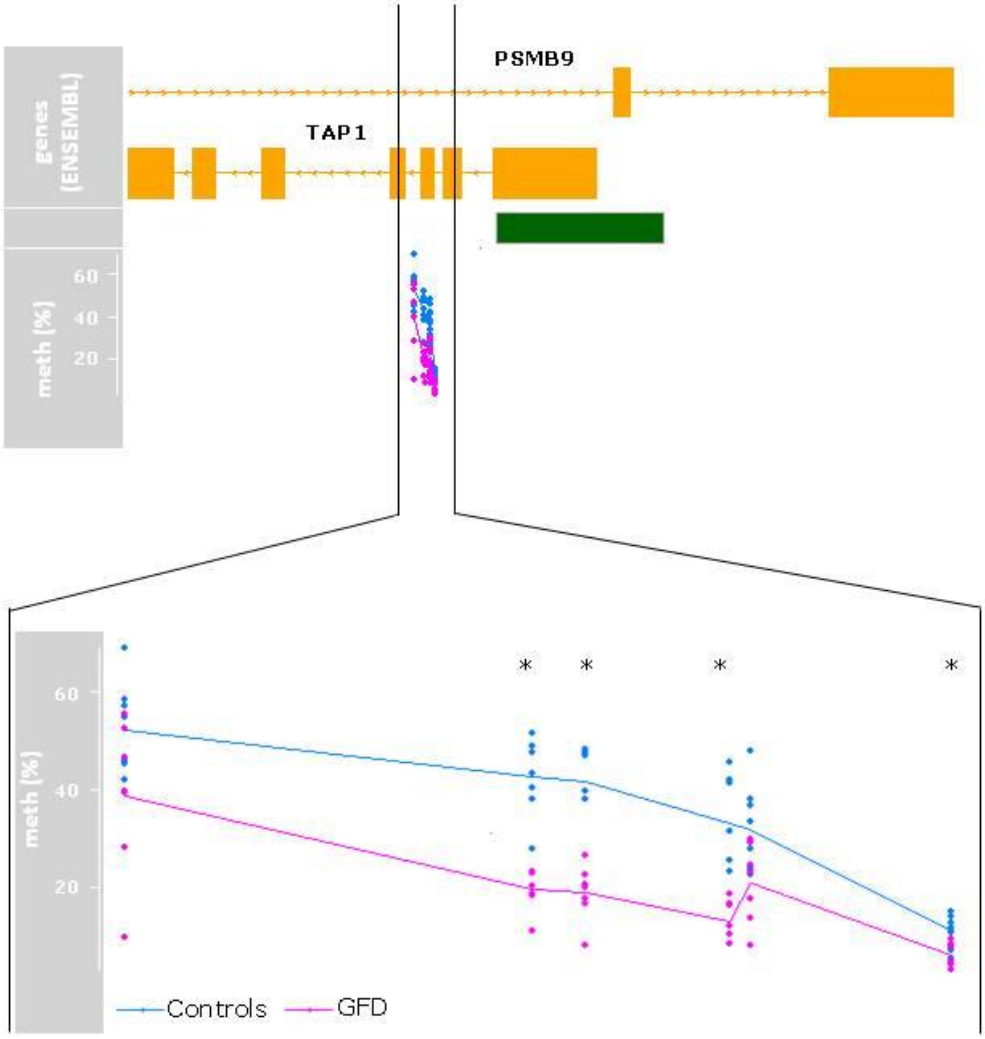


Figure 39. Methylation profiles of the *TAP1* region comparing unchallenged biopsies from CD patients on GFD and from non-celiac controls. Asterisks represent significant DMPs (*P < 0.05).

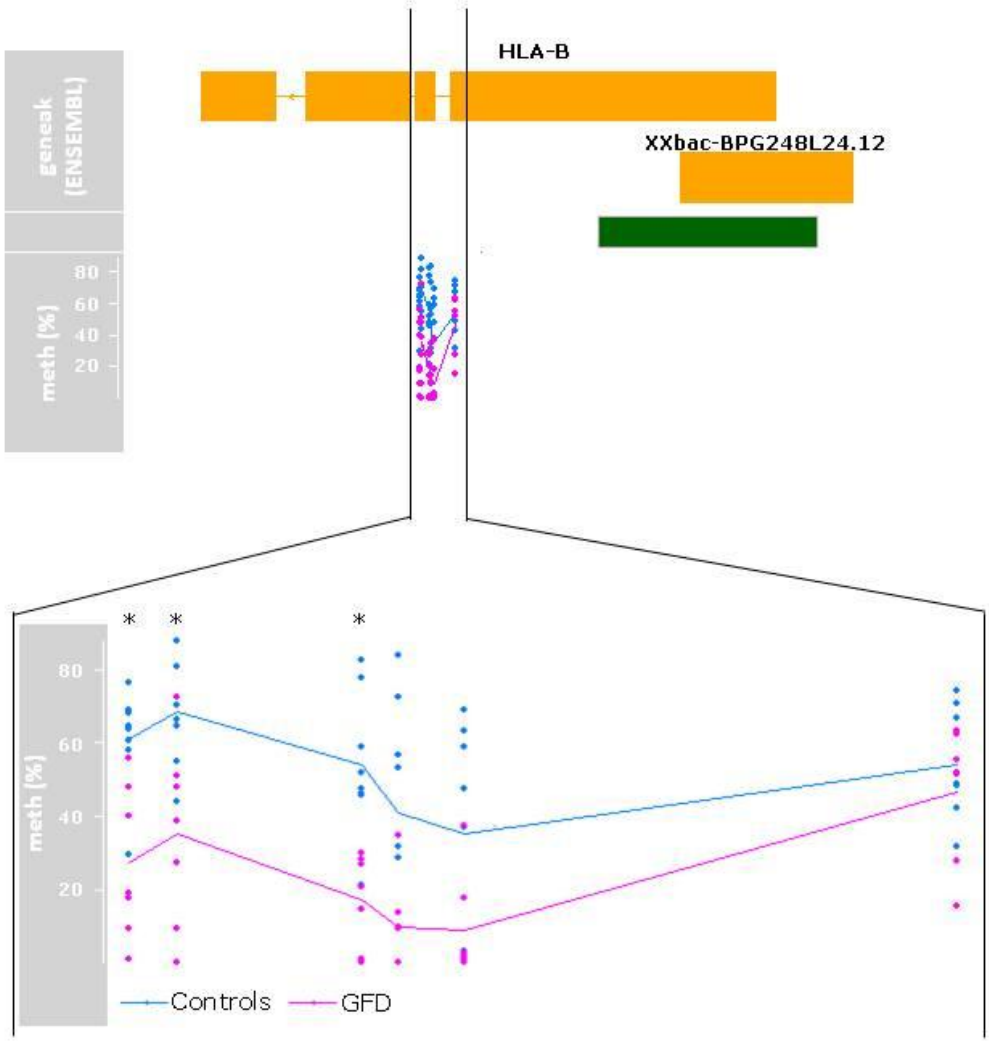


Figure 40. Methylation profiles of the *HLA-B* region comparing unchallenged biopsies from CD patients on GFD and from non-celiac controls. Asterisks represent significant DMPs (*P < 0.05).

Discussion

Celiac disease is a complex autoimmune disease that develops in genetically predisposed individuals upon exposure to dietary gluten. HLA-DQ2 and HLA-DQ8 haplotypes account for around 40% of the genetic contribution to CD and are present in almost all celiac patients. The majority of efforts aimed at identifying the genetic predisposition to the disease have relied on SNP association studies, and the contribution of common genetic variation identified is able to explain around 50% of the heritability (Trynka et al., 2011). However, other layers of genomic information that are independent from DNA sequence variation could also contribute to the pathogenesis of CD but have been left unscrutinized.

In this aspect, epigenetic mechanisms could have a key role in the disease, since they regulate gene expression and are sensitive to external stimuli, bridging the gap between environmental and genetic factors (Gupta and Hawkins, 2015).

In this thesis, publicly available data from different omic layers have been used to identify genetic and epigenetic gene regulation mechanisms that could be involved in CD pathogenesis. For example, previous expression microarray experiments from our group, results from the Immunochip SNP association project, published information on TADs and recently generated methylation and expression data have been collected and reanalyzed, allowing us to identify novel regulatory elements and genomic regions involved in the development of CD.

Whole genome co-expression in CD

Transcriptomic approaches other than the classical differential expression analyses (such as co-expression studies) are useful and could help to identify genes that are simultaneously active, often participating in the same biological processes (de la Fuente, 2010). Co-transcribed genes identified through co-expression analyses might share common regulatory elements, which in turn

could be responsible for the changes in expression and co-expression that occur upon interaction with the environment, and be relevant to the disease.

Several attempts to conduct comprehensive and comparative analyses of gene co-expression at the network level have been performed. One of the first studies to be published on this topic aimed to reveal similarities and differences between Hepatitis B Virus- and Hepatitis C Virus-derived hepatocellular carcinoma, focusing on the inflammatory processes driven by the viral infections (He et al., 2012). That work demonstrated the power of a network-based Systems Biology approach in identifying different oncogenic and dysfunctional modules compared to previous differential expression analyses and viral protein target-based methods.

To take advantage of the strength of co-expression-based approaches, our group has investigated co-expression in several pathways in CD, and has elucidated that gliadin is able to disrupt co-expression in certain groups of genes and provoke a coordinated response in others (Fernandez-Jimenez et al., 2014; Plaza-Izurieta et al., 2015). However, co-expression analyses performed in CD so far have been limited to small numbers of genes or pathways, apart from a single work that studied genome-wide co-expression in peripheral blood (Quinn et al., 2015). This is the first whole genome co-expression analysis that tests the effect of gliadin in duodenum biopsies of CD patients, and is able to identify regulatory elements that could play a role in the development of the disease.

In this thesis, co-expression was analyzed in data from expression microarray experiments resembling the acute and chronic effects of gliadin in the celiac gut (Castellanos-Rubio et al., 2008, 2010). A complete reorganization of co-expressed gene groups was observed in both experiments, but the acute insult provoked more dramatic co-expression changes. Noteworthy, gliadin-free conditions, mimicking health (or treatment), showed more unstable co-expression modules than those related to the disease stage (with gliadin),

suggesting that the immunogenic insult dysregulates health-related coordination, whereas the co-expression modules formed in disease seem to be more stable across stages. As gene regulation plays a key role in complex disease pathogenesis (Cookson et al., 2009), identification of the potential regulatory elements involved in this alteration of coordination could be a good strategy to better understand CD pathogenesis and develop medical applications in the future.

Therefore, based on the differentially co-expressed genes identified, enrichment analyses for TFBSs and miRNAs were performed in order to find potential co-expression modifiers. TFs control gene expression, and approximately 10% are implicated in human diseases (Bouhlel et al., 2015). TFBSs enrichment analyses revealed that several TFs could be involved in the observed changes in co-expression. However, the analysis of the DCG sets did not produce significant results in the majority of cases, suggesting that there must be additional mechanisms other than TFs that control co-expression in CD.

As expected, among the biological functions related to DCGs we found annotations related to the immune response and apoptosis, both key elements in CD pathogenesis, since it is an autoimmune disease in which apoptotic activity is increased in the intestinal mucosa (Green and Cellier, 2007; Shalimar et al., 2013). In addition, other CD-related GO terms such as vasodilatation, lower in celiac patients compared to controls (Sari et al., 2012), or lysosomal transport, which is altered in CD (Lebreton et al., 2012) were also found. This could suggest that gliadin induced co-expression changes could affect the proper functioning of those biological processes.

