European Journal of Human Genetics (2015) 23, 1100–1105

www.nature.com/ejhg

EJHG Open

ARTICLE

Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease

Leticia Plaza-Izurieta[1], Nora Fernandez-Jimenez[1], Iñaki Irastorza[2], Amaia Jauregi-Miguel[1], Irati Romero-Garmendia[1], Juan Carlos Vitoria[2] and Jose Ramon Bilbao*[,1]

Celiac disease is a chronic immune-mediated disorder with an important genetic component. To date, there are 57 independent association signals from 39 non-HLA loci, and a total of 66 candidate genes have been proposed. We aimed to scrutinize the functional implication of 45 of those genes by analyzing their expression in the disease tissue of celiac patients (at diagnosis/treatment) compared with non-celiac controls. Moreover, we investigated the SNP genotype effect in gene expression and performed coexpression analyses. Several genes showed differential expression among disease groups, most of them related to immune response. Multiple *trans*-eQTLs but only four *cis*-eQTLs were found, and surprisingly the genotype effect seems to be stimulus dependent as it differs among groups. Coexpression levels vary from higher to lower levels in active patients at diagnosis, treated patients and non-celiac controls respectively. A subset of 18 genes tightly correlated in both groups of patients but not in controls was identified. Interestingly, this subset of genes was influenced by the genotype of three SNPs. One of the SNPs, rs1018326 on chromosome two is on top of a known lincRNA whose function is not yet described, and whose expression seems to be upregulated in active disease when comparing biopsy pairs from the same individuals. Our results strongly suggest that the effects of disease-associated SNPs go far beyond the oversimplistic idea of transcriptional control at a nearby locus. Further investigations are needed to determine how each variant disrupts fine-tuning mechanisms in the genome that eventually lead to disease.

European Journal of Human Genetics (2015) **23,** 1100–1105; doi:10.1038/ejhg.2014.244; published online 12 November 2014

INTRODUCTION

Celiac disease (CD) is a common (prevalence 1:100) chronic immune-mediated enteropathy caused by intolerance to ingested gluten that develops in genetically predisposed individuals. The typical histological findings in active CD comprise villous atrophy, crypt hyperplasia and lymphocytic infiltration of the small intestinal mucosa, and the only effective treatment is strict lifelong gluten-free diet (GFD).[1] The major CD susceptibility locus maps to the MHC region on chromosome 6p21 and has been estimated to be responsible for 40% of the genetic contribution to CD; in fact, virtually all patients are HLA–DQ2- or HLA–DQ8-positive.[2] However, risk HLA variants are necessary but not sufficient for CD development, as those alleles are also common in general population, pointing to the contribution of other loci to the genetic predisposition to develop the disease.

To date, two genome-wide association studies (GWAS) have been performed in CD, revealing 26 regions of genetic susceptibility to the disease.[3–5] More recently, 13 additional susceptibility loci have been discovered with the Immunochip genotyping array, where immune-mediated disease loci containing markers that had achieved genome-wide significance ($P<5\times10^{-8}$) in 12 diseases (autoimmune thyroid disease, ankylosing spondylitis, Crohn's disease, CD, IgA deficiency, multiple sclerosis, primary biliary cirrhosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes and ulcerative colitis) were densely genotyped.[6] Many of the loci identified are also associated with other autoimmune or chronic immune-mediated diseases, with particular overlapping between CD, type 1 diabetes[7] and rheumatoid arthritis.[8]

Several genes within those regions have been proposed as etiological candidates, most of them previously related to the immune response or to T-cell maturation, and it has been suggested that they might participate in the different stages of the pathogenesis of CD. However, association studies are only able to pinpoint the location of susceptibility loci and the subsequent selection of candidate genes is often aprioristic and biased by the current paradigm of CD pathogenesis, with no robust experimental results to support any functional involvement of those candidate genes in the target tissue of CD patients. So far, the large-scale studies performed in CD have discovered a total of 57 independent CD association signals from 39 non-HLA loci.[6] Twenty-nine of those regions map to a single protein-coding gene, whereas the majority seem to localize to intergenic regions, suggesting more than one possible causal gene or some yet unidentified functional elements of the genome. Overall, 66 candidate genes have been proposed based on their localization under the association peaks, but there is a need to perform functional studies in the disease target tissue to prove the causative mechanism suggested for each association signal.

