A celiac diasease associated lncRNA named *HCG14* regulates *NOD1* expression in intestinal cells.

Izortze **Santin**[1*], Amaia **Jauregi-Miguel**[2], Teresa **Velayos**[1], Ainara **Castellanos-Rubio**[2], Koldo **Garcia-Etxebarria**[3], Irati **Romero-Garmendia**[2], Nora **Fernandez-Jimenez**[2], Iñaki **Irastorza**[4], Luis **Castaño**[1] and Jose Ramón **Bilbao**[2].

[1] Endocrinology and Diabetes Research Group, BioCruces Health Research Institute, UPV-EHU, CIBERDEM, Barakaldo, Basque Country, Spain.

[2] Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV-EHU), BioCruces Health Research Institute, Leioa, Spain.

[3]Department of Gastrointestinal and Liver Diseases, Biodonostia Health Research Institute, San Sebastian, Spain.

[4] Department of Pediatrics, University of the Basque Country (UPV-EHU), Cruces University Hospital, Barakaldo, Spain.

Word count: 2990 words, 1 table and 3 figures

Conflicts of interest: None declared.

***To whom correspondence should be addressed:**

Dr. Izortze Santin

Endocrinology and Diabetes Research Group,

BioCruces Health Research Institute, UPV-EHU

E48903-Barakaldo

Bizkaia, Basque Country, Spain

Tel: +34946006000, ext.6376

Fax: +34946006076

Email address: izortze.santingomez@osakidetza.eus

Author's contribution

I.S., A.C.R. and J.R.B. contributed to the general idea, design of experiments, research data, contributed to discussion and wrote, revised and edited the manuscript.

A.J.M., T.V., K.G.E., I.R.G., N. F.J., I.I. and L.C. contributed to research data, and revised and edited the manuscript.

ABSTRACT

Objective: To identify additional celiac disease associated *loci* in the Major Histocompatibility Complex independent from classical HLA risk alleles (HLA-DR3-DQ2) and to characterize their potential functional impact in celiac disease pathogenesis at the intestinal level.

Methods: We performed a high resolution SNP genotyping of the MHC region, comparing HLA-DR3 homozygous celiac patients and non-celiac controls carrying a single copy of the B8-DR3-DQ2 conserved extended haplotype . Expression level of potential novel risk genes was determined by RT-PCR in intestinal biopsies and in intestinal and immune cells isolated from control and celiac individuals. Small interfering RNA-driven silencing of selected genes was performed in the intestinal cell line T84.

Results: MHC genotyping revealed two associated SNPs, one located in *TRIM27* gene and another in the non-coding gene *HCG14*. After stratification analysis, only *HCG14* showed significant association independent from *HLA-DR-DQ loci* Expression of *HCG14* was slightly downregulated in epithelial cells isolated from duodenal biopsies of celiac patients, and eQTL analysis revealed that polymorphisms in *HCG14* region were associated with decreased *NOD1* expression in duodenal intestinal cells.

Conclusions: We have sucessfully employed a conserved extended haplotype-matching strategy and identified a novel additional celiac disease risk variant in the lncRNA *HGC14.* This lncRNA seems to regulate the expression of *NOD1* in an allele-specific manner. Further functional studies are needed to clarify the role of *HCG14* in the regulation of gene expression and to determine the molecular mechanisms by which the risk variant in *HCG14* contributes to celiac disease pathogenesis.

KEYWORDS

Genetic association study, susceptibility genes, Major Histocompatibility Complex, intestinal biopsies, functional characterization, celiac disease.

WHAT IS KNOWN

- The major susceptibility locus for celiac disease is located within the Major Histocompatibility Complex (HLA-DQ2) and evidence suggest the presence of additional CeD *loci* within or close to the MHC.
- Long non-coding (lnc)RNAs are emerging as important players in the regulation of disease-associated pathways.

WHAT IS NEW

- The lncRNA *HCG14* located at the MHC harbors CeD-associated variants that confer susceptibility independently of the HLA-DQ genotype.
- *HCG14* expression is slightly downregulated in epithelial cells from duodenal biopsies of patients with CeD.
- Expression of *NOD1*, a candidate gene for several inflammatory diseases, is modulated by *HCG14* in intestinal cells.