When the original gene modules were considered as a background, many GO annotations were identified. Interestingly, some of them have been previously related to CD, namely cell-cell adhesion involved in intestinal permeability (Jauregi-Miguel et al., 2014), inflammation mediators (protein kinase cascade)

(Yao et al., 2018), or apoptosis (Shalimar et al., 2013). Surprisingly, when the DCGs were removed from their original modules, no GO annotations remained significant. These results confirm a pivotal role of the DCGs in the functions defined for the modules where they were identified and support the biological relevance of the outcome of our enrichment analyses.

Four TFs showed expression changes in CD patients. IRF1 and CREB1, both of them related to the innate and adaptive immune responses (Guo et al., 2010; Wen et al., 2010) showed increased expression in active patients compared to the control group. This is the first work to show overexpression of *CREB1* in CD, while the increase of *IRF1* confirms previous reports (Lahdenperä et al., 2011). On the other hand, overexpression of the *NFKB1* subunit in disease is in accordance with the involvement of the NFκB pathway in CD (Fernandez-Jimenez et al., 2014). *ELK1* a TF that is known to be involved in intestinal permeability (Al-Sadi et al., 2013), was overexpressed in active CD. It is known that intestinal permeability is an important feature of CD (Jauregi-Miguel et al., 2014). More interestingly, most of the target genes of the TFs analyzed followed the same expression trends as the TFs themselves.

As DCGs were identified through gliadin-induced co-expression changes, we wanted to know whether it could alter the TFs at the protein level. Upon gliadin stimulation, increased nuclear translocation of CREB1 and IRF1 was observed in the C2BBe1 cell line. In a recent work, IRF1 was shown to be upregulated by both interferon-γ and CD-associated bacteria challenges in intestinal epithelial cells of active CD (Pietz et al., 2017). Taken together, these results confirm an increase of IRF1 activity in situations that promote the disease. On the other hand, there are no previous data on CREB1 in CD, but it has been related to other diseases like human colorectal cancer, where CREB1 acts as an activating TF for tumor driver *RRM2,* or glioblastoma, where CREB1 acts as a mediator of the induction of TGFβ2 (Fang et al., 2016; Rodón et al., 2014). The increase in

nuclear translocation suggests a role for gliadin in the upregulation of target genes through the activation of TFs, also in non-celiac individuals.

Encouraged by those results, we investigated the interaction of the candidate TFs and their target genes in the context of gliadin stimulation. With that purpose, ChIP-PCR was performed in C2BBe1 cells. Although we had observed altered expression patterns of several of the TF targets in different stages of the disease, we could not find evidences of changes *in vitro*. However, we confirmed the binding to the putative target genes selected from the DCG sets, at least in the case of IRF1. We cannot rule out that our cell model might be limited because the cell lines used are derived from a non-celiac, cancerous tissue. In this context, this study also stresses the limitations of the *in vitro* models available for the investigation of complex diseases and in particular, of CD. However, the C2BBe1 cell line represents the best available *in vitro* model of absorptive enterocytes and it is widely used as an *in vitro* model in the disease of our interest (Rauhavirta et al., 2011).

On the other hand, several miRNAs were also identified after enrichment analysis. miRNAs are able to regulate gene expression at post-transcriptional level and are predicted to regulate the translation of up to 60% of protein-coding genes (Esteller, 2011). Non-coding RNAs, including miRNAs, modulate gene expression and have previously been related to CD (Castellanos-Rubio et al., 2016; Felli et al., 2017). Different reports showing miRNA alterations in CD (Buoli Comani et al., 2015; Capuano et al., 2011; Magni et al., 2014; Vaira et al., 2014) suggest a role in disease pathogenesis and a potential use for the diagnosis of CD.

Nine miRNAs were prioritized for biological validation, and their expression in duodenal biopsies of active and treated CD patients, as well as in non-celiac controls was analyzed. We analyzed the primary form of nine miRNAs due to sample constraints, since more RNA is needed for mature miRNA analyses, and

RNA of biopsies was a limiting factor. Nevertheless, we studied the expression of two mature miRNAs, and both mature and pri-miRNA forms showed a concordant pattern as it has been observed in other cases (Powrózek et al., 2018). Additionally, it has been shown that pri-miRNAs can also contribute to target repression (Trujillo et al., 2010; Yue et al., 2011), supporting the adequacy of our approach. In the end, six previously unidentified pri-miRNAs were shown to be downregulated in CD when compared to non-celiac controls.

Since the dysregulation in the expression of each miRNA is believed to affect many mRNAs (Chen et al., 2016; Erson and Petty, 2008), the expression of the target genes of the downregulated pri-miRNAs was queried in our RNA-seq dataset of intestinal cell fractions. Twenty-one target genes showed significantly different mRNA levels in active disease. In general, overexpression was observed in the celiac epithelial fraction, coherent with the downregulation of the pri-miRNAs. In contrast, both overexpressed and downregulated genes were found in the celiac immune fraction. Interestingly, the epithelial fraction is somehow more susceptible to changes in miRNA target expression.

Only four arget genes were differentially expressed in both duodenal cell fractions; *JAG1*, *SP1* and *CYP2J2* showed the same trend in both fractions, while *ARL4C* had different patterns in each, indicating that cell specific regulation is an important aspect to take into account. However, although several miRNA targets showed alterations in their mRNA levels, we cannot rule out the possibility that the changes observed are derived from other factors independent from miRNA regulation.

Additional mechanistic studies will be necessary to completely understand the molecular and cellular mechanisms underlying our observations, but our work shows that co-expression studies can complement classical differential expression analyses, and are particularly useful for the identification of regulatory elements that could be relevant to human diseases. Moreover, these

approaches can take advantage of publicly available genome-wide datasets and reuse raw experimental results. Nevertheless, even though this approach was useful to identify DCGs, TFs and miRNAs, the results obtained through *in vitro* experiments were not able to explain the totality of the observed co-expression alterations.

Topologically associating domains in CD

Gene regulation is strongly influenced by chromosome organization (Lupiáñez et al., 2016; Sexton et al., 2007), and TADs are key elements in the conformation of three-dimensional chromatin. TADs are functional units of chromatin that are enriched for chromatin-chromatin interactions, and appear to be stable across cell types and conserved across species in mammals (Dixon et al., 2012). CTCF sites are found flanking such domains, and define TAD boundaries and contribute to chromatin looping (Ong and Corces, 2014).

Aberrations in TAD boundaries have been described in several pathogenic events. Human limb malformations have been associated with TAD boundary disruptions, including deletions, duplications and inversions altering the structure of the TAD spanning the WNT6/IHH/EPHA4/PAX3 *locus* (Lupiáñez et al., 2015). TAD disruption is often also found in cancer cells. For example, hypermethylation at cohesin and CTCF binding sites lead to reduced CTCF binding, and inactivation of TAD boundaries in gliomas. This switches on cancer drivers such as *PDGFRA* by enhancers located outside their TADs (Flavahan et al., 2016).

In addition, TADs are crucial chromosome structural units of long-range regulation (Dekker and Heard, 2015). In one of the first studies relating TADs with gene expression, it was found that TADs also align with coordinately regulated gene clusters, showing that the expression levels of genes located in the

same TADs are more correlated among them than the levels of those located in different ones (Nora et al., 2012).

As previously mentioned, co-expressed genes might share regulatory elements responsible for the expression and co-expression changes observed in different stages of the disease. In this sense, TADs could pay an important role in CD, since they are important players in gene regulation. In this thesis, a new approach to identify TAD disruption and merge events underlying changes at a co-expression level is proposed. In summary, candidate regulatory regions that could be involved in the coordination of gene expression have been selected using co-expression data obtained in this study, together with published data of TADs and CD-associated SNPs.

We hypothesized that alterations in TAD organization could explain, at least in part, the dysregulation of co-expression in CD. Therefore, co-expression patterns were defined from RNA-seq data and co-expressed genes that overlapped with conserved TADs (Dixon et al., 2012) were identified. In healthy individuals, 739 co-expressed genes were located in 486 TADs, and in the case of CD, 613 genes were located in 430 TADs. To further prioritize TADs potentially involved in CD, we selected those cases in which SNPs associated with the disease were present (Garcia-Etxebarria et al., 2016). In particular, among all the regions identified, we selected those where the CD-associated SNPs fell between adjacent TADs or within TADs. We postulated that risk alleles could particpate in domain disruption through mutations in TAD boundaries leading to either a merge of two adjacent TADs, or to a break-up of a TAD into two new domains.

In general, insertions, deletions or inversions are necessary to alter the 3D organization of chromatin. However, single-base changes at specific positions have also been found to alter TAD boundaries and CTCF binding sites. For example, frequent mutations have been reported at CTCF/cohesin-binding sites, identifying them as major mutational hotspots in the noncoding cancer genome

(Katainen et al., 2015). Also, mutations in the CTCF motif at a boundary of a TAD containing the *NOTCH1* gene have been discovered in ovarian cancer. These nucleotide changes lead to *NOTCH1* dysregulation, probably through altered enhancer action following TAD disruption (Ji et al., 2016). In a more recent work, a SNP-mediated disruption of a CTCF binding site has been observed to be associated with severe influenza. In summary, the authors found that the genotype of rs34481144 influences CTCF binding, and this was correlated with expression of genes related to the response viral infection at the *locus* surrounding *IFITM3* (Allen et al., 2017).

In this thesis, we selected two regions for further characterization; *HSCB-XBP1* region (a potential disruption of a TAD) and *PROCR-ROMO1* region (a potential merge of two TADs). The two selected regions contained a group of SNPs associated to CD that could be implicated in the changes in co-expression between adjacent genes, namely *PROCR* and *ROMO1*, and *HSCB* and *XBP1*, respectively.

Three of these genes have been implicated in pathways associated to CD: *PROCR* acts as a negative regulator of the Th17 response and is specifically expressed on the surface of Th17 cells. The reduction in the expression of *PROCR* leads to higher Th17 pathogenicity and enhances experimental autoimmune encephalomyelitis (EAE) *in vivo* (Kishi et al., 2016). This is concordant with microarray data published by our group, where the expression of *PROCR* is reduced when active CD patients are compared to GFD CD patients (Castellanos-Rubio et al., 2008). Moreover, the Th17 response has previously been described to be upregulated in CD (Castellanos-Rubio et al., 2009; Cicerone et al., 2015).

In turn, *ROMO1* is the key regulator of the release of mitochondrial reactive oxygen species (ROS) (Bae et al., 2011). It has been described to be crucial for cancer cell proliferation and invasion, and has been associated with unfavorable

prognosis in colorectal cancer (Kim et al., 2017). *ROMO1* has been reported to activate NFκB, a pathway that is overexpressed in CD (Fernandez-Jimenez et al., 2014). Overexpression of *ROMO1* promotes nuclear translocation of p65, and it could be an essential regulatory factor in the maintenance of constitutive NFκB activation in tumor cells (Chung et al., 2014). Therefore, *ROMO1* could be a promising therapeutic target for diseases characterized by NFκB dysregulation (Lee et al., 2015).

Finally, *XBP1* is a TF that is involved in the innate immune response (Jheng et al., 2012). Moreover, it was found that XBP1 is required for production of some inflammatory factors such as IL-6 after activation by toll-like receptors (TLRs) (Martinon et al., 2010; Savic et al., 2014) and endoplasmic reticulum stress (Gargalovic et al., 2006; Toosi et al., 2012). Furthermore, it is involved in some diseases including type 2 diabetes (Özcan et al., 2006), cancer (Jin et al., 2016), and inflammatory bowel disease (IBD) (Kaser et al., 2008).