[1]Immunogenetics Research Laboratory, Department of Genetics, Physical Anthropology and Animal Physiology, BioCruces Research Institute, University of the Basque Country-UPV/EHU, Leioa, Spain; [2]Department of Pediatrics, Cruces University Hospital, University of the Basque Country-UPV/EHU, Barakaldo, Spain
*Correspondence: Dr JR Bilbao, Faculty of Medicine and Odontology, Immunogenetics Research Laboratory, Department of Genetics, Physical Anthropology and Animal Physiology, BioCruces Research Institute, University of the Basque Country-UPV/EHU, Bizkaiko Campusa, Leioa 48940, Spain. Tel: +34 946015289; Fax: +34 946013145; E-mail: joseramon.bilbao@ehu.es
Received 2 July 2014; revised 30 September 2014; accepted 4 October 2014; published online 12 November 2014

In a previous work, our group analyzed the expression of the 10 candidate genes proposed in the first GWAS in intestinal biopsies from patients and controls,[4] to determine the influence of associated SNP genotypes in their expression and their possible implication on CD development. We observed that several genes were differentially expressed depending on disease status, and found different functional relationships between the expression of candidate genes and SNP genotypes.[9]

In the present work, we wanted to question the implication of the additional proposed candidate genes in disease development. To investigate their putative role in the disease process we selected an additional set of 45 candidate genes with known function (Supplementary Table 1) and analyzed their expression in the disease tissue of celiac patients at diagnosis and after more than 2 years on GFD, and compared it with non-celiac controls. We also aimed to determine whether disease-associated variants have any influence on gene expression, considering the genotypes of the top-associated SNPs in the Immunochip project for each candidate gene. Moreover, we performed coexpression analyses in order to reveal possible common regulatory elements, which could be altered in celiac patients on account of inflammation or owing to predisposing genetic determinants.

MATERIALS AND METHODS

Patients and biopsies

CD was diagnosed according to the European Society of Pediatric Gastroenterology Hepatology and Nutrition criteria in force at the time of recruitment, including anti-gliadin, anti-endomysium and anti-transglutaminase antibody determinations as well as a confirmatory small bowel biopsy. The study was approved by the Institutional Boards (Cruces University Hospital code CEIC-E09/10 and Basque Clinical Trials and Ethics Committee code PI2013072) and analyses were performed after informed consent was obtained from all subjects or their parents. Biopsy specimens from the distal duodenum of each patient were obtained during routine diagnosis endoscopy.

The sample set consisted of 15 CD children at diagnosis (on a gluten-containing diet, with CD-associated antibodies, atrophy of intestinal villi and crypt hyperplasia), and the same patients in remission after being treated with GFD for >2 years (asymptomatic, antibody negative and normalized intestinal epithelium at that time), plus 15 tissue samples from non-celiac individuals not suffering from inflammation at the time of endoscopy used as controls. Total RNA was extracted from small bowel biopsies using the NucleoSpin microRNA kit (Macherey-Nagel, Düren, Germany) following manufacturer's instructions.

RNA samples and gene expression

RNA was normalized to 8 ng/μl and converted to cDNA using the AffinityScript cDNA Synthesis kit (Agilent Technologies, Santa Clara, CA, USA) following manufacturer's protocol. Gene expression analyses were performed using Fluidigm Biomark 48.48 dynamic arrays (Fluidigm Corp., South San Francisco, CA, USA) and commercially available TaqMan Gene Expression assays. Housekeeping gene *RPLPO* was simultaneously quantified and used as an endogenous control of input RNA (Life Technologies, Thermo Fisher Scientific Inc., Waltham, MA, USA). Relative expression in each sample was calculated using the accurate Ct method[10] and normalized to the average expression value of the 15 control samples as previously described. Gene expression results are publicly available at the Gene Expression Omnibus data repository (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE61849.