INTRODUCTION

Celiac disease (CeD) is a chronic, autoimmune disorder caused by intolerance to ingested gluten that develops in genetically susceptible individuals (1). It usually presents as a histologic lesion of the jejunum, consisting in atrophy of the intestinal villi and crypt hyperplasia, with a very wide range of clinical and functional expression (2). Familial clustering of CeD has been known for a long time, with concordance rates between identical twins close to 70%, and around 10% of first-degree relatives of CeD patients affected (3). The major susceptibility locus is located in the MHC region on chromosome 6p21 and more than 90% of patients with CeD express the *HLA-DQ2* heterodimer, encoded by the risk haplotype *DQA1*0501/*02* in *cis* or in *trans* in HLA-DR5/DR7 heterozygous patients. The minority of patients lacking HLA-DQ2 presents the HLA-DQ8 molecule, encoded by haplotype *DQA1*0301/DQB1*0302* (4). Allele HLA-DQ2 is common in the general population, being present in approximately 30% of Caucasians (5).

Overall, DQ2 seems necessary but yet not sufficient for the development of CeD, and this idea has encouraged the search for non-HLA and other HLA genetic variants predisposing to disease. Conserved extended haplotypes (CEH) are long genomic extensions that have been maintained unaltered during human evolution and are a characteristic of the MHC genomic region (6). The HLA-B8-DR3-DQ2 CEH or AH8.1 is one of the most common extended DR3 haplotypes in almost all Caucasian populations and consists of the HLA-A1, HLA-Cw7, HLA-B8, MICA-5.1, HLA-DR3, and HLA-DQ2 alleles. HLA-B8-DR3-DQ2 CEH is highly conserved throughout more than 4 Mb, and this conservation spans at least from *HLA-A* to*HLA-DQ locus* (7,8). Interestingly, this CEH has been associated with several autoimmune diseases,

including CeD and type 1 diabetes (T1D) (9,10). Recently, five independent association signals have been detected in the extended MHC: amino acid position 9 in HLA-DPB1, allele HLA-B*08:01 and HLA-B*39:06 and two SNPs, rs1611710 (near HLA-F) and rs2301226 (close to HLA-DPB1) (11). We have previously shown that, at least in in the Basque population, HLA-B8-DR3-DQ2 CEH discriminates HLA-DR3-DQ2 homozygous CeD patients from DR3-DQ2-homozygous T1D patients (characterized by the presence of another CEH, B18-DR3-DQ2, or AH18.1), suggesting the presence of additional CeD predisposing *loci* in the B8-DR3-DQ2 CEH (10). Using a CEH-matching strategy we detected novel association signals for type 1 diabetes in the MHC which had an independent effect from classical risk alleles in the HLA-DRB1 gene (12).

Under this hypothesis and to identify those additional CeD loci, we performed high resolution SNP genotyping of the MHC region, comparing HLA-DR3 homozygous CeD patients and non-celiac controls carrying a single copy of the B8-DR3-DQ2 CEH, in order to investigate putatively celiac-predisposing variation at the other HLA-DR3-DQ2 chromosome. Once again, the CEH-matching strategy that controls for the major effects of HLA loci was able to detec an additional risk variant with a modest contribution to CeD risk, located in the long non-coding RNA (lncRNA) gene HCG14.

We have shown that HCG14 is preferentially located in the nucleus of intestinal epithelial cells, suggesting a potential gene expression regulatory function. In fact, the CeD risk genotype of SNPs located in the HCG14 region correlates with decreased NOD1 expression in intestinal epithelial cells. NOD1 is an innate immune receptor that upon recognition of exogenous and endogenous ligands induce the activation of several pro-inflammatory pathways. Interestingly, accumulating evidence points to the implication of the innate immune system in the initiation of CeD and it has been shown

that gluten peptides act as ligands for several innate receptors, inducing innate immune activation (13,14)

In addition, our results show that HCG14 is preferentially located in the nuclei of intestinal cells, suggesting that in this cell type, *HCG14* may have a regulatory role in gene expression. In this sense, a CeD-associated lncRNA that is preferentially expressed in the nuclei of intestinal cells has been shown to modulate the expression of several pro-inflammatory genes (15). Although further studies are needed to characterize the functional implication of *HCG14* in CeD pathogenesis, our results suggest that in intestinal epithelial cells, *HCG14* is implicated in transcriptional regulation.

METHODS

Human Biological samples and cell separation

All studies involving human samples were approved by the Ethics Committee of Cruces University Hospital and samples were collected after informed consent had been obtained from participants or their parents. CeD patients were diagnosed according to the European Society for Paediatric Gastroenterology and Nutrition criteria in force at the time of recruitment, including the determination of anti-gliadin, anti-endomysium, and anti-tissue transglutaminase antibodies as well as a confirmatory small bowel biopsy.