The two candidate regions were studied in cell lines in order to determine whether they are accessible to DNase and therefore have an open chromatin conformation. Furthermore, we also sought to create a cellular model with stable modifications for additional studies. As previously mentioned, the C2BBe1 cell line is widely used as an *in vitro* model in CD; however, previous attempts performed in our laboratory showed the difficulties to edit this cell line. Thus, HCT116, HCT15 and HEK293FT human epithelial cell lines were used for the following experiments. HCT116 and HCT15 were chosen for being colon-derived cell lines that have been used to study colorectal cancer (Bessa et al., 2018; Gong et al., 2018). Moreover, HCT116 cell line has also been used to study tight junction regulation and intestinal permeability (Kolodziej et al., 2011), features known to be altered in CD (Jauregi-Miguel et al., 2014). On the other hand, the HEK293FT cell line is originally derived from human embryonic kidney cells, and has been extensively used for genome editing using CRISPR-Cas9 (He et al., 2016).

We next wanted to identify the CTCF binding sites that could act as TAD boundaries (Ghirlando and Felsenfeld, 2016). We chose CTCF binding sites that were located inside DNase I hypersensitive regions, since those sites reveal chromatin stretches that are accessible to protein binding (Gross and Garrard, 1988). DNase digestion followed by quantitative PCR revealed high DNase accessibility in both regions in cultured cell lines. In general, CTCF binding has been related to transcriptionally active *loci* (Batlle-López et al., 2015), although DNase I sensitivity signals in CTCF binding sites at TAD boundaries have been reported to be weaker than those located inside TADs (Hong and Kim, 2017).

We used CRISPR-Cas9 gene editing technology in order to obtain precise and permanent alterations in CTCF binding sites in epithelial cell lines. This technique enables long-term studies in the edited cell lines, which allows studying phenotypes in deep. Some studies have already deleted CTCF binding sites using CRISPR-Cas9 editing. In some studies the deletion of a single CTCF site was enough to perturb a TAD boundary (Lupiáñez et al., 2015), while in others it was not, suggesting that apart from CTCF binding, additional mechanisms may play a role in establishing TAD boundary formation (Barutcu et al., 2018).

In this work we have generated clonal mutant HEK293FT cell lines and a mixed population in the HCT15 cell line for one of the regions. In particular, we deleted a 377-bp region located between two adjacent TADs containing *PROCR* and *ROMO1* genes, respectively, that harbors a DNase I hypersensitive site and a CTCF binding site. The deletion produced expression changes in both *PROCR* and *ROMO1* genes in the mutant HEK293FT cell line, and changes in the co-expression between the two genes in both the clonal mutant HEK293FT cell lines and in the HCT15 mixed population. However, the co-expression changes that occur in each of the two cell lines upon deletion of the CTCF binding site seem to be contradictory; co-expression is disrupted in HEK293FT cells while a new co-expression occurs between the genes in HCT15 cells.

Considering those results, we wanted to know whether different genotypes could contribute to differences in CTCF binding and TAD organization, and consequently explain the different co-expression relationships observed in the two cell lines. With that purpose, rs6060369, rs224371 and rs2104417 SNPs were genotyped in the HCT15 and HEK293FT cell lines. The three SNPs represented the haplotypes of the SNPs associated with CD that were located between the two TADs in the *PROCR-ROMO1* region, outside the deleted CTCF binding region. Different genotypes were found in the two cell lines, but we are not able to conclude whether this variation is enough to explain the differences in co-expression observed upon genomic edition.

However, it has been shown that the genotype of one of the SNPs (rs6060369) is associated with changes in *PROCR* gene expression (https://www.gtexportal.org/home/), indicating that it is an expression quantitative trait *locus* (eQTL). We have observed that the genotype of the rs6060369 SNP differs between the HCT15 and HEK293FT cell lines; therefore, it could well be that the different expression and co-expression signatures in the cell lines depend on the identified eQTL. Moreover, we must also take into account the fact that HEK293FT cells were homozygous for the deletion while HCT15 were a mixed population, and this could also explain the different co-expression results obtained in HEK293FT and HCT15 cell lines upon genomic edition.

In conclusion, the selected regions could be important in CD pathogenesis since mutations in the *PROCR-ROMO1* region, located between the TADs containing *PROCR* and *ROMO1* affect the co-expression of those genes, which in turn are related to pathways altered in the disease. Nevertheless, taking into account the technical limitations, as well as the lack of a perfect *in vitro* model, the development of novel approaches will be necessary to collect more evidence relating gene expression, TADs and GWAs signals in the early future (Pervjakova and Prokopenko, 2017).

Acute changes in methylation patterns in CD

DNA methylation has an important role in many biological processes including development and disease through modulation of gene expression (Wan et al., 2015). Aberrant DNA methylation patterns are frequently observed in disease, especially in cancer (Jones et al., 2007), and it is also known to have a role in immune diseases like IBD or T1D (Agardh et al., 2015; McDermott et al., 2016). Many environmental factors, such as pollution (Madrigano et al., 2011), temperature and humidity (Bind et al., 2014) or smoking (Zeilinger et al., 2013) have been implicated in DNA methylation changes.

Regarding CD, a candidate-gene methylation analysis in duodenal biopsies of patients was able to detect changes in the promoters of several NFκB-related genes (Fernandez-Jimenez et al., 2014). In this thesis we wanted to know whether 4 hours of gliadin exposition could be enough to cause methylation changes in CD. With that objective, six candidate regions selected using previous expression (Castellanos-Rubio et al., 2008, 2010) and methylation data (Scientific Reports, under review) were studied in gliadin-challenged and unchallenged biopsy portions from CD patients on GFD and in gliadin-challenged and unchallenged biopsy portions from non-celiac controls.

To search for the differentially methylated cytosines, bisulfite treated DNA was amplified by methylation-specific PCR, and amplicons of selected regions were sequenced using NGS. Reads were mapped to a reference genome and 71.56% were properly aligned, a much higher proportion than normally achieved, which is usually around 40% (Krueger and Andrews, 2011; Tran et al., 2014). Besides, in this work non-CpG cytosines were detected as thymines in more than 98.12% of reads in all samples, highlighting the high efficiency of bisulfite conversion achieved in this work (Holmes et al., 2014).

Gliadin-challenged and unchallenged biopsy portions from GFD CD patients and non-celiac controls were compared in order to determine the acute effects of gliadin on DNA methylation in candidate regions. Samples from GFD CD patients are very informative since genomic alterations could reflect constitutive characteristics related to their genetic or epigenetic predisposition, that have been either inherited or acquired very early in childhood (before CD onset) or even in prenatal stages. Otherwise, it could be that some of the disease-related changes at the DNA methylation do not revert despite tissue recovery after GFD.

We did not observe significant acute methylation changes except for *HDAC4*, a gene known to be related to inflammatory processes (Yang et al., 2018). Hypomethylation of the *HDAC4* region was present in gliadin-challenged biopsies from controls, but not from GFD CD patients. Methylation differences provoked by acute stimuli are difficult to detect if previous genetic or epigenetic signatures are already present. Indeed, the inability to detect methylation differences between gliadin-challenged and unchallenged GFD CD biopsies could be due to the fact that *HDAC4* was already hypomethylated in samples from GFD CD patients. These results suggest that alterations in *HDAC4* mehtylation could be either constitutive (inherited or acquired very early in life in CD-prone individuals) or be related to the disease and persist even upon GFD-treatment (pointing to a non-reversible epigenetic signature). Very interestingly, *HDAC4*, a target of the candidate TF IRF1 identified in this thesis, is upregulated in active disease. Athough not sinificant, the experssion in GFD patients followed the same trend compared to controls, which seems coherent with the persistant differential methylation around *HDAC4*.

Furthermore, our results show that mehtylation alterations in *HDAC4* do occur upon an acute exposition to gliadin in controls, but do not persist, and point to a reversible epigenetic signature in non-celiac individuals. This is the first work analyzing methylation changes induced by acute gliadin exposure in the non-

celiac intestine, but previous studies have demonstrated gliadin effects in non-celiac individuals at an mRNA level (Fernandez-Jimenez et al., 2014).

In a recent work carried out by our group, methylation signatures were studied in the intestinal epithelial and immune cell populations of active CD patients and non-celiac controls (Scientific Reports, under review). In that work, among other alterations of methylation, *HLA-B* and *TAP1* gene promoters were found to be hypomethylated in active CD. In particular, *TAP1* was hypomethylated in both cell populations, while *HLA-B* was hypomethylated in the epithelial fraction. Moreover, RNA-seq results showed that gene expression was coherent with the cell-type specific methylation differences observed.

In the present thesis, we could confirm the hypomethylation of the *HLA-B* and *TAP1* promoters in unchallenged biopsies from GFD patients, pointing to either a constitutive or a non-reversible alteration in CD that prevails even after more than two years on GFD. Both *HLA-B* and *TAP1* have been previously related to CD (Bratanic et al., 2010; Pietz et al., 2017). Particularly, TAP1 is an important HLA class-I surface peptide transporter, and recent works showed a downregulation of *TAP1* expression in colorectal cancer, which was inversely correlated with methylation at sites in close proximity to the promoter region (Ling et al., 2017). Altered methylation patterns could indicate an anomalous adaptive immune response in CD.

FINAL REMARKS

This thesis points out the importance of taking into account different layers of genomic information when studying complex diseases, stressing the idea that regulatory elements such as TFs, miRNAs, TADs and DNA methylation, among others, affect both the onset and the development of CD, and by extension, of other common disorders.

Until now, the transcriptomic landscape on CD has been analyzed focusing mainly on the differences in gene expression levels (Castellanos-Rubio et al., 2008, 2010). In the present thesis the transcriptome has been analyzed considering changes in co-expression at the genome-wide level. Thanks to this approach, a complete picture of CD-related alterations of co-expression has been depicted in CD.

This changes in co-expression could be explained, on the one hand, by TFs. In this work we have observed that CREB1 and IRF1 are likely to be involved in the regulation of co-expression in CD, since the expression of a significant number of their targets is affected, and translocation to the nucleus of these TFs is enhanced upon gliadin stimulation *in vitro*.

On the other hand, other genomic mechanisms, such as chromatin organization, could also be implicated in the changes in co-expression observed in the disease. In this work we have investigated the overlap between conserved TADs, and CD-related DCGs and SNPs to find TAD remodeling events implicated in CD. This is the case of the *PROCR-ROMO1* region, where a partial deletion of TAD boundaries altered the local co-expression pattern.

DNA methylation is an important epigenetic mechanism involved in gene regulation (Lou et al., 2014). Up to now, only a few works have addressed the effect of methylation in CD (Fernandez-Jimenez et al., 2014; Scientific Reports,

under review). In this thesis, the relevance of methylation is confirmed by a sensitive analysis, incorporating NGS technology to the study. Interestingly, this study shows that gliadin is able to provoke acute changes in DNA methylation of intestinal cells in non-celiac individuals, in regions that are irreversibly modified in patients, even after more than two years on GFD.

It is also worth mentioning that throughout the whole work we have reused existing data from different genomic layers that we have reanalyzed with innovative bioinformatics approaches to make original proposals on the pathogenesis of CD. This type of data-recycling approaches have proven successful and could be further exploited in the context of CD and of other complex traits.

In this aspect, the past decade has witnessed the generation of vast amounts of genomic data. Traditional biochemical methods are time-consuming and inefficient, while omic technologies perform global and high-throughput analyses, producing data at a large-scale. Following GWA studies that revealed disease associated *loci*, other disciplines such as epigenomics (Gupta and Hawkins, 2015), transcriptomics (Mchale et al., 2013), proteomics (Breker and Schuldiner, 2014) or metabolomics (Fearnley and Inouye, 2016) have emerged. In order to draw a comprehensive view of biological processes, experimental data obtained from different layers have to be integrated and analyzed, as has been the case in this thesis. Multi-omics approaches integrate data obtained for different biological layers using high-throughput analytical approaches and bioinformatics in order to understand their relationships and the functioning of larger systems (Kohl et al., 2014).