Differences in gene expression levels were analyzed with nonparametric Wilcoxon matched pairs rank test (diagnosis *vs* treated) and Mann–Whitney *U*-test (non-celiac *vs* both disease groups). Coexpression was calculated using Pearson correlation. All statistic calculations were performed in GraphPad Prism 5 (GraphPad Software, La Jolla, CA, USA). Extreme outliers exceeding >3 SD from the mean of each group were considered methodological errors and were removed from statistical comparisons.

SNP genotyping

Genotyping of 44 top-associated SNPs from the Immunochip project was performed with a Fluidigm Biomark dynamic array (48.48) and SNPtype assays (Fluidigm Corp.) in 26 samples with expression results in which DNA was available. Eight samples were already genotyped in the Immunochip sample set and were used as quality control for the new genotyping. Three samples had to be removed from the study due to failed genotyping, resulting in a total number of 23 samples, 14 controls and 9 celiac patients. The assay design was performed by the Fluidigm Assay Design Group. Seven of the target SNPs did not fulfill the established assay design requirements due to adjacent SNPs within 20–30 bases on each side of the target SNP, GC content >65% or triallelic SNPs. After an in-deep analysis of those seven SNPs, taking into account the allelic frequencies of the target SNP and the adjacent SNPs and the frequency of each allele in the case of the unique triallelic SNP (rs61907765) in Ensembl, we decided to omit this obstacle in the design of six SNPs and to remove the SNP rs60215663 from the analysis due to smaller minor-allele frequency than adjacent SNPs. Complete genotyping results are available as Supplementary Material.

Coexpression analysis

Merlin 1.1.2 software was used to test association between SNP genotype and candidate gene expression.[11] The association was tested independently in each of the studied groups in order to avoid false associations due to duplicated genotypes in CD sample pairs.

RESULTS

Differentially expressed genes in CD

Fifteen out of the forty-five genes analyzed were differentially expressed when comparing the fold change between active disease samples and non-celiac controls. Nine of the genes were significantly overexpressed in active CD (*CTLA4, ICOS, CIITA, FASLG, PLEK, PVT1, CD28, UBASH3A* and *SOCS1*), whereas the other six genes (*ATXN2, ICOSLG, ARHGAP31, ZFP36L1, CCR2* and *TREH*) were downregulated (Figure 1a). As could be expected due to the aprioristic selection of the candidate genes, GO-term analysis of the altered genes showed enrichment of immune response related processes such as regulation of T cells, lymphocyte and leukocyte activation and proliferation, lymphocyte costimulation and so on. The most relevant genes behind this enrichment are *ICOSLG* (inducible T-cell co-stimulator ligand); *CCR2* (chemokine (C–C motif) receptor 2), a receptor for a chemokine which specifically mediates monocyte chemotaxis and is involved in monocyte infiltration in inflammatory diseases; *PLEK* (pleckstrin); *CTLA4* (cytotoxic T-lymphocyte-associated protein 4), a member of the immunoglobulin superfamily that encodes a protein which transmits an inhibitory signal to T cells; *CD28*, an essential protein for T-cell proliferation and survival, cytokine production and T-helper type-2 development and *ICOS* (inducible T-cell co-stimulator), which also belongs to the *CD28* and *CTLA4* cell-surface receptor family and has an important role in cell–cell signaling, immune response and regulation of cell proliferation. *ICOS*, *CD28* and *CTLA4* are located on the CELIAC3 locus, a well-known region that has been linked to several autoimmune disorders, including CD, originally identified by Holopainen *et al*[12] and that has been replicated several times in posterior studies. When treated patients and non-celiac controls were compared, only three genes showed significant expression differences (*ATXN2*, *CCR2* and *CCR4*), being constitutively downregulated in the disease group (Figure 1b).