The high resolution SNP scan was performed in a discovery sample set comprising 10 CeD patients and 39 non-celiac blood donors of Caucasian origin who were homozygous for HLA-DR3-DQ2 and carried a single copy of the B8-DR3-DQ2 CEH.

A replication study of the most significantly associated SNPs was performed in an independent sample of 525 CeD patients and 563 healthy controls, without *HLA* matching.

Biopsy specimens from the distal duodenum were obtained during routine diagnostic endoscopy, and were either snap frozen and stored in liquid nitrogen (whole biopsies) until they were used for gene expression studies or immediately processed for the isolation of epithelial ($CD326^{+}$) cells. Expression analyses were performed in a sample set of 54 duodenal mucosa samples from 2 categories: biopsy specimens from 28 patients diagnosed with active CeD who were on a non-restricted diet at that time, and from 26 patients without CeD who had undergone endoscopy but had no inflammation of the gut. Epithelial cells were separated using MACS magnetic cell separation technology (Miltenyi Biotec, Bergisch Gladbach, Germany) following the manufacturer's protocol (See Supplemental Methods for details, Supplemental Digital Content 1).

MHC genotyping and high risk HLA-DR-DQ genotyping

To analyze 2,360 SPNs across a region of 4.91bp in the MHC, a commercially available MHC Panel Set (Illumina Inc., San Diego, CA) was genotyped in the discovery samples using the Golden Gate protocol (Illumina Inc.). After filtering for quality control parameters (see Supplemental Methods, Supplemental Digital Content 1), Fisher's exact test was used for single marker allelic association analyses, and p-values below 10^{-4} were regarded significant.

In order to determine whether the risk attributed to the newly identified variants was independent from HLA class II, SNPs rs2187668 and rs7454108, proxies for *HLA-DQA1*0501-DQB1*0201* (DQ2) and *HLA-DQA1*0301-DQB1*0302* (DQ8),

respectively, were genotyped in all samples (16). Linkage disequilibrium between each of these SNPs and the novel risk markers was measured using Haploview (17). Independent contribution was also assesed by stratified analysis.

Cell culture and gliadin exposure

C2BBe1 and T84 intestinal cells were maintained in DMEM supplemented with 10% heat inactivated FBS, 1% non-essential amino acids, and 1% penicillin-streptomycin in tissue culture flasks.

For *in vitro* stimulation of cultured cells, pepsin-trypsin digest of gliadin (PT-G) was prepared as described previously (18). In addition, an enzymatic digest of BSA (PT-BSA) (ThermoFisher Scientific) prepared in the same way was used as a negative control of stimulation. C2BBe1 and T84 cells cells were exposed to 1 mg/ml of PT-G or PT-BSA for 4h.

Cellular fractionation

To determine the distribution of *HCG14* in the nuclear and cytoplasmic cell compartments, nuclei were isolated using C1 lysis buffer as previously described (15). The amounts of *HCG14*, *Lnc13* (nuclear control) and *RPLPO* (cytoplasmic control) were measured by RT-PCR and compared to the total amount of those RNAs in the whole cell.

Expression analysis

Total RNA was extracted from whole duodenal biopsies or from immune and epithelial cell fractions isolated from biopsies using RNeasy Micro Kit (Qiagen) and cDNA was synthesized with the iScript cDNA Synthesis Kit (Biorad). Expression of *HCG14*, *TRIM27*, *NOD1* and *Lnc13* was determined by real-time (RT)-PCR using specific

Taqman Gene Expression Assays (ThermoFisher Scientific) in an Eco Real Time PCR machine (Illumina). Expression values were corrected for the housekeeping gene *RPLPO*.

HCG14 expression was also measured in a commercially available RNA panel set of different human tissues (Human total RNA master panel II, Clontech, France).