Hence, the use of different omic layers and their integrative analysis improves the understanding of the pathological processes of the disease, and reveals key players, pathways and mechanisms that can lead to more precise diagnosis and better treatments. In addition, this approach can help to prioritize candidates in

functional analyses and extend our knowledge in complex diseases as CD, and lead to future therapies and drug development for different complex diseases.

In conclusion, these new findings should encourage the scientific community towards a novel point of view, in which complex and integrative regulatory mechanisms affect gene regulation in a multidimensional way, contributing to the molecular explanation of the genetic variation associated to complex diseases.

Conclusions

1. Gliadin provokes a genome-wide remodeling of co-expression, both upon acute and long-term exposures. However, health-related co-expression modules are extremely disrupted compared to disease-related stages, while co-expression modules identified under the gliadin stimulus seem more stable across stages, revealing basic coordination units.

These findings highlight co-expression as another layer of genomic information involved in CD development.

2. Enrichment analyses of the DCGs identified 10 TFs and 46 miRNAs as potential regulators of part of the co-expression changes observed in CD.

a. Out of the 5 TFs selected for biological validation, IRF1, CREB1, NFKB1 and ELK1 were upregulated in biopsies from active CD patients, and the majority of their target genes were also differentially expressed. On the other hand, 6 pri-miRNAs (hsa-mir-33a, hsa-mir-520b, hsa-mir-520e, hsa-let-7b, hsa-mir-655, hsa-mir-26b) were downregulated in CD, and their targets more often showed upregulation in the epithelial fraction of biopsies from active CD patients than in the immune compartment, suggesting cell-type specificity of the mechanism.

Altogether, these results support the effectiveness of the in silico approach and the participation of the selected candidate regulators in CD.

b. In intestinal epithelial cells (C2BBe1), TFs CREB1 and IRF1 translocate to the nucleus upon gliadin challenge. Additionally, IRF1 binds to the promoter of two newly proposed targets that were identified from a DCG set, namely *HDAC4* and *WDR43*, although gliadin did not cause differences in binding.

These results point to the ability of gliadin to influence gene expression through its effect on TFs, but also stress the limitations of currently available cell models.

3. The overlap among conserved TADs, and CD-related DCGs and SNPs is able to identify 25 regions in which the merge or the disruption of TADs could be involved in CD. CRISPR-Cas9 deletion of the TAD boundaries in two such regions (*PROCR-ROMO1* region as a merge and *HSCB-XBP1* region as a disruption) in cell models provokes changes at the co-expression level, although the different cell lines studied show opposite trends.

These results propose the 3D genome as another potential regulator of local gene co-expression that can be altered in disease. Additionally, they suggest that the genotype of disease-associated SNPs located in the candidate regions could explain the cell line-dependent differences.

4. Gliadin induces hypomethylation of the *HDAC4 locus* in duodenal biopsies from non-celiac individuals. The change is not observed in biopsies from GFD-treated CD patients upon gliadin challenge, because there is a chronic hypomethylation at this *locus* when compared to controls. On the other hand, the DMRs in the promoters of *HLA-B* and *TAP1* previously identified in intestinal cell populations from active CD patients were replicated when GFD patients were compared to controls.

Overall, these data show a methylation signature in CD that is either constitutive (and therefore heritable or acquired very early in development) or non-reversible by GFD. In any case, DNA methylation must be taken into account when addressing the genomics of CD.

Bibliography

Abadie, V., Sollid, L.M., Barreiro, L.B., and Jabri, B. (2011). Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis. Annu. Rev. Immunol. *29*, 493–525.

Agardh, E., Lundstig, A., Perfilyev, A., Volkov, P., Freiburghaus, T., Lindholm, E., Rönn, T., Agardh, C.D., and Ling, C. (2015). Genome-wide analysis of DNA methylation in subjects with type 1 diabetes identifies epigenetic modifications associated with proliferative diabetic retinopathy. BMC Med. *13*, 182.

Al-hassany, M. (1975). Coeliac disease in Iraqi children. J. Trop. Pediatr. *21*, 178–179.

Al-Sadi, R., Guo, S., Ye, D., and Ma, T.Y. (2013). TNF-α modulation of intestinal epithelial tight junction barrier is regulated by ERK1/2 activation of Elk-1. Am. J. Pathol. *183*, 1871–1884.

Al-Tawaty, A.I., and Elbargathy, S.M. (1998). Coeliac disease in north-eastern Libya. Ann. Trop. Paediatr. *18*, 27–30.

Allen, E.K., Randolph, A.G., Bhangale, T., Dogra, P., Ohlson, M., Oshansky, C.M., Zamora, A.E., Shannon, J.P., Finkelstein, D., Dressen, A., et al. (2017). SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. Nat. Med. *23*, 975–983.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anderson, R.P., Degano, P., Godkin, A.J., Jewell, D.P., and Hill, A.V.S. (2000). In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. Nat. Med. *6*, 337–342.

Arentz-Hansen, H., Körner, R., Molberg, O., Quarsten, H., Vader, W., Kooy, Y.M., Lundin, K.E., Koning, F., Roepstorff, P., Sollid, L.M., et al. (2000). 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. *191*, 603–612.

Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851–1864.

Bae, Y.S., Oh, H., Rhee, S.G., and Yoo, Y.D. (2011). Regulation of reactive oxygen species generation in cell signaling. Mol Cells *32*, 491–509.

Banka, S., de Goede, C., Yue, W.W., Morris, A.A.M., von Bremen, B., Chandler, K.E., Feichtinger, R.G., Hart, C., Khan, N., Lunzer, V., et al. (2014). Expanding the clinical and molecular spectrum of thiamine pyrophosphokinase deficiency:

A treatable neurological disorder caused by TPK1 mutations. Mol. Genet. Metab. *113*, 301–306.

Barutcu, A.R., Maass, P.G., Lewandowski, J.P., Weiner, C.L., and Rinn, J.L. (2018). A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. Nat. Commun. *9*, 1444.

Batlle-López, A., Cortiguera, M.G., Rosa-Garrido, M., Blanco, R., Del Cerro, E., Torrano, V., Wagner, S.D., and Delgado, M.D. (2015). Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus. Oncogene *34*, 246–256.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321–1325.

van Belzen, M.J., Koeleman, B.P.C., Crusius, J.B.A., Meijer, J.W.R., Bardoel, A.F.J., Pearson, P.L., Sandkuijl, L.A., Houwen, R.H.J., and Wijmenga, C. (2004). Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients. Genes Immun. *5*, 215–220.

Van Belzen, M.J., Meijer, J.W.R., Sandkuijl, L.A., Bardoel, A.F.J., Mulder, C.J.J., Pearson, P.L., Houwen, R.H.J., and Wijmenga, C. (2003). A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology *125*, 1032–1041.

Bessa, C., Soares, J., Raimundo, L., Loureiro, J.B., Gomes, C., Reis, F., Soares, M.L., Santos, D., Dureja, C., Chaudhuri, S.R., et al. (2018). Discovery of a small-molecule protein kinase Cδ-selective activator with promising application in colon cancer therapy. Cell Death Dis. *9*, 23.

Bhagwat, A.S., and Vakoc, C.R. (2015). Targeting Transcription Factors in Cancer. Trends in Cancer *1*, 53–65.

Bickmore, W.A. (2013). The Spatial Organization of the Human Genome. Annu. Rev. Genomics Hum. Genet. *14*, 67–84.

Bind, M.A., Zanobetti, A., Gasparrini, A., Peters, A., Coull, B., Baccarelli, A., Tarantini, L., Koutrakis, P., Vokonas, P., and Schwartz, J. (2014). Effects of temperature and relative humidity on DNA methylation. Epidemiology *25*, 561–569.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bodé, S., and Gudmand-Høyer, E. (1996). Incidence and prevalence of adult coeliac disease within a defined geographic area in Denmark. Scand. J. Gastroenterol. *31*, 694–699.

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

Borchert, G.M., Lanier, W., and Davidson, B.L. (2006). RNA polymerase III transcribes human microRNAs. Nat. Struct. Mol. Biol. *13*, 1097–1101.

Bouhlel, M.A., Lambert, M., and David-Cordonnier, M.-H. (2015). Targeting Transcription Factor Binding to DNA by Competing with DNA Binders as an Approach for Controlling Gene Expression. Curr. Top. Med. Chem. *15*, 1323–1358.

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

Bratanic, N., Smigoc Schweiger, D., Mendez, A., Bratina, N., Battelino, T., and Vidan-Jeras, B. (2010). An influence of HLA-A, B, DR, DQ, and MICA on the occurrence of Celiac disease in patients with type 1 diabetes. Tissue Antigens *76*, 208–215.

Breker, M., and Schuldiner, M. (2014). The emergence of proteome-wide technologies: Systematic analysis of proteins comes of age. Nat. Rev. Mol. Cell Biol. *15*, 453–464.

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

Capuano, M., Iaffaldano, L., Tinto, N., Montanaro, D., Capobianco, V., Izzo, V., Tucci, F., Troncone, G., Greco, L., and Sacchetti, L. (2011). MicroRNA-449a overexpression, reduced NOTCH1 signals and scarce goblet cells characterize the small intestine of celiac patients. PLoS One *6*, e29094.

Caputo, I., Barone, M.V., Martucciello, S., Lepretti, M., and Esposito, C. (2009). Tissue transglutaminase in celiac disease: Role of autoantibodies. Amino Acids *36*, 693–699.

Castellanos-Rubio, A., and Bilbao, J.R. (2017). Profiling Celiac Disease-Related Transcriptional Changes. Int. Rev. Cell Mol. Biol. *336*, 149–174.

Castellanos-Rubio, A., Martin-Pagola, A., Santin, I., Hualde, I., Aransay, A.M., Castaño, L., Vitoria, J.C., and Bilbao, J.R. (2008). Combined Functional and Positional Gene Information for the Identification of Susceptibility Variants in Celiac Disease. Gastroenterology *134*, 738–746.

Castellanos-Rubio, A., Santin, I., Irastorza, I., Castaño, L., Carlos Vitoria, J., and Bilbao, J.R. (2009). TH17 (and TH1) signatures of intestinal biopsies of CD

patients in response to gliadin. Autoimmunity *42*, 69–73.

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

Castellanos-Rubio, A., Fernandez-Jimenez, N., Kratchmarov, R., Luo, X., Bhagat, G., Green, P.H.R., Schneider, R., Kiledjian, M., Bilbao, J.R., and Ghosh, S. (2016). A long noncoding RNA associated with susceptibility to celiac disease. Science (80-.). *352*, 91–95.

Cataldo, F., and Montalto, G. (2007). Celiac disease in the developing countries: a new and challenging public health problem. World J. Gastroenterol. *13*, 2153–2159.

Catassi, C., Gatti, S., and Fasano, A. (2014). The new epidemiology of celiac disease. J. Pediatr. Gastroenterol. Nutr. *59*, 7–9.

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

Chung, J.S., Lee, S., and Yoo, Y. Do (2014). Constitutive NF-κB activation and tumor-growth promotion by Romo1-mediated reactive oxygen species production. Biochem. Biophys. Res. Commun. *450*, 1656–1661.

Cicerone, C., Nenna, R., and Pontone, S. (2015). Th17, intestinal microbiota and the abnormal immune response in the pathogenesis of celiac disease. Gastroenterol. Hepatol. from Bed to Bench *8*, 117–122.

Collin, P., Reunala, T., Rasmussen, M., Kyrönpalo, S., Pehkonen, E., Laippala, P., and Mäki, M. (1997). High incidence and prevalence of adult coeliac disease: Augmented diagnostic approach. Scand. J. Gastroenterol. *32*, 1129–1133.

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. Nat Rev Genet *10*, 184–194.