The comparison between active and treated disease mucosa-identified differential expression in ten genes, nine of which were upregulated in the active disease (*CIITA, POU2AF1, IRF4, SOCS1, PVT1, ICOS, CTLA4, YDJC* and *PLEK*) and one was downregulated (*TREH*) (Figure 1c). As in the case of active disease *vs* controls, the

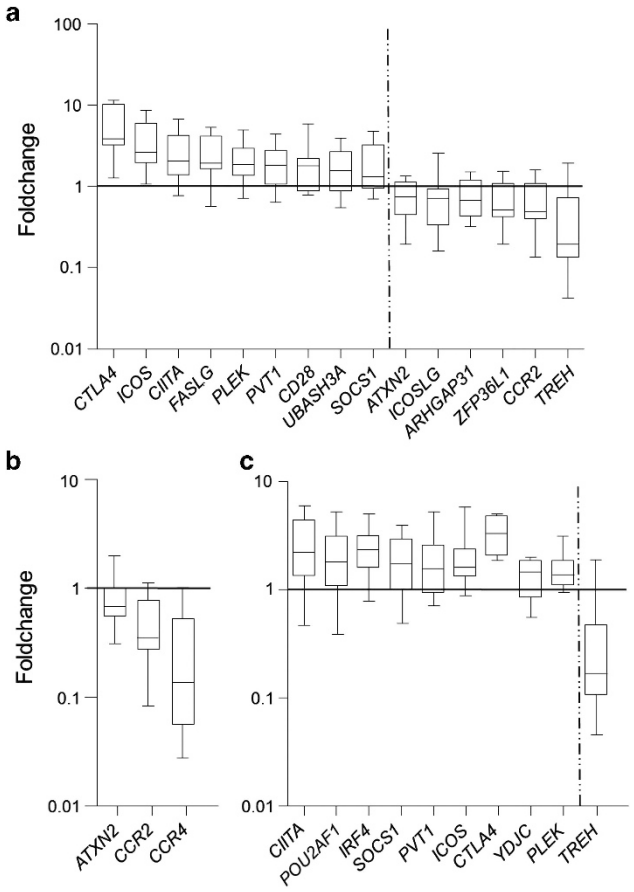

Figure 1 Expression fold change of differentially expressed genes. (**a**) Active CD *vs* controls, (**b**) treated CD *vs* controls and (**c**) active *vs* treated CD.

enriched GO terms are related to the regulation of immune cell activation, due to the altered expression of *CTLA4*, *ICOS* and *PLEK* as previously, plus *IRF4* (interferon regulatory factor 4), an important transcription factor in the regulation of interferon in response to infection by viruses, which is lymphocyte specific and negatively regulates TLR signaling, a pathway that is central to the activation of innate immune system. Apart from that, GO terms related to interferon–gamma response are also enriched in this case, due to three genes that are upregulated in the active disease attributable to the inflammatory process, *CIITA* (class II MHC transactivator), *IRF4* and *SOCS1* (suppressor of cytokine signaling 1).

Genotype effect in gene expression

Despite the limited number of biological samples in our study, we also searched for relationships between SNP genotypes and gene expression levels. We were able to include 14 individuals from the control group and 9 sample pairs from the disease group, for whom both genotypes and expression results were available. For this reason, it was often impossible to have all three genotypes present in every group; heterozygous and minor-allele homozygous samples were combined in order to increase statistical power.

We detected genotype effects of a number of SNPs on the expression of several genes, but surprisingly, the effect seemed to be stimulus dependent, as it was different among the groups. Moreover, most eQTLs were in *trans* and only four candidate genes located under the association peak were influenced by its putative regulatory SNP; rs1980422-*ICOS* in debuts, rs79758729-*ELMO1* in treated patients, rs12068671-*TNSF18* and rs13397-*TMEM187* in controls (Figure 2). In an attempt to explain this result, we scrutinized the genomic region around each associated SNP in search for putative regulatory elements that could be altering the expression of genes *in trans*. We conducted searches in different databases available online, such as Haploreg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php),[13] Ensembl (http://www.ensembl.org)[14] and the UCSC Genome browser (http://genome.ucsc.edu).[15] As expected, elements affected by the potentially regulatory SNPs included open chromatin regions, novel protein-coding sequences, processed antisense transcripts, pseudogenes, microRNAs, novel lincRNAs and altered protein-binding motifs. This finding opens the door for further studies in order to determine whether any of those sequences could have a real functional role in gene regulation and development of CD.