RESULTS

A high resolution SNP genotyping of the MHC in CEH-matched individuals identifies novel independent CeD association signals

In the present study, we have taken advantage of the extraordinary conservation of the B8-DR3-DQ2 CEH and performed a high resolution SNP genotyping within the MHC, in order to search for additional CeD predisposing variants independent from *HLA-DR-DQ loci*. As is shown in Figure 1, the allelic association analysis revealed two association peaks in the most telomeric side of the xMHC (extended MHC): rs2523765 (dbSNP build 150; chr6: 29,849,329) in the most telomeric part of the classical MHC, downstream *HLA-G locus*, and rs3130838 (chr6: 28,898,751) located in an intergenic region between the coding gene *TRIM27* and the lncRNA gene *HCG14*. Genotyping results of both SNPs were confirmed using an alternative genotyping technique in an independent larger sample without *HLA*-matching. Although there was a possiblility that the associations initially identified could actually be false positives due to the limited sample size, the associations were replicated in the larger sample set [rs2523765: MAF cases/controls = 0.44/0.36; $p=2x10^{-4}$; OR=1.39 (95% CI=1.17-1.66),

and rs3135316: MAF cases/controls=0.13/0.04; $p=6x10^{-10}$; OR=3.21 (95%CI=2.16-4.77)].

Both associated SNPs are located in the telomeric side of the MHC, far enough from the classical *HLA* risk *loci* (*DR* and *DQ*). However, we wanted to check whether the associated variants were in linkage disequilibrium (LD) with the classical *HLA* risk alleles (*HLA-DR3-DQ2* and *HLA-DR4-DQ8*), so samples were genotyped for two tagSNPs that are proxies for these two *HLA* risk haplotypes. Results from the LD analysis revealed that neither rs2523765 nor rs3135316 were in LD with HLA-DR3-DQ2 or HLA-DR4-DQ8, and correlation values (r^2) between the associated variants and *HLA-DR3-DQ2*/*HLA-DR4-DQ8* was close to zero.

In addition, stratified analysis was performed in order to test whether the effect of the associated variants to disease risk was independent from *HLA-DR-DQ loci* effect (Table 1). Genotyping results for rs2523765 were stratified according to the risk HLA-DR3-DQ2 haplotype, the association analysis did not reach statistical significance. Failure of the stratification analysis was probably due to an insufficient sample-size to detect such a small contribution of the associated variant. In contrast, the effect of rs3135316 in disease risk was more pronounced, so that stratified analysis of rs3135316 (tagSNP for rs3130838) suggested that the risk conferred by this variant was indeed independent from classical HLA risk alleles.

Linkage disequilibrium analysis of rs3130838 revealed that it is in absolute LD (r2=1) with two SNPs that are located in intron 4 of *TRIM27* gene (rs3135293) and in the intronic region of the lncRNA gene *HCG14* (rs3135316) (Figure 2A).

Expression of HCG14 is slightly dowregulated in epithelial cells isolated from duodenal biopsies of CeD patients.

We analyzed the expression of *TRIM27* and *HCG14* in duodenal biopsies of active CeD patients and non-CeD controls. Both *TRIM27* and *HCG14* were expressed in whole duodenal biopsies, although we did not detect significant differences between groups (Supplemental Digital Content 2). As expected, the expression of *TRIM27* was higher than *HCG14*, confirming that in general, expression of non-coding RNAs is lower than expression of coding genes (19).

The expression levels of *HCG14* were also measured in immune and epithelial cellular fractions from duodenal biopsies of celiac and non-celiac control individuals. As shown in Figures 2B and 2C, the expression of *HCG14* was higher in $CD45^+$ immune cells than in $CD326^+$ epithelial cells. While in immune cells expression levels of *HCG14* were similar in celiac and control groups (Figure 2B); in epithelial cells, *HCG14* was slightly dowregulated in the celiac group, although differences did not reach statistical significance (Figure 2C). The expression levels of *TRIM27* did not differ between celiac and non-celiac controls in none of the fractions (data not shown).

HCG14 is ubiquotiusly expressed in human tissues and is preferentially located in the nucleus of intestinal cells

To characterize the role of *HCG14*, we analyzed its expression in a set of human tissues and in the intestinal cell line T84 (Figure 2D). We found that *HCG14* is ubiquotiusly expressed in all the human tissues analyzed, with the highest levels in brain, liver, lung and kidney. The expression level of *HCG14* is higher in tissues related to the immune system (e.g. spleen) than in colon and intestine. These results confirmed our previous

expression results in which we observed a higher expression of *HCG14* in $CD45^+$ immune cells than in intestinal $CD326^+$ cells (Figures 2B and C).