Dalmay, T. (2008). Identification of genes targeted by microRNAs. Biochem. Soc. Trans. *36*, 1194–1196.

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. FEBS Lett. *589*, 2877–2884.

Delpu, Y., Cordelier, P., Cho, W.C., and Torrisani, J. (2013). DNA methylation and cancer diagnosis. Int. J. Mol. Sci. *14*, 15029–15058.

Dicke, W.K. (1950). Coeliake. MD Thesis.

Dieli-Crimi, R., Cénit, M.C., and Núñez, C. (2015). The genetics of celiac disease: A comprehensive review of clinical implications. J. Autoimmun. *64*, 26–41.

Le Dily, F.L., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H.G., Ballare, C., Filion, G., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. *28*, 2151–2162.

Diop-Bove, N.K., Wu, J., Zhao, R., Locker, J., and Goldman, I.D. (2009). Hypermethylation of the human proton-coupled folate transporter (SLC46A1) minimal transcriptional regulatory region in an antifolate-resistant HeLa cell line. Mol. Cancer Ther. *8*, 2424–2431.

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

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol. Cell *62*, 668–680.

Dubé, C., Rostom, A., Sy, R., Cranney, A., Saloojee, N., Garritty, C., Sampson, M., Zhang, L., Yazdi, F., Mamaladze, V., et al. (2005). The prevalence of celiac disease in average-risk and at-risk Western European populations: A systematic review. Gastroenterology *128*, S57-67.

Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A.R., Ádány, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. *42*, 295–302.

Erson, A.E., and Petty, E.M. (2008). MicroRNAs in development and disease. Clin. Genet. *74*, 296–306.

Escudero-Hernández, C., Plaza-Izurieta, L., Garrote, J.A., Bilbao, J.R., and Arranz, E. (2017). Association of the IL-15 and IL-15Rα genes with celiac disease. Cytokine *99*, 73–79.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat. Rev. Genet. *12*, 861–874.

Fang, Z., Lin, A., Chen, J., Zhang, X., Liu, H., Li, H., Hu, Y., Zhang, X., Zhang, J., Qiu, L., et al. (2016). CREB1 directly activates the transcription of ribonucleotide reductase small subunit M2 and promotes the aggressiveness of

human colorectal cancer. Oncotarget *7*, 78055–78068.

Fasano, A., and Catassi, C. (2001). Current approaches to diagnosis and treatment of celiac disease: An evolving spectrum. Gastroenterology *120*, 636–651.

Fearnley, L.G., and Inouye, M. (2016). Metabolomics in epidemiology: from metabolite concentrations to integrative reaction networks. Int. J. Epidemiol. *45*, 1319–1328.

Feighery, C. (1999). Coeliac disease. Bmj *319*, 236–239.

Feighery, C., Weir, D.G., Whelan, A., Willoughby, R., Youngprapakorn, S., Lynch, S., O'Moráin, C., McEneany, P., and O'Farrelly, C. (1998). Diagnosis of gluten-sensitive enteropathy: is exclusive reliance on histology appropriate? Eur. J. Gastroenterol. Hepatol. *10*, 919–925.

Felli, C., Baldassarre, A., and Masotti, A. (2017). Intestinal and circulating micrornas in coeliac disease. Int. J. Mol. Sci. *18*.

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

Fina, D., Sarra, M., Caruso, R., Del Vecchio Blanco, G., Pallone, F., MacDonald, T.T., and Monteleone, G. (2008). Interleukin 21 contributes to the mucosal T helper cell type 1 response in coeliac disease. Gut *57*, 887–892.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Folk, J.E., and Chung, S.I. (1985). Transglutaminases. Methods Enzymol. *113*, 358–375.

Folk, J.E., and Cole, P.W. (1966). Transglutaminase: Mechanistic features of the active site as determined by kinetic and inhibitor studies. BBA - Enzymol. Biol. Oxid. *122*, 244–264.

Freeman, H.J. (2013). Non-dietary forms of treatment for adult celiac disease. World J. Gastrointest. Pharmacol. Ther. *4*, 108–112.

Friendly, M. (2002). Corrgrams: Exploratory displays for correlatigon matrices. Am. Stat. *56*, 316–324.

Frietze, S., and Farnham, P.J. (2011). Transcription factor effector domains.

Subcell. Biochem. *52*, 261–277.

Fulton, D.L.L., Sundararajan, S., Badis, G., Hughes, T.R.R., Wasserman, W.W.W., Roach, J.C.C., and Sladek, R. (2009). TFCat: The curated catalog of mouse and human transcription factors. Genome Biol. *10*, R29.

Galvao, L.C., Gomes, R.C., and Ramos, A.M. (1992). [Celiac disease: report of 20 cases in Rio Grande do Norte, Brazil]. Arq Gastroenterol *29*, 28–33.

Garcia-Etxebarria, K., Jauregi-Miguel, A., Romero-Garmendia, I., Plaza-Izurieta, L., Legarda, M., Irastorza, I.I., and Bilbao, J.R.J.R. (2016). Ancestry-based stratified analysis of Immunochip data identifies novel associations with celiac disease. Eur. J. Hum. Genet. *24*, 1831–1834.

Gargalovic, P.S., Gharavi, N.M., Clark, M.J., Pagnon, J., Yang, W.P., He, A., Truong, A., Baruch-Oren, T., Berliner, J.A., Kirchgessner, T.G., et al. (2006). The unfolded protein response is an important regulator of inflammatory genes in endothelial cells. Arterioscler. Thromb. Vasc. Biol. *26*, 2490–2496.

Garten, A., Petzold, S., Schuster, S., Körner, A., Kratzsch, J., and Kiess, W. (2011). Nampt and its potential role in inflammation and type 2 diabetes. Handb. Exp. Pharmacol. *203*, 147–164.

Garvie, C.W., Stagno, J.R., Reid, S., Singh, A., Harrington, E., and Boss, J.M. (2007). Characterization of the RFX complex and the RFX5(L66A) mutant: Implications for the regulation of MHC class II gene expression. Biochemistry *46*, 1597–1611.

Gee, S. (1888). On the celiac disease. St Bart Hosp Rep. *24*, 17–20.

Geertz, M., Shore, D., and Maerkl, S.J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. Proc. Natl. Acad. Sci. *109*, 16540–16545.

Ghirlando, R., and Felsenfeld, G. (2016). CTCF: Making the right connections. Genes Dev. *30*, 881–891.

Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., Vaula, G., Stamoulis, G., Santoni, F., Atzori, C., et al. (2014). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: A second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). Hum. Mol. Genet. *24*, 3143–3154.

Goel, G., King, T., Daveson, A.J., Andrews, J.M., Krishnarajah, J., Krause, R., Brown, G.J.E., Fogel, R., Barish, C.F., Epstein, R., et al. (2017). Epitope-specific immunotherapy targeting CD4-positive T cells in coeliac disease: two randomised, double-blind, placebo-controlled phase 1 studies. Lancet Gastroenterol. Hepatol. *2*, 479–493.

Gong, J., Tian, J., Lou, J., Wang, X., Ke, J., Li, J., Yang, Y., Gong, Y., Zhu, Y., Zou, D., et al. (2018). A polymorphic MYC response element in KBTBD11 influences colorectal cancer risk, especially in interaction with an MYC-regulated SNP rs6983267. Ann. Oncol. *29*, 632–639.

Greco, L., Corazza, G., Babron, M.C., Clot, F., Fulchignoni-Lataud, M.C., Percopo, S., Zavattari, P., Bouguerra, F., Dib, C., Tosi, R., et al. (1998). Genome search in celiac disease. Am. J. Hum. Genet. *62*, 669–675.

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M., Paparo, F., Gasperi, V., Limongelli, M.G., Cotichini, R., et al. (2002). The first large population based twin study of coeliac disease. Gut *50*, 624–628.

Green, P.H.R., and Cellier, C. (2007). Celiac disease. N. Engl. J. Med. *357*, 1731–1743.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. *34*, D140–D144.

Gross, D., and Garrard, W.T. (1988). Nuclease Hypersensitive Sites In Chromatin. Annu. Rev. Biochem. *57*, 159–197.

Guo, M., Mao, X., Ji, Q., Lang, M., Li, S., Peng, Y., Zhou, W., Xiong, B., and Zeng, Q. (2010). Inhibition of IFN regulatory factor-1 down-regulate Th1 cell function in patients with acute coronary syndrome. J. Clin. Immunol. *30*, 241–252.

Gupta, B., and Hawkins, R.D. (2015). Epigenomics of autoimmune diseases. Immunol. Cell Biol. *93*, 271–276.

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

Hang, L.-W., Hsia, T.-C., Chen, W.-C., Chen, H.-Y., and Tsai, F.-J. (2003). TAP1 gene AccI polymorphism is associated with atopic bronchial asthma. J. Clin. Lab. Anal. *17*, 57–60.

Harris, K.M., Fasano, A., and Mann, D.L. (2010). Monocytes differentiated with IL-15 support Th17 and Th1 responses to wheat gliadin: Implications for celiac disease. Clin. Immunol. *135*, 430–439.

He, D., Liu, Z.-P., Honda, M., Kaneko, S., and Chen, L. (2012). Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. J. Mol. Cell Biol. *4*, 140–152.

He, Z., Proudfoot, C., Whitelaw, C.B.A., and Lillico, S.G. (2016). Comparison of CRISPR/Cas9 and TALENs on editing an integrated EGFP gene in the genome of HEK293FT cells. Springerplus *5*, 814.

van Heel, D.A., Franke, L., Hunt, K.A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M.C., Barnardo, M.C., Bethel, G., Holmes, G.K., et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat. Genet. *39*, 827–829.

Hogen Esch, C.E., Rosen, A., Auricchio, R., Romanos, J., Chmielewska, A., Putter, H., Ivarsson, A., Szajewska, H., Koning, F., Wijmenga, C., et al. (2010). The PreventCD Study design: towards new strategies for the prevention of coeliac disease. Eur. J. Gastroenterol. Hepatol. *22*, 1424–1430.

Holmes, E.E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L.A., Uhl, B., Kristiansen, G., et al. (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. PLoS One *9*, e93933.

Holopainen, P., Naluai, A.T., Moodie, S., Percopo, S., Coto, I., Clot, F., Ascher, H., Sollid, L., Ciclitira, P., Greco, L., et al. (2004). Candidate gene region 2q33 in European families with coeliac disease. Tissue Antigens *63*, 212–222.

Hong, S., and Kim, D. (2017). Computational characterization of chromatin domain boundary-associated genomic elements. Nucleic Acids Res. *45*, 10403–10414.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Wright, M.W., et al. (2004). Gene map of the extended human MHC. Nat. Rev. Genet. *5*, 889–899.

Hüe, S., Mention, J.J., Monteiro, R.C., Zhang, S.L., Cellier, C., Schmitz, J., Verkarre, V., Fodil, N., Bahram, S., Cerf-Bensussan, N., et al. (2004). A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. Immunity *21*, 367–377.

Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: New tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. *43*, D117–D122.

Hunt, K.A., Mistry, V., Bockett, N.A., Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O., et al. (2013). Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature *498*, 232–235.

Husby, S., Koletzko, S., Korponay-Szabo´, I., Mearin, M., Phillips, A., Shamir, R., Troncone, R., Giersiepen, K., Branksi, D., Catasssi, C., et al. (2012).

European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Guidelines for the Diagnosis of Coeliac Disease. J. Pediatr. Gastroenterol. Nutr. *54*, 136–160.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Ivarsson, A., Persson, L., Nyström, L., Ascher, H., Cavell, B., Danielsson, L., Dannaeus, A., Lindberg, T., Lindquist, B., Stenhammar, L., et al. (2000). Epidemic of coeliac disease in Swedish children. Acta Paediatr. Int. J. Paediatr. *89*, 165–171.