Coexpressed gene patterns in CD

Coexpression analyses were performed to identify possible common regulation signatures that could be altered in celiac patients on account of inflammation or owing to genetic determinants. Interestingly, we observed different correlation patterns among genes in the three study groups, from higher to lower coexpression levels in gluten-consuming celiac patients at diagnosis, treated patients and non-celiac controls, respectively (Supplementary Figure 1). The selection of those genes that were coexpressed in both groups of patients, but not in non-celiac controls, identified a subset of 18 genes that were tightly correlated in patients that seemed to be putatively under the control of three SNPs (Figure 3).

One of those SNPs, rs1018326, is located on chromosome two, in an intergenic region between *UBE2E3* and *ITGA4*, on top of a known lincRNA (*AC104820.2*) whose function has not been described yet. This RNA gene has five transcripts (spliced variants), ranging from 342 to 1771 base pairs length. The expression of *AC104820.2* was significantly altered between biopsy pairs from the same patients in different stages of the disease, being upregulated in active biopsies (Figure 4). We did not observe these differences when comparing unpaired biopsies from independent active and treated CD patients, stressing the enormous variability among CD patients and the need for strict sample pairing for efficient comparisons (data not shown).

DISCUSSION

Candidate gene selection following large-scale SNP association studies is often aprioristic and greatly influenced by the current knowledge of the pathogenic mechanisms that are thought to be involved in the disease, but functional studies are the only unbiased approach to identify real functional players. Until now, only a small number of studies have performed deep analyses of associated regions prior to proposing candidate-susceptibility genes: a genetic and functional analysis of *THEMIS* and *PTPRK*, the two candidate genes located on the CD association peak chr6: 127.99–128.38 Mb found a significant correlation between the expression levels of both genes in CD patients that was absent in the control group.[16] Although this finding could suggest a possible role for both of the genes, it shows the existence of a common regulatory relationship that could reside in the noncoding albeit functional intergenic region. Using a different approach, fine mapping of the *LPP* locus to identify possible functional variants revealed six SNPs that overlap regulatory sites, with rs4686484 having a possible effect on *LPP* gene expression in CD patients.[17] Finally, Östensson M *et al*[18] recently performed pathway analyses and two-locus interaction studies to further investigate association signals. They found some differentially expressed genes in the small intestine mucosa from CD patients, and identified susceptibility genes from