We analyzed *HCG14* expression in nuclear fractions and whole cell lysates of T84 clonal intestinal cells. As shown in Figure 2E, *HCG14* was preferentially expressed in the nucleus of intestinal cells as is the case of the previously described nuclear, CeD-associated, regulatory *Lnc13* (15). These results show that *HCG14* is preferentially expressed in the nucleus of intestinal cells, suggesting a potential regulatory role in the trancription of other genes.

In order to assess whether *HCG14* expression is modulated by gliadin in intestinal cells, we next determined its expression in the intestinal cell lines T84 and C2BBe1 upon exposure to PT-BSA (as control condition) or PT-gliadin for 4 hours (Supplemental Digital Content 3). We did not observe any effect in the expression of *HCG14* in C2BBe1 and T84 cells after 4h of gliadin exposure. Nevertheless, to completely discard an effect of gliadin in the expression of *HCG14*, other exposure times and gliadin concentrations should be tested.

Polymorphisms in HCG14 region are associated with decreased NOD1 expression in duodenal intestinal cells

Preliminary eQTL analyses in duodenal epithelial cells (unpublished data) suggested that the differential expression of *NOD1* (*Nucleotide Binding Oligomerization Domain Containing 1*) gene correlates with the genotype of a group of 8 SNPs located in the genomic region of *HCG14.* Indeed, the expression of *NOD1* was around 2-fold higher in intestinal cells homozygous for the common allele compared to those carrying the heterozygous genotype (Figure 3A).

In order to determine whether *HCG14* could be a potential regulator of *NOD1* expression in intestinal cells, we analyzed *NOD1* expression in *HCG14*-silenced T84 cells. Specific small interfering (si)RNAs were able to significantly reduce *HCG14* expression by around 70% in T84 cells (Figure 3B). As shown in Figure 3C, *NOD1* expression was slightly decreased in siHCG14-transfected intestinal cells in comparison with cells transfected with an irrelevant siRNA (siCTRL), suggesting a potential regulation of *NOD1* expression by *HCG14*.

DISCUSSION

Celiac disease is considered a model immunogenetic disease, where several of the participants (genetic as well as environmental) are already known (20). From a genetic point of view, it is clear that the presence of DQ2 (or DQ8) allele is an almost absolute requirement for the development of CeD, although it is not sufficient, since the frequency of HLA-DQ2 allele is also high among non-celiac individuals from the general population (5). Different studies have shown that not all DQ2-harboring haplotypes confer equal susceptibility and have pointed out the presence of additional risk factors on certain HLA-DR3-DQ2 haplotypes (21). Studies comparing CeD patients and non-celiac individuals selected for the presence of DR3-DQ2 have proposed association at several *HLA*-linked loci (22-24) and have supported the presence of additional CeD susceptibility factors that are independent of *HLA-DQ* in the B8-DR3-DQ2 CEH, suggesting their localization around the *HLA* class I region (22-24).

In this study, we have detected two novel variants associated with CeD risk that are not in LD with classical *HLA-DQ* risk variants and could modulate the risk conferred by

different HLA-DR3-DQ2 haplotypes to the disease. One of these polymorphisms, rs3130838, seems to have a risk effect that is independent from DR3-DQ2 alleles.

The CeD-associated SNP rs3130838 is located in an intergenic region flanked by two genes, *TRIM27* and *HCG14*, a protein-coding and a non-coding gene, respectively. The *TRIM27* (or *RFP*) gene encodes a protein of the tripartite motif (TRIM) family that is localized in the nuclear matrix and is involved in the regulation of gene expression and cell proliferation (25). Interestingly, TRIM27 negatively regulates NFkB activation and thus, participates in the regulation of innate immune responses and inflammatory processes (26). Several studies have implicated the NFkB pathway in the development of CeD (27,28), so that the genes related to the regulation of this pathway are potential candidate genes for CeD.

On the other hand, *HCG14* gene encodes an antisense lncRNA with unknown function. In the last years, lncRNAs have emerged as important regulatory molecules of gene expression and several studies have implicated lncRNAs in a wide range of biological processes (29-32). Importantly, lncRNAs are known to modulate gene expression in both *cis* and *trans* acting manners or by direct interaction with the chromatin-modifying proteins and transcription factors (33-36). In contrast to protein-coding genes, lncRNAs are preferentially located in the nucleus and loss-of-function studies have suggested that they have broad effects on gene expression (37). In the present work, we have shown that *HCG14* is preferentially located in the nucleus of intestinal cells, suggesting a potential regulatory function in gene expression.