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

Jauregi-Miguel, A., Fernandez-Jimenez, N., Irastorza, I., Plaza-Izurieta, L., Vitoria, J.C., and Bilbao, J.R. (2014). Alteration of tight junction gene expression in celiac disease. J. Pediatr. Gastroenterol. Nutr. *58*, 762–767.

Jheng, J.R., Lin, C.Y., Horng, J.T., and Lau, K.S. (2012). Inhibition of enterovirus 71 entry by transcription factor XBP1. Biochem. Biophys. Res. Commun. *420*, 882–887.

Ji, X., Dadon, D.B., Powell, B.E., Misteli, T., Jaenisch, R., Young, R.A., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells Cell Stem Cell Resource 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262–275.

Jin, C., Jin, Z., Chen, N.Z., Lu, M., Liu, C.B., Hu, W. Le, and Zheng, C.G. (2016). Activation of IRE1α-XBP1 pathway induces cell proliferation and invasion in colorectal carcinoma. Biochem. Biophys. Res. Commun. *470*, 75–81.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388.

Jones, P.A., Baylin, S.B., Erdjument-Bromage, H., Tempst, P., Bird, A., Reinberg, D., Sato, F., Meltzer, S.J., Sidransky, D., Badenhorst, P., et al. (2007). The Epigenomics of Cancer. Cell *128*, 683–692.

Juuti-Uusitalo, K., Mäki, M., Kaukinen, K., Collin, P., Visakorpi, T., Vihinen, M., and Kainulainen, H. (2004). cDNA microarray analysis of gene expression in

coeliac disease jejunal biopsy samples. J. Autoimmun. *22*, 249–265.

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

Karell, K., Louka, A.S., Moodie, S.J., Ascher, H., Clot, F., Greco, L., Ciclitira, P.J., Sollid, L.M., and Partanen, J. (2003). 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. *64*, 469–477.

Kaser, A., Lee, A.-H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E.S., Higgins, D.E., Schreiber, S., Glimcher, L.H., et al. (2008). XBP1 Links ER Stress to Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease. Cell *134*, 743–756.

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A.E., Ristolainen, H., Hänninen, U.A., Cajuso, T., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. Nat. Genet. *47*, 818–821.

de Kauwe, A.L., Chen, Z., Anderson, R.P., Keech, C.L., Price, J.D., Wijburg, O., Jackson, D.C., Ladhams, J., Allison, J., and McCluskey, J. (2009). Resistance to Celiac Disease in Humanized HLA-DR3-DQ2-Transgenic Mice Expressing Specific Anti-Gliadin CD4+ T Cells. J. Immunol. *182*, 7440–7450.

Khuffash, F.A., Barakat, M.H., Shaltout, A.A., Farwana, S.S., Adnani, M.S., and Tungekar, M.F. (1987). Coeliac disease among children in Kuwait: Difficulties in diagnosis and management. Gut *28*, 1595–1599.

Kim, H.J., Jo, M.J., Kim, B.R., Kim, J.L., Jeong, Y.A., Na, Y.J., Park, S.H., Lee, S.Y., Lee, D.H., Lee, H.S., et al. (2017). Reactive oxygen species modulator-1 (Romo1) predicts unfavorable prognosis in colorectal cancer patients. PLoS One *12*, e0176834.

Kim, M.S., Chang, X., Yamashita, K., Nagpal, J.K., Baek, J.H., Wu, G., Trink, B., Ratovitski, E.A., Mori, M., and Sidransky, D. (2008). Aberrant promoter methylation and tumor suppressive activity of the DFNA5 gene in colorectal carcinoma. Oncogene *27*, 3624–3634.

Kishi, Y., Kondo, T., Xiao, S., Yosef, N., Gaublomme, J., Wu, C., Wang, C., Chihara, N., Regev, A., Joller, N., et al. (2016). Protein C receptor (PROCR) is a negative regulator of Th17 pathogenicity. J. Exp. Med. *213*, 2489–2501.

Kitamura, K., Seike, M., Okano, T., Matsuda, K., Miyanaga, A., Mizutani, H., Noro, R., Minegishi, Y., Kubota, K., and Gemma, A. (2014). MiR-134/487b/655 cluster regulates TGF-β-induced epithelial-mesenchymal transition and drug resistance to gefitinib by targeting MAGI2 in lung adenocarcinoma cells. Mol. Cancer Ther. *13*, 444–453.

Kohl, M., Megger, D.A., Trippler, M., Meckel, H., Ahrens, M., Bracht, T., Weber, F., Hoffmann, A.C., Baba, H.A., Sitek, B., et al. (2014). A practical data processing workflow for multi-OMICS projects. Biochim. Biophys. Acta *1844*, 52–62.

Kolodziej, L.E., Lodolce, J.P., Chang, J.E., Schneider, J.R., Grimm, W.A., Bartulis, S.J., Zhu, X., Messer, J.S., Murphy, S.F., Reddy, N., et al. (2011). TNFAIP3 maintains intestinal barrier function and supports epithelial cell tight junctions. PLoS One *6*.

Krueger, F., and Andrews, S.R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Krumm, A., and Duan, Z. (2018). Understanding the 3D genome: Emerging impacts on human disease. Semin. Cell Dev. Biol.

Kuo, P.-L., Liao, S.-H., Hung, J.-Y., Huang, M.-S., and Hsu, Y.-L. (2013). MicroRNA-33a functions as a bone metastasis suppressor in lung cancer by targeting parathyroid hormone related protein. Biochim. Biophys. Acta *1830*, 3756–3766.

de la Fuente, A. (2010). From "differential expression" to "differential networking" - identification of dysfunctional regulatory networks in diseases. Trends Genet. *26*, 326–333.

Lahdenperä, A., Ludvigsson, J., Fälth-Magnusson, K., Högberg, L., and Vaarala, O. (2011). The effect of gluten-free diet on Th1-Th2-Th3-associated intestinal immune responses in celiac disease. Scand. J. Gastroenterol. *46*, 538–549.

Lai, L., Song, Y., Liu, Y., Chen, Q., Han, Q., Chen, W., Pan, T., Zhang, Y., Cao, X., and Wang, Q. (2013). MicroRNA-92a negatively regulates toll-like receptor (TLR)-triggered inflammatory response in macrophages by targeting MKK4 kinase. J. Biol. Chem. *288*, 7956–7967.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lebreton, C., Ménard, S., Abed, J., Moura, I.C., Coppo, R., Dugave, C., Monteiro, R.C., Fricot, A., Traore, M.G., Griffin, M., et al. (2012). Interactions among secretory immunoglobulin A, CD71, and transglutaminase-2 affect permeability of intestinal epithelial cells to gliadin peptides. Gastroenterology *143*, 698–707.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lee, S., Park, Y.H., Chung, J.S., and Yoo, Y.D. (2015). Romo1 and the NF-κB pathway are involved in oxidative stress-induced tumor cell invasion. Int. J. Oncol. *46*, 2021–2028.

Leffler, D.A., Kelly, C.P., Green, P.H.R., Fedorak, R.N., Dimarino, A., Perrow, W., Rasmussen, H., Wang, C., Bercik, P., Bachir, N.M., et al. (2015). Larazotide acetate for persistent symptoms of celiac disease despite a gluten-free diet: A randomized controlled trial. Gastroenterology *148*, 1311–1319.

Lenna, S., Assassi, S., Farina, G.A., Mantero, J.C., Scorza, R., Lafyatis, R., Farber, H.W., and Trojanowska, M. (2015). The HLA-B*35 allele modulates ER stress, inflammation and proliferation in PBMCs from Limited Cutaneous Systemic Sclerosis patients. Arthritis Res. Ther. *17*, 363.

León, A.J., Garrote, J.A., Blanco-Quirós, A., Calvo, C., Fernández-Salazar, L., Del Villar, A., Barrera, A., and Arranz, E. (2006). Interleukin 18 maintains a long-standing inflammation in coeliac disease patients. Clin. Exp. Immunol. *146*, 479–485.

Lerner, A. (2010). New therapeutic strategies for celiac disease. Autoimmun. Rev. *9*, 144–147.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J.J., Chen, S., Wang, Q., and Pan, H. (2018). Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. Sci. Rep. *8*, 622.

Liang, H., Cheung, L.W.T., Li, J., Ju, Z., Yu, S., Stemke-Hale, K., Dogruluk, T., Lu, Y., Liu, X., Gu, C., et al. (2012). Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. Genome Res. *22*, 2120–2129.

Lindfors, K., Kaukinen, K., and Mäki, M. (2009). A role for anti-transglutaminase 2 autoantibodies in the pathogenesis of coeliac disease? Amino Acids *36*, 685–691.

Ling, A., Löfgren-Burström, A., Larsson, P., Li, X., Wikberg, M.L., Öberg, Å., Stenling, R., Edin, S., and Palmqvist, R. (2017). TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. Oncoimmunology *6*, e1356143.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lou, S., Lee, H.M., Qin, H., Li, J.W., Gao, Z., Liu, X., Chan, L.L., Lam, V.K.L., So, W.Y., Wang, Y., et al. (2014). Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome Biol. *15*, 408.

Louka, a S., and Sollid, L.M. (2003). HLA in coeliac disease: unravelling the complex genetics of a complex disorder. Tissue Antigens *61*, 105–117.

Ludwig, H., Polymenidis, Z., Granditsch, G., and Wick, G. (1973). Association of HL-A1 and HL-A8 with childhood celiac disease. Z. Immunitatsforsch. Exp. Klin. Immunol. *146*, 158–167.

Lundin, K.E., Sollid, L.M., Qvigstad, E., Markussen, G., Gjertsen, H. a, Ek, J., and Thorsby, E. (1990). T lymphocyte recognition of a celiac disease-associated cis- or trans-encoded HLA-DQ alpha/beta-heterodimer. J. Immunol. *145*, 136–139.

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

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet. *32*, 225–237.

Madrigano, J., Baccarelli, A., Mittleman, M.A., Wright, R.O., Sparrow, D., Vokonas, P.S., Tarantini, L., and Schwartz, J. (2011). Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. Environ. Health Perspect. *119*, 977–982.

Magni, S., Comani, G.B., Elli, L., Vanessi, S., Ballarini, E., Nicolini, G., Rusconi, M., Castoldi, M., Meneveri, R., Muckenthaler, M.U., et al. (2014). miRNAs Affect the Expression of Innate and Adaptive Immunity Proteins in Celiac Disease. Am. J. Gastroenterol. *109*, 1662–1674.

Maiuri, L., Troncone, R., Mayer, M., Coletta, S., Picarelli, A., De Vincenzi, M., Pavone, V., and Auricchio, S. (1996). In vitro Activities of A-Gliadin-Related Synthetic Peptides: Damaging Effect on the Atrophic Coeliac Mucosa and Activation of Mucosal Immune Response in the Treated Coeliac Mucosa. Scand. J. Gastroenterol. *31*, 247–253.

Maiuri, L., Ciacci, C., Ricciardelli, I., Vacca, L., Raia, V., Auricchio, S., Picard,

J., Osman, M., Quaratino, S., and Londei, M. (2003). Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. Lancet *362*, 30–37.

Maiuri, L., Luciani, A., Villella, V.R., Vasaturo, A., Giardino, I., Pettoello-Mantovani, M., Guido, S., Cexus, O.N., Peake, N., Londei, M., et al. (2010). Lysosomal accumulation of gliadin p31-43 peptide induces oxidative stress and tissue transglutaminase-mediated PPARγ downregulation in intestinal epithelial cells and coeliac mucosa. Gut *59*, 311–319.

Mäki, M. (1995). The humoral immune system in coeliac disease. Baillieres. Clin. Gastroenterol. *9*, 231–249.