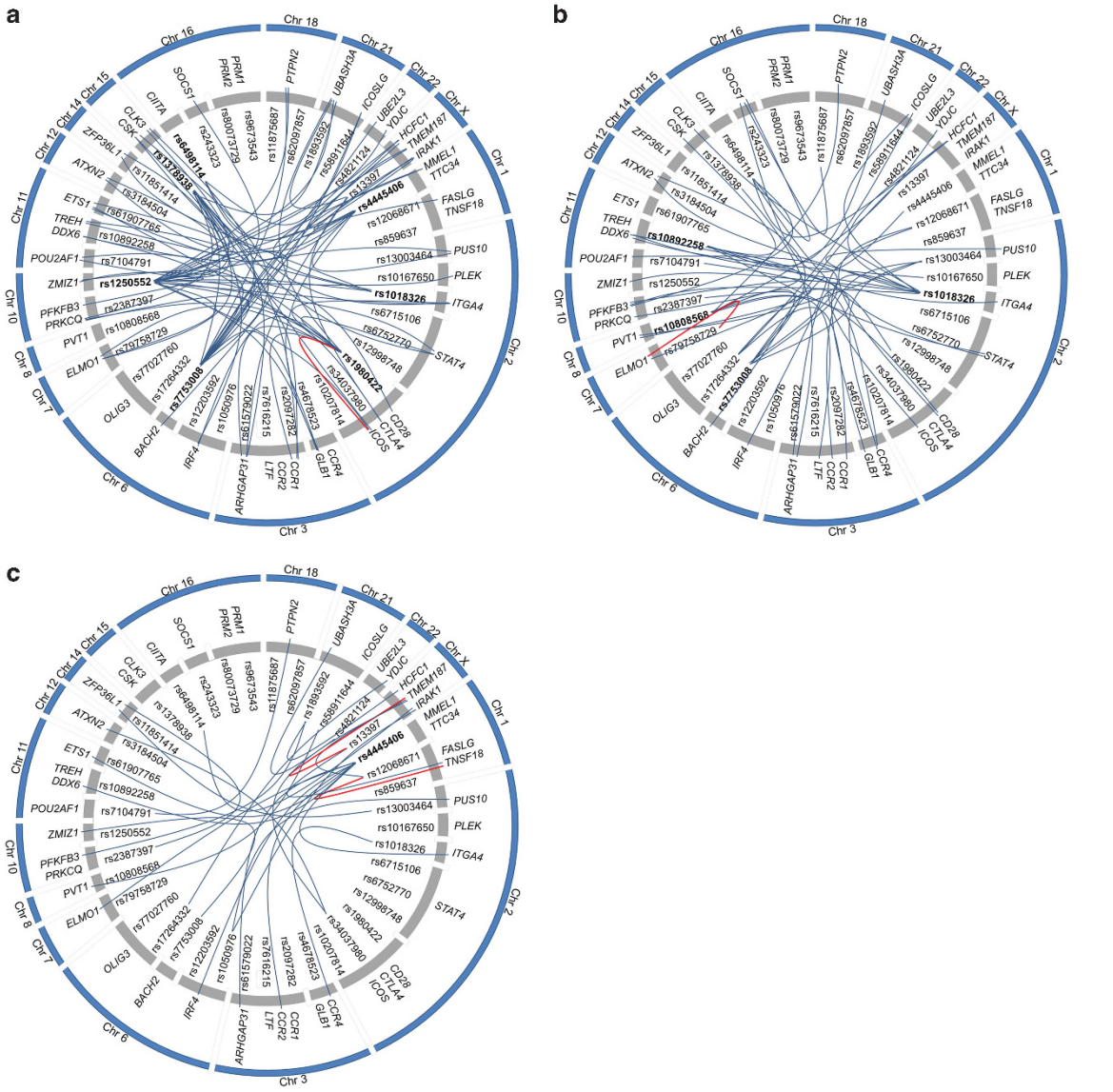

Figure 2 SNP genotype effect on candidate gene expression for the different disease statuses. (**a**) At diagnosis, (**b**) >2 years GFD and (**c**) controls. Statistical analyses performed with Merlin 1.1.2 software; *P*-value <0.005 was fixed as significant SNP effect. SNPs with an effect on multiple genes are shown in bold. Blue lines indicate *trans*-eQTLs and red lines *cis*-eQTLs.

top-scoring regions that could be gathered into several categories. They suggested that those genes and pathways together could reveal a new potential biological mechanism that could influence the genesis of CD and other chronic inflammatory disorders.

Although the effects of associated SNPs on gene expression has been previously studied in CD, and several *cis*- and *trans*- eQTLs have been found, expression data have always been obtained from peripheral blood samples.[18] A recently published work analyzed the effect of regulatory variants upon monocyte activation and concludes that a significant proportion of variants may show activity only in a context-specific manner, proposing that only considering the genetic, cellular and environmental context relevant to the disease will make it possible to resolve functional genetic variants more extensively.[19] This is the case of the results obtained in the present study, where we are able to distinguish different eQTLs determined by prolonged gliadin insult and inflammation that are present in CD patients at diagnosis. Furthermore into the complexity of the mechanisms involved in the functional translation of associated genetic variants, we observe that SNP rs1018326 seems to exert its effect through a lincRNA, *AC104820.2,* for whom disease-related expression changes are only evident when biopsies from the same patients taken at different stages of the disease are compared.