Interestingly, we have observed that *HCG14* is slightly downregulated in intestinal epithelial cells isolated from CeD patients in comparison to cells isolated from biopsies of non-celiac individuals. Preliminary genome-wide eQTL data (unpublished data)

show that the genotype of 8 SNPs in *HCG14* region correlate with differential expression of *NOD1* gene in intestinal cells from duodenal biopsies. Thus, intestinal cells homozygous for the common alleles (protective for CeD) had higher *NOD1* expression than cells harbouring the heterozygous genotypes. These data suggest that CeD-associated polymorphisms in *HCG14* may alter its function, affecting its capacity to regulate *NOD1* expression. Indeed, silencing of *HCG14* expression by siRNAs led to decreased *NOD1* expression in intestinal cells. Based on published expression data (38), it seems that *NOD1* is preferentially expresed in immune tissues (e.g. spleen or lymph nodes), however NOD1 acts as the first barrier of defense against pathogens in several other tissues, including intestine (39). Indeed, the immune profile of NOD1-deficient ileum suggested a more susceptible epithelial barrier, characterized by decreased expression of antibacterial proteins, such as α- and β-defensins (40).

NOD1 is a member of the NOD-like receptors (NLRs) family, one of the most studied pathogen recognition receptors (PRRs). It is widely expressed in many cell types and is activated upon recognition of specific motifs present in bacterial peptidoglycan. Interestingly, genetic variants in *NOD1* have been associated with susceptibility to inflammatory diseases. Thus, *NOD1* polymorphisms have been associated with the development of atopic eczema, asthma and increased serum IgE concentration (41), while polymorphisms in the intronic region of *NOD1* have been linked with the age at onset of inflammatory bowel disease (IBD) (42).

Our data suggest that CeD-associated variants in *HCG14* region alter *NOD1* expression in intestinal cells, however the exact molecular mechanism underlying this effect and its impact on CeD pathogenesis remains to be clarified. Several studies have suggested that disease-associated SNPs located within non-coding RNAs could perturb their function by disrupting their secondary structure (43). It has been recently shown that a CeD-

associated SNP in *Lnc13* alters its capacity to bind to the ribonucleoprotein hnRNPD, affecting the expression of several pro-inflammatory genes, such as the trascription factors STAT1 and STAT3, or the pro-inflammatory cytokine CCL2 (15). Interestingly, *Lnc13* is less expressed in biopsies from CeD patients compared to controls and as *HCG14* is preferentially located in the nucleus (15). Moreover, specific GWAS SNPs in and around an antisense lncRNA named *ANRIL* can alter the transcription and processing of *ANRIL* transcripts associated with increased susceptibility to coronary disease, intracranial aneurysm, type 2 diabetes, as well as several tumor types (44-46).

In the present work, we have successfullyemployed a CEH-matching strategy to idenfify a novel additional CeD risk variant in the Major Histocompatibility Complex. We have identified a CeD risk variant in total LD with two SNPs, one in the coding gene TRIM27 and the other intronic region of the lncRNA gene *HCG14.* While *TRIM27* expression is not modified in biopsies of CeD patients when compared to controls, suggesting that if TRIM27 is somehow implicated in NFκB signaling pathway alteration observed in the intestine of CeD patients, it might be by a molecular mechanism in which its expression level is not relevant. On the other hand, the variant in the lncRNA *HGC14* seems to regulate the expression of *NOD1* in an allele-specific manner. Further functional studies will allow to clarify the role of *HCG14* in the regulation of gene expression and to determine the molecular mechanisms by which the risk variant in *HCG14* contributes to CeD pathogenesis.

ACKNOWLEDGMENTS

IRG and AJM are supported by Predoctoral Fellowship grants from the UPV/EHU and the Basque Department of Education, respectively. NFJ is a Postdoctoral Fellow supported by the Basque Department of Education. ACR is an Ikerbasque Research Fellow. JRB is funded by Project ISCIII-PI16/00258 and co-funded by the European Union ERDF/ESF "A way to make Europe". ISG is funded by a Research Project Grant 2015111068 of the Basque Department of Health.

REFERENCES

1. Sollid LM. Molecular Basis of Celiac Disease. Annu Rev Immunol 2000;18:53-81.

2. Feighery C. Fortnightly review: coeliac disease. BMJ 1999;319:236-239.

3. Vitoria JC, Arrieta A, Astigarraga I, García-Masdevall D, Rodriguez-Soriano J. Use of serological markers as a screening test in family members of patients with celiac disease. J Pediatr Gastroenterol Nutr 1994;19:304-309.