Mäki, M., and Collin, P. (1997). Coeliac disease. Lancet *349*, 1755–1759.

Marsh, M.N. (1992). Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology *102*, 330–354.

Martin-Pagola, A., Ortiz, L., De Nanclares, G.P., Vitoria, J.C., Castaño, L., and Bilbao, J.R. (2003). Analysis of the Expression of MICA in Small Intestinal Mucosa of Patients with Celiac Disease. J. Clin. Immunol. *23*, 498–503.

Martín-Pagola, A., Pérez-Nanclares, G., Ortiz, L., Vitoria, J.C., Hualde, I., Zaballa, R., Preciado, E., Castaño, L., and Bilbao, J.R. (2004). MICA response to gliadin in intestinal mucosa from celiac patients. Immunogenetics *56*, 549–554.

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., et al. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat. Commun. *6*, 10069.

Martinon, F., Chen, X., Lee, A.H., and Glimcher, L.H. (2010). TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. Nat. Immunol. *11*, 411–418.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115.

Matys, V., Kel-Margoulis, O. V, Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. *34*, D108-10.

Mazzarella, G., Maglio, M., Paparo, F., Nardone, G., Stefanile, R., Greco, L., Van de Wal, Y., Kooy, Y., Koning, F., Auricchio, S., et al. (2003). 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 *52*, 57–62.

McDermott, E., Ryan, E.J., Tosetto, M., Gibson, D., Burrage, J., Keegan, D., Byrne, K., Crowe, E., Sexton, G., Malone, K., et al. (2016). DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. J. Crohn's Colitis *10*, 77–86.

McHale, C.M., Zhang, L., Thomas, R., and Smith, M.T. (2013). Analysis of the transcriptome in molecular epidemiology studies. Environ. Mol. Mutagen. *54*, 500–517.

Meddens, C.A., Harakalova, M., van den Dungen, N.A.M., Foroughi Asl, H., Hijma, H.J., Cuppen, E.P.J.G., Björkegren, J.L.M., Asselbergs, F.W., Nieuwenhuis, E.E.S., and Mokry, M. (2016). Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. Genome Biol. *17*, 247.

Medina, I., Carbonell, J., Pulido, L., Madeira, S.C., Goetz, S., Conesa, A., Tárraga, J., Pascual-Montano, A., Nogales-Cadenas, R., Santoyo, J., et al. (2010). Babelomics: An integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res. *38*, 210–213.

Miklos, G.L.G., and Rubin, G.M. (1996). The role of the genome project in determining gene function: Insights from model organisms. Cell *86*, 521–529.

Molberg, Ø., Mcadam, S.N., Körner, R., Quarsten, H., Kristiansen, C., Madsen, L., Fugger, L., Scott, H., Norën, O., Roepstorff, P., et al. (1998). Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. Nat. Med. *4*, 713–717.

Monteleone, G., Pender, S.L., Alstead, E., Hauer, A.C., Lionetti, P., McKenzie, C., and MacDonald, T.T. (2001). Role of interferon alpha in promoting T helper cell type 1 responses in the small intestine in coeliac disease. Gut *48*, 425–429.

Monteleone, I., Sarra, M., Del Vecchio Blanco, G., Paoluzi, O.A., Franzè, E., Fina, D., Fabrizi, A., MacDonald, T.T., Pallone, F., and Monteleone, G. (2010). Characterization of IL-17A–Producing Cells in Celiac Disease Mucosa. J. Immunol. *184*, 2211–2218.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.

Murray, J.A., Van Dyke, C., Plevak, M.F., Dierkhising, R.A., Zinsmeister, A.R., and Melton, L.J. (2003). Trends in the identification and clinical features of

celiac disease in a North American community, 1950-2001. Clin. Gastroenterol. Hepatol. *1*, 19–27.

Murray, J.A., Watson, T., Clearman, B., and Mitros, F. (2004). Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease. Am. J. Clin. Nutr. *79*, 669–673.

Nanayakkara, M., Lania, G., Maglio, M., Auricchio, R., De Musis, C., Discepolo, V., Miele, E., Jabri, B., Troncone, R., Auricchio, S., et al. (2018). P31–43, an undigested gliadin peptide, mimics and enhances the innate immune response to viruses and interferes with endocytic trafficking: a role in celiac disease. Sci. Rep. *8*, 10821.

Nilsen, E.M., Lundin, K.E., Krajci, P., Scott, H., Sollid, L.M., and Brandtzaeg, P. (1995). Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. Gut *37*, 766–776.

Nistico, L., Fagnani, C., Coto, I., Percopo, S., Cotichini, R., Limongelli, M.G., Paparo, F., D'Alfonso, S., Giordano, M., Sferlazzas, C., et al. (2006). Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut *55*, 803–804.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Ong, C.T., and Corces, V.G. (2014). CTCF: An architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Özcan, U., Yilmaz, E., Özcan, L., Furuhashi, M., Vaillancourt, E., Smith, R.O., Görgün, C.Z., and Hotamisligil, G.S. (2006). Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. Science (80-.). *313*, 1137–1140.

Paolella, G., Lepretti, M., Barone, M.V., Nanayakkara, M., Di Zenzo, M., Sblattero, D., Auricchio, S., Esposito, C., and Caputo, I. (2017). Celiac anti-type 2 transglutaminase antibodies induce differential effects in fibroblasts from celiac disease patients and from healthy subjects. Amino Acids *49*, 541–550.

Parmar, A., Greco, D., Venäläinen, J., Gentile, M., Dukes, E., and Saavalainen, P. (2013). Gene Expression Profiling of Gliadin Effects on Intestinal Epithelial Cells Suggests Novel Non-Enzymatic Functions of Pepsin and Trypsin. PLoS One *8*, e66307.

Paterson, B.M., Lammers, K.M., Arrieta, M.C., Fasano, A., and Meddings, J.B. (2007). The safety, tolerance, pharmacokinetic and pharmacodynamic effects of

single doses of AT-1001 in coeliac disease subjects: a proof of concept study. Aliment. Pharmacol. Ther. *26*, 757–766.

Pervjakova, N., and Prokopenko, I. (2017). The TAD-pathway for GWAS signals. Eur. J. Hum. Genet. *25*, 1179–1180.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Picarelli, A., Di Tola, M., Sabbatella, L., Anania, M.C., Di Cello, T., Greco, R., Silano, M., and De Vincenzi, M. (1999). 31-43 amino acid sequence of the alpha-gliadin induces anti-endomysial antibody production during in vitro challenge. Scand. J. Gastroenterol. *34*, 1099–1102.

Pietz, G., De, R., Hedberg, M., Sjöberg, V., Sandström, O., Hernell, O., Hammarström, S., and Hammarström, M.-L. (2017). Immunopathology of childhood celiac disease—Key role of intestinal epithelial cells. PLoS One *12*, e0185025.

Plaza-Izurieta, L., Fernandez-Jimenez, N., Irastorza, I., Jauregi-Miguel, A., Romero-Garmendia, I., Vitoria, J.C.J.C., and Bilbao, J.R.J.R. (2015). Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease. Eur. J. Hum. Genet. *23*, 1–6.

Ploski, R., Ek, J., Thorsby, E., and Sollid, L.M. (1993). On the HLA-DQ(α1*0501, β1*0201)-associated susceptibility in celiac disease: A possible gene dosage effect of DQB1*0201. Tissue Antigens *41*, 173–177.

Plot, L., and Amital, H. (2009). Infectious associations of Celiac disease. Autoimmun. Rev. *8*, 316–319.

Powrózek, T., Mlak, R., Dziedzic, M., Małecka-Massalska, T., and Sagan, D. (2018). Investigation of relationship between precursor of miRNA-944 and its mature form in lung squamous-cell carcinoma - the diagnostic value. Pathol. Res. Pract. *214*, 368–373.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinn, E.M., Coleman, C., Molloy, B., Dominguez Castro, P., Cormican, P., Trimble, V., Mahmud, N., and McManus, R. (2015). Transcriptome Analysis of CD4+ T Cells in Coeliac Disease Reveals Imprint of BACH2 and IFNγ Regulation. PLoS One *10*, e0140049.

Rabassa, E.B., Sagaró, E., Fragoso, T., Castañeda, C., and Gra, B. (1981). Coeliac disease in Cuban children. Arch. Dis. Child. *56*, 128–131.

Ráki, M., Schjetne, K.W., Stamnaes, J., Molberg, Jahnsen, F.L., Issekutz, T.B., Bogen, B., and Sollid, L.M. (2007). 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. *65*, 213–220.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rauhavirta, T., Qiao, S.W., Jiang, Z., Myrsky, E., Loponen, J., Korponay-Szabó, I.R., Salovaara, H., Garcia-Horsman, J.A., Venäläinen, J., Männistö, P.T., et al. (2011). Epithelial transport and deamidation of gliadin peptides: A role for coeliac disease patient immunoglobulin A. Clin. Exp. Immunol. *164*, 127–136.

Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. *6*, 597–610.

Rodón, L., Gonzàlez-Juncà, A., Del Mar Inda, M., Sala-Hojman, A., Martínez-Sáez, E., and Seoane, J. (2014). Active creb1 promotes a malignant TGFβ2 autocrine loop in glioblastoma. Cancer Discov. *4*, 1230–1241.

Rodríguez-Fraticelli, A.E., Bagwell, J., Bosch-Fortea, M., Boncompain, G., Reglero-Real, N., García-León, M.J., Andrés, G., Toribio, M.L., Alonso, M.A., Millán, J., et al. (2015). Developmental regulation of apical endocytosis controls epithelial patterning in vertebrate tubular organs. Nat. Cell Biol. *17*, 241–250.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. Genome Res. *14*, 1902–1910.

Rosenfeld, M.G., Lunyak, V. V., and Glass, C.K. (2006). Sensors and signals: A coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes Dev. *20*, 1405–1428.

Rubio-Tapia, A., and Murray, J.A. (2010). Classification and management of refractory coeliac disease. Gut *59*, 547–557.

Di Sabatino, A., and Corazza, G.R. (2009). Coeliac disease. Lancet *373*, 1480–1493.

Sagaro, E., and Jimenez, N. (1981). Family studies of coeliac disease in Cuba. Arch. Dis. Child. *56*, 132–133.

Samasca, G., Sur, G., Lupan, I., and Deleanu, D. (2014). Gluten-free diet and

quality of life in celiac disease. Gastroenterol. Hepatol. from Bed to Bench *7*, 139–143.

Sari, C., Bayram, N.A., Doğan, F.E.A., Baştuğ, S., Bolat, A.D., Sarı, S.Ö., Ersoy, O., and Bozkurt, E. (2012). The evaluation of endothelial functions in patients with celiac disease. Echocardiography *29*, 471–477.

Sarra, M., Cupi, M.L., Monteleone, I., Franzè, E., Ronchetti, G., Di Sabatino, A., Gentileschi, P., Franceschilli, L., Sileri, P., Sica, G., et al. (2013). IL-15 positively regulates IL-21 production in celiac disease mucosa. Mucosal Immunol. *6*, 244–255.

Savic, S., Ouboussad, L., Dickie, L.J., Geiler, J., Wong, C., Doody, G.M., Churchman, S.M., Ponchel, F., Emery, P., Cook, G.P., et al. (2014). TLR dependent XBP-1 activation induces an autocrine loop in rheumatoid arthritis synoviocytes. J. Autoimmun. *50*, 59–66.

Sawle, A.D., Kebschull, M., Demmer, R.T., and Papapanou, P.N. (2016). Identification of Master Regulator Genes in Human Periodontitis. J. Dent. Res. *95*, 1010–1017.

Schmitges, F.W., Radovani, E., Najafabadi, H.S., Barazandeh, M., Campitelli, L.F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T., et al. (2016). Multiparameter functional diversity of human C2H2 zinc finger proteins. Genome Res. *26*, 1742–1752.