Concerning the coexpression patterns identified, the higher correlation observed in patients could suggest a coordinated alteration that again points towards complex regulatory mechanisms. In the case of active disease, a higher degree of correlation could be explained as a consequence of the inflammatory milieu provoked by the ingestion of gliadin, taking into account that all the candidate genes proposed are related to the immune response. However, this correlation is maintained in the group of treated patients, even after gliadin withdrawal from the diet for >2 years, suggesting at least two possible explanations: an intrinsic, constitutive alteration of those genes in celiac individuals that is independent from gliadin ingestion or a response pattern caused by the gliadin insult, that is not reversible after

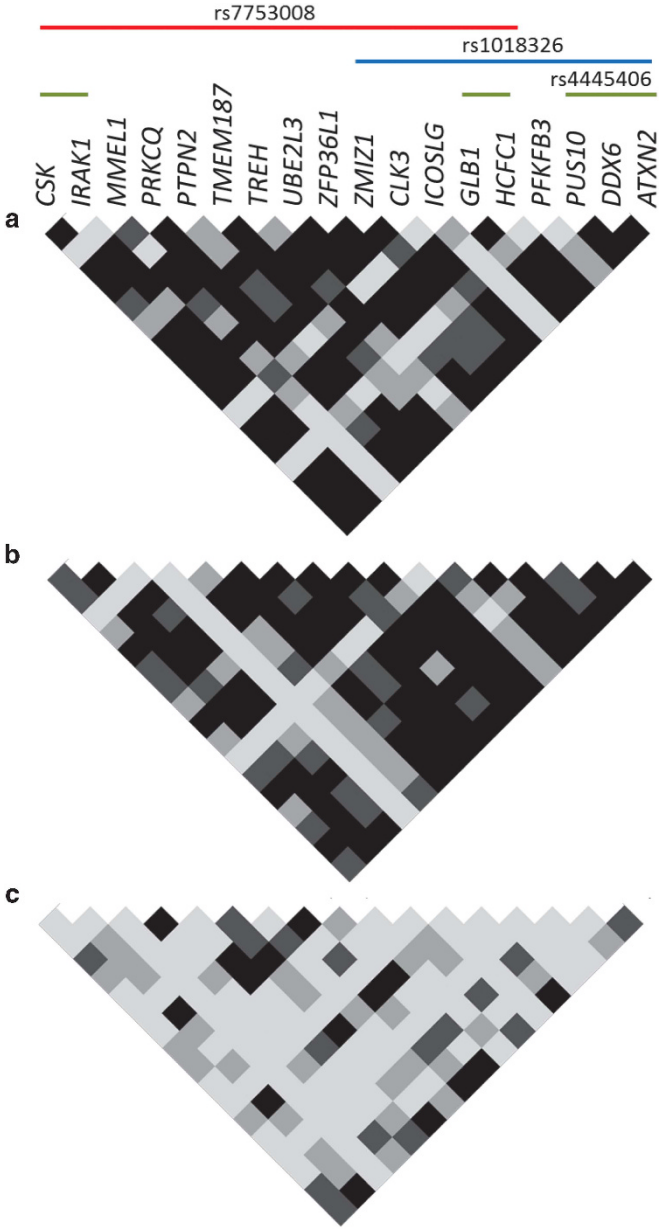

Figure 3 Gene pair coexpression matrixes for the different disease statuses on a subset of genes correlated in patients but not in controls. (**a**) At diagnosis, (**b**) >2 years GFD and (**c**) controls. Each small square represents the *P*-value for the correlation of the expression level in a specific gene pair. Black, dark gray, light gray and white indicate Pearson's correlation *P*-value of $P<0.0001$, $P<0.001$, $P<0.01$ and $P>0.01$, respectively. SNPs with *trans*-eQTLs for those genes are shown.

2 years on GFD and probably requires a longer time to reach basal expression levels.

An opposite coexpression scenario has recently been described by our group in the case of the NFκB pathway in CD.[20] In that case, the strongest correlation was found in the control group, suggesting a very tight regulatory control of the pathway in a healthy gut, and an alteration of this pathway in the disease. These opposed results make sense if we take into account that NFκB coexpression is indeed expected to be the normal situation because genes that are part of the same pathway are expected to be under the same regulatory mechanisms. In the case of the disease-associated loci, even though enriched