4. Karell K, Louka AS, Moodie SJ, et al. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. Hum Immunol 2003;64:469-477.

5. 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.

6. Yunis EJ, Larsen CE, Fernandez-Viña M, et al. Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. Tissue Antigens 2003;62:1-20.

7. Aly TA, Eller E, Ide A, et al. Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype. Diabetes 2006;55:1265-1269.

8. Ide A, Babu SR, Robles DT, et al. "Extended" A1, B8, DR3 haplotype shows remarkable linkage disequilibrium but is similar to nonextended haplotypes in terms of diabetes risk. Diabetes 2005;54:1879-1883.

9. Valdes AM, Wapelhorst B, Concannon P, Erlich HA, Thomson G, Noble JA. Extended DR3-D6S273-HLA-B haplotypes are associated with increased susceptibility to type 1 diabetes in US Caucasians. Tissue Antigens 2005;65:115-119.

10. Bilbao JR, Calvo B, Aransay AM, et al. Conserved extended haplotypes discriminate HLA-DR3-homozygous Basque patients with type 1 diabetes mellitus and celiac disease. Genes Immun 2006;7:550-554.

11. Gutierrez-Achury J, Zhernakova A, Pulit SL, et al. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nat Genet 2015;47:577-578.

12. Santin I, Castellanos-Rubio A, Aransay AM, et al. Exploring the diabetogenicity of the HLA-B18-DR3 CEH: Independent association with T1D genetic risk close to HLA-DOA. Genes Immun 2009;10.

13. Araya RE, Gomez Castro MF, Carasi P, et al. Mechanisms of innate immune activation by gluten peptide p31-43 in mice. Am J Physiol Liver Physiol 2016;311:G40-G49.

14. Barone M, Troncone R, Auricchio S. Gliadin Peptides as Triggers of the Proliferative and Stress/Innate Immune Response of the Celiac Small Intestinal Mucosa. Int J Mol Sci 2014;15:20518-20537.

15. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, et al. A long noncoding RNA associated with susceptibility to celiac disease. Science 2016;352:91-95.

16. Monsuur AJ, de Bakker PIW, Zhernakova A, et al. Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. Heutink P, ed. PLoS One 2008;3:e2270.

17. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-265.

18. 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 2014;22:358-362.

19. Mercer TR, Gerhardt DJ, Dinger ME, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol 2011;30:99-104.

20. Kagnoff MF. Celiac disease: pathogenesis of a model immunogenetic disease. J Clin Invest 2007;117:41-49.

21. Karell K, Holopainen P, Mustalahti K, Collin P, Mäki M, Partanen J. Not all HLA DR3 DQ2 haplotypes confer equal susceptibility to coeliac disease: transmission analysis in families. Scand J Gastroenterol 2002;37:56-61.

22. Bolognesi E, Karell K, Percopo S, et al. Additional factor in some HLA DR3/DQ2 haplotypes confers a fourfold increased genetic risk of celiac disease. Tissue Antigens 2003;61:308-316.

23. Lie BA, Sollid LM, Ascher H, et al. A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. Tissue Antigens 1999;54:162-168.

24. McManus R, Moloney M, Borton M, et al. Association of celiac disease with

microsatellite polymorphisms close to the tumor necrosis factor genes. Hum Immunol 1996;45:24-31. 25.

25. Cao T, Borden KL, Freemont PS, Etkin LD. Involvement of the rfp tripartite motif in protein-protein interactions and subcellular distribution. J Cell Sci 1997;110:1563-1571.

26. Zha J, Han K-J, Xu L-G, et al. The Ret finger protein inhibits signaling mediated by the noncanonical and canonical IkappaB kinase family members. J Immunol 2006;176:1072-1080.

27. Maiuri MC, De Stefano D, Mele G, et al. Nuclear factor kappa B is activated in small intestinal mucosa of celiac patients. J Mol Med (Berl) 2003;81:373-379.

28. 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 2014;23:1298-1310.

29. Ponting CP, Oliver PL, Reik W. Evolution and Functions of Long Noncoding RNAs. Cell 2009;136:629-641.

30. Klattenhoff CA, Scheuermann JC, Surface LE, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 2013;152:570-583.

31. Meola N, Pizzo M, Alfano G, Surace EM, Banfi S. The long noncoding RNA Vax2os1 controls the cell cycle progression of photoreceptor progenitors in the mouse retina. RNA 2012;18:111-123.

32. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a

growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 2010;3:ra8.

33. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10:155-159.

34. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904-914.

35. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482:339-346.

36. Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. Cell 2011;145:178-181.

37. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol 2016;17:756-770.

38. Uhlen M, Fagerberg L, Hallstrom BM, et al. Tissue-based map of the human proteome. Science 2015;347:1260419-1260419.

39. Caruso R, Warner N, Inohara N, Núñez G. NOD1 and NOD2: signaling, host defense, and inflammatory disease. Immunity 2014;41:898-908.

40. Robertson SJ, Zhou JY, Geddes K, et al. Nod1 and Nod2 signaling does not alter the composition of intestinal bacterial communities at homeostasis. Gut Microbes 2013;4:222-231.

41. Hysi P, Kabesch M, Moffatt MF, et al. NOD1 variation, immunoglobulin E and

asthma. Hum Mol Genet 2005;14:935-941.

42. McGovern DPB, Hysi P, Ahmad T, et al. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Hum Mol Genet 2005;14:1245-1250.

43. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. Gojobori T, ed. PLoS Genet 2010;6:e1001074.

44. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. Chang HY, ed. PLoS Genet 2010;6:e1001233.

45. Harismendy O, Notani D, Song X, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature 2011;470:264-268.

46. Pasmant E, Sabbagh A, Vidaud M, Bieche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J 2011;25:444-448.

FIGURE LEGENDS

Figure 1. Genetic association analysis of DR3-DQ2-matched celiac disease (CeD) patients and non-celiac controls revealed two associated polymorphisms. Graphical representation of allelic association results obtained comparing CeD patients and non-celiac controls homozygous for DR3-DQ2 and carrying only one copy of the A*1-B*8-MICA*5.1-DRB1*0301-DQB1*0201-(B8-DR3). Results are represented as the –log (p-value).

Figure 2. Expression of the lncRNA *HCG14* is slightly downregulated in duodenal epithelial cells of celiac patients. A) The CeD-associated SNP rs3130838 is located in an intergenic region in chromosome 6 and is in linkage disequilibrium (r^2=1) with two SNPs, one located in the long non-coding RNA *HCG14* (rs3135316) and the other in the *TRIM27* coding gene (rs3135293). B) Relative *HCG1*4 RNA expression levels in $CD45^+$ immune cells purified from intestinal biopsies of non-celiac (Ctrl; n=6) and celiac (CeD; n=10) individuals. C) Relative *HCG1*4 RNA expression levels in $CD326^+$ intestinal cells purified from intestinal biopsies of non-celiac (Ctrl; n=6) and celiac (CeD; n=7). D) Relative *HCG14* RNA expression in a set of human tissues and in the human colon cell line T84. Results are means ± SEM of 3 experimental replicates.E) Relative *HCG14* expression in nuclear and whole extracts of T84 cells. Expression of *Lnc13* and *RPLPO* was used as a control for subcellular fractions. Results are means ± SEM of 4 independent experiments.

Figure 3. Celiac disease-associated SNP genotypes in *HCG14* gene correlate with differential expression of *NOD1* gene. A) Expression of *NOD1* gene stratified by the genotype of a set of 8 SNPs located in the genomic region of *HCG14.* Results are means ± SEM of 12 samples with homozygous genotype and 6 samples with heterozygous

genotype;*p<0.05; Student's *t* test. B and C) Expression level of *HCG14* (B) and *NOD1* (C) in siCTRL and siHCG14-transfected T84 intestinal cells. Results are means ± SEM of 3 independent experiment;*p<0.05; Student's *t* test.

LEGENDS OF SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 2. *HCG14* and *TRIM27* are not differentially expressed in whole intestinal biopsies of control and celiac individuals. *HCG14* (A) and *TRIM27* (B) RNA expression was analyzed by RT-PCR and normalized by the housekeeping gene *RPLO* in intestinal biopsies of control (Ctrl; n=19-21) and celiac (CeD; n=16-20) individuals.

Supplemental Digital Content 3. Short exposure to gliadin does not modify *HCG14* expression in intestinal cells. T84 (A) and C2BBe1 (B) cells were exposed to of pepsin-trypsine-digested BSA or Gliadin for 4 hours. Expression level of *HCG14* was assayed by RT-PCR and normalized by the housekeeping gene *RPLPO*. Results are means ± SEM of 3-6 independent experiments.