Schuppan, D., Junker, Y., and Barisani, D. (2009). Celiac Disease: From Pathogenesis to Novel Therapies. Gastroenterology *137*, 1912–1933.

Sexton, T., Schober, H., Fraser, P., and Gasser, S.M. (2007). Gene regulation through nuclear organization. Nat. Struct. Mol. Biol. *14*, 1049–1055.

Shalimar, Das, P., Sreenivas, V., Gupta, S.D., Panda, S.K., and Makharia, G.K. (2013). Mechanism of villous atrophy in celiac disease: Role of apoptosis and epithelial regeneration. Arch. Pathol. Lab. Med. *137*, 1262–1269.

Shan, L., Molberg, Ø., Parrot, I., Hausch, F., Filiz, F., Gray, G.M., Sollid, L.M., and Khosla, C. (2002). Structural basis for gluten intolerance in Celiac Sprue. Science (80-.). *297*, 2275–2279.

Siggers, T., and Gordân, R. (2014). Protein-DNA binding: Complexities and multi-protein codes. Nucleic Acids Res. *42*, 2099–2111.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., Van Steensel, B., and De Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Sjöström, H., Lundin, K.E.A., Molberg, Körner, R., Mcadam, S.N., Anthonsen,

D., Quarsten, H., Norén, O., Roepstorff, P., Thorsby, E., et al. (1998). Identification of a gliadin T-cell epitope in coeliac disease: General importance of gliadin deamidation for intestinal T-cell recognition. Scand. J. Immunol. *48*, 111–115.

Sollid, L.M. (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J. Exp. Med. *169*, 345–350.

Sollid, L.M. (2002). Coeliac disease: Dissecting a complex inflammatory disorder. Nat. Rev. Immunol. *2*, 647–655.

Sollid, L.M., and Thorsby, E. (1993). HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. Gastroenterology *105*, 910–922.

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

Steinman, L. (2007). A brief history of TH17, the first major revision in the TH1/TH2 hypothesis of T cell-mediated tissue damage. Nat. Med. *13*, 139–145.

Stenman, S.M., Lindfors, K., Korponay-Szabo, I.R., Lohi, O., Saavalainen, P., Partanen, J., Haimila, K., Wieser, H., Mäki, M., and Kaukinen, K. (2008). Secretion of celiac disease autoantibodies after in vitro gliadin challenge is dependent on small-bowel mucosal transglutaminase 2-specific IgA deposits. BMC Immunol. *9*, 6.

Sturgess, R., Day, P., Ellis, H.J., Kontakou, M., Ciclitira, P.J., Lundin, K.E.A., and Gjertsen, H.A. (1994). Wheat peptide challenge in coeliac disease. Lancet *343*, 758–761.

Suliman, G.I. (1978). Coeliac disease in Sudanese children. Gut *19*, 121–125.

Sun, B., Hu, L., Luo, Z.Y., Chen, X.P., Zhou, H.H., and Zhang, W. (2016). DNA methylation perspectives in the pathogenesis of autoimmune diseases. Clin. Immunol. *164*, 21–27.

Sun, W., Julie Li, Y.S., Huang, H. Da, Shyy, J.Y., and Chien, S. (2010). microRNA: A Master Regulator of Cellular Processes for Bioengineering Systems. Annu. Rev. Biomed. Eng. *12*, 1–27.

Takeuchi, A., Badr, M.E.S.G., Miyauchi, K., Ishihara, C., Onishi, R., Guo, Z., Sasaki, Y., Ike, H., Takumi, A., Tsuji, N.M., et al. (2016). CRTAM determines the $CD4^+$ cytotoxic T lymphocyte lineage. J. Exp. Med.

Teng, G. gen, Wang, W. hong, Dai, Y., Wang, S. jun, Chu, Y. xiang, and Li, J. (2013). Let-7b Is Involved in the Inflammation and Immune Responses Associated with Helicobacter pylori Infection by Targeting Toll-Like Receptor 4. PLoS One *8*, e56709.

Toosi, S., Orlow, S.J., and Manga, P. (2012). Vitiligo-inducing phenols activate the unfolded protein response in melanocytes resulting in upregulation of IL6 and IL8. J. Invest. Dermatol. *132*, 2601–2609.

Tovar, H., García-Herrera, R., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Transcriptional master regulator analysis in breast cancer genetic networks. Comput. Biol. Chem. *59*, 67–77.

Tran, H., Porter, J., Sun, M.A., Xie, H., and Zhang, L. (2014). Objective and comprehensive evaluation of bisulfite short read mapping tools. Adv. Bioinformatics *2014*, 472045.

Trenkmann, M., Brock, M., Gay, R.E., Michel, B.A., Gay, S., and Huber, L.C. (2013). Tumor necrosis factor α-induced microRNA-18a activates rheumatoid arthritis synovial fibroblasts through a feedback loop in NF-κB signaling. Arthritis Rheum. *65*, 916–927.

Troncone, R., Gianfrani, C., Mazzarella, G., Greco, L., Guardiola, J., Auricchio, S., and De Berardinis, P. (1998). Majority of gliadin-specific T-cell clones from celiac small intestinal mucosa produce interferon-γ and interleukin-4. Dig. Dis. Sci. *43*, 156–161.

Trujillo, R.D., Yue, S.B., Tang, Y., O'Gorman, W.E., and Chen, C.Z. (2010). The potential functions of primary microRNAs in target recognition and repression. EMBO J. *29*, 3272–3285.

Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. *43*, 1193–1201.

Vaira, V., Roncoroni, L., Barisani, D., Gaudioso, G., Bosari, S., Bulfamante, G., Doneda, L., Conte, D., Tomba, C., Bardella, M.T., et al. (2014). microRNA profiles in coeliac patients distinguish different clinical phenotypes and are modulated by gliadin peptides in primary duodenal fibroblasts. Clin. Sci. (Lond). *126*, 417–423.

Vallejo-Diez, S., Bernardo, D., De Moreno, M.L., Muñoz-Suano, A., Fernández-Salazar, L., Calvo, C., Sousa, C., Garrote, J.A., Cebolla, Á., and Arranz, E. (2013). Detection of specific IgA antibodies against a novel deamidated 8-mer gliadin peptide in blood plasma samples from celiac patients. PLoS One *8*, e80982.

Valton, A.L., and Dekker, J. (2016). TAD disruption as oncogenic driver. Curr. Opin. Genet. Dev. *36*, 34–40.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: Function, expression and

evolution. Nat. Rev. Genet. *10*, 252–263.

Vriezinga, S.L., Auricchio, R., Bravi, E., Castillejo, G., Chmielewska, A., Crespo Escobar, P., Kolaček, S., Koletzko, S., Korponay-Szabo, I.R., Mummert, E., et al. (2014). Randomized Feeding Intervention in Infants at High Risk for Celiac Disease. N. Engl. J. Med. *371*, 1304–1315.

Wal, Y. Van De, Kooy, Y., Veelen, P. Van, Peña, S., Mearin, L., Papadopoulos, G., and Koning, F. (1998). Selective Deamidation by Tissue Transglutaminase Strongly Enhances Gliadin-Specific T Cell Reactivity. J. Immunol. *161*, 1585–1588.

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L., and Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics *16*, 49.

Wang, J., Zhuang, J., Iyer, S., Lin, X.Y., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. *22*, 1798–1812.

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

Wapenaar, M.C., Monsuur, A.J., Bodegraven, A.A. Van, Weersma, R.K., Bevova, M.R., Linskens, R.K., Howdle, P., Holmes, G., Mulder, C.J., Dijkstra, G., et al. (2008). Associations with tight junction genes PARD3 and MAGI2 in Dutch patients point to a common barrier defect for coeliac disease and ulcerative colitis Associations with tight junction genes PARD3 and MAGI2 in Dutch patients point to a common barrier defect. *57*, 463–467.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. Cell *158*, 1431–1443.

Weischenfeldt, J., Dubash, T., Drainas, A.P., Mardin, B.R., Chen, Y., Stütz, A.M., Waszak, S.M., Bosco, G., Halvorsen, A.R., Raeder, B., et al. (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat. Genet. *49*, 65–74.

Wen, A.Y., Sakamoto, K.M., and Miller, L.S. (2010). The Role of the Transcription Factor CREB in Immune Function. J. Immunol. *185*, 6413–6419.

Widschwendter, M., Jiang, G., Woods, C., Müller, H.M., Fiegl, H., Goebel, G., Marth, C., Müller-Holzner, E., Zeimet, A.G., Laird, P.W., et al. (2004). DNA hypomethylation and ovarian cancer biology. Cancer Res. *64*, 4472–4480.

Wingender, E. (2000). TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. *28*, 316–319.

Wingender, E., Schoeps, T., Haubrock, M., and Dönitz, J. (2015). TFClass: A classification of human transcription factors and their rodent orthologs. Nucleic Acids Res. *43*, D97–D102.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: MicroRNA biogenesis pathways and their regulation. Nat. Cell Biol. *11*, 228–234.

Yadav, D., Ngolab, J., Lim, R.S.-H., Krishnamurthy, S., and Bui, J.D. (2008). Cutting Edge: Down-Regulation of MHC Class I-Related Chain A on Tumor Cells by IFN- -Induced MicroRNA. J. Immunol. *182*, 39–43.

Yang, D., Xiao, C., Long, F., Su, Z., Jia, W., Qin, M., Huang, M., Wu, W., Suguro, R., Liu, X., et al. (2018). HDAC4 regulates vascular inflammation via activation of autophagy. Cardiovasc. Res. *114*, 1016–1028.

Yang, J., Yu, H., Liu, B.H., Zhao, Z., Liu, L., Ma, L.X., Li, Y.X., and Li, Y.Y. (2013). DCGL v2.0: an R package for unveiling differential regulation from differential co-expression. PLoS One *8*, e79729.

Yao, K., Lee, S.Y., Peng, C., Lim, D.Y., Yamamoto, H., Ryu, J., Lim, T.G., Chen, H., Jin, G., Zhao, Z., et al. (2018). RSK2 is required for TRAF6 phosphorylation-mediated colon inflammation. Oncogene *37*, 3501–3513.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (80-.). *356*.

Yue, S.B., Trujillo, R.D., Tang, Y., O'Gorman, W.E., and Chen, C.Z. (2011). Loop nucleotides control primary and mature miRNA function in target recognition and repression. RNA Biol. *8*, 1115–1123.

Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., et al. (2013). Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. PLoS One *8*, e63812.

Zhang, S., Shan, C., Kong, G., Du, Y., Ye, L., and Zhang, X. (2012). MicroRNA-520e suppresses growth of hepatoma cells by targeting the NF-κB-inducing kinase (NIK). Oncogene *31*, 3607–3620.

Zhang, Y., Schottker, B., Ordonez-Mena, J., Holleczek, B., Yang, R.X., Burwinkel, B., Butterbach, K., and Brenner, H. (2015). F2RL3 methylation, lung cancer incidence and mortality. Int. J. Cancer *137*, 1739–1748.

Zhao, N., Wang, R., Zhou, L., Zhu, Y., Gong, J., and Zhuang, S.M. (2014). MicroRNA-26b suppresses the NF-κB signaling and enhances the chemosensitivity of hepatocellular carcinoma cells by targeting TAK1 and TAB3. Mol. Cancer *13*, 35.

Zhou, R., Gong, A.Y., Chen, D., Miller, R.E., Eischeid, A.N., and Chen, X.M. (2013). Histone Deacetylases and NF-kB Signaling Coordinate Expression of CX3CL1 in Epithelial Cells in Response to Microbial Challenge by Suppressing miR-424 and miR-503. PLoS One *8*, e65153.