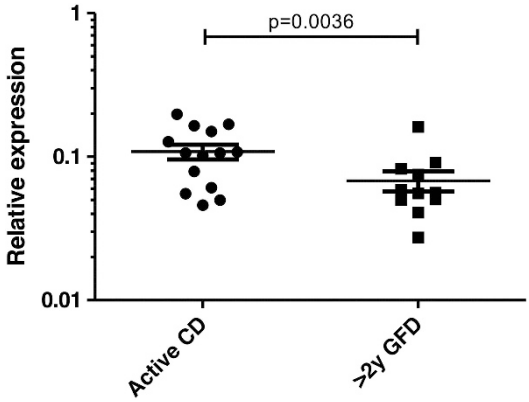

Figure 4 Relative expression of *AC104820.2* lincRNA in a set of 11 biopsy pairs. Paired *t*-test was applied for statistical analysis.

in immune-related genes, they would not be expected to react in a coordinated manner upon an environmental challenge unless they are related to the same regulatory variation. In this work we are analyzing the expression of many candidate genes that have in common the implication in the immune response, which is altered in CD, so the coordinated alteration of those genes could be understood.

The idea put forward in the present study needs robust experimental confirmation to be proven, and there are still many pieces to be put together in the puzzle of the common disease genetic susceptibility. However, it is clear that the effects of associated variants go far beyond the oversimplistic idea of transcriptional control at a nearby locus. The complex interactions that maintain a coordinated, healthy response to an environmental challenge are written on our genome, and the disruption of those subtle fine-tuning mechanisms emerge as the initial cause of a series of events that eventually lead to disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was partially funded by Research Project grants from the Spanish Ministry of Science and Innovation (10/0310) and Basque Department of Health (2011/111034) JRB. LP-I, NF-J and A J-M are predoctoral fellows supported by grants from the Basque Department of Education.

1 Abadie V, Sollid LM, Barreiro LB, Jabri B: Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annu Rev Immunol* 2011; **29**: 493–525.
2 Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E: Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* 1989; **169**: 345–350.
3 van Heel DA, Franke L, Hunt KA *et al*: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007; **39**: 827–829.
4 Hunt KA, Zhernakova A, Turner G *et al*: Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008; **40**: 395–402.
5 Dubois PC, Trynka G, Franke L *et al*: Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010; **42**: 295–302.
6 Trynka G, Hunt KA, Bockett NA *et al*: Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; **43**: 1193–1201.
7 Smyth DJ, Plagnol V, Walker NM *et al*: Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 2008; **359**: 2767–2777.
8 Zhernakova A, Stahl EA, Trynka G *et al*: Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet* 2011; **7**: e1002004.

9 Plaza-Izurieta L, Castellanos-Rubio A, Irastorza I, Fernandez-Jimenez N, Gutierrez G, Bilbao JR: Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes. *J Med Genet* 2011; **48**: 493–496.
10 Martin-Pagola A, Perez-Nanclares G, Ortiz L *et al*: MICA response to gliadin in intestinal mucosa from celiac patients. *Immunogenetics* 2004; **56**: 549–554.
11 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
12 Holopainen P, Naluai AT, Moodie S *et al*: Candidate gene region 2q33 in European families with coeliac disease. *Tissue Antigens* 2004; **63**: 212–222.
13 Ward LD, Kellis M: HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40**: D930–D934.
14 Flicek P, Ahmed I, Amode MR *et al*: Ensembl 2013. *Nucleic Acids Res* 2013; **41**: D48–D55.
15 Kent WJ, Sugnet CW, Furey TS *et al*: The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.
16 Bondar C, Plaza-Izurieta L, Fernandez-Jimenez N *et al*: THEMIS and PTPRK in celiac intestinal mucosa: coexpression in disease and after in vitro gliadin challenge. *Eur J Hum Genet* 2013; **22**: 358–362.
17 Almeida R, Ricaño-Ponce I, Kumar V *et al*: Fine mapping of the celiac disease-associated LPP locus reveals a potential functional variant. *Hum Mol Genet* **23**: 2481–24892013.
18 Östensson M, Montén C, Bacelis J *et al*: A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease. *PLoS One* 2013; **8**: e70174.
19 Fairfax BP, Humburg P, Makino S *et al*: Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 2014; **343**: 1246949.
20 Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L *et al*: Coregulation and modulation of NFκB-related genes in celiac disease: uncovered aspects of gut mucosal inflammation. *Hum Mol Genet* 2013; **23**: 1298–1310.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)