Finding the missing 'LiNCs' in celiac disease
Hrdlicková, Barbara

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 01-11-2019
THE ROLE OF CELIAC DISEASE LOCI-ASSOCIATED IncRNA GENES: AN EXPLORATIVE STUDY

Barbara Hrdlickova¹, Vinod Kumar¹, Mathieu Platteel¹, Isis Ricaño-Ponce¹, Renske Veenendaal¹, Cisca Wijmenga⁹, Sebo Withoff⁰

¹University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands.

In preparation
ABSTRACT

Genome-wide association studies (GWAS) have identified common single nucleotide polymorphisms (SNPs) at hundreds of loci for common disease phenotypes and quantitative traits. The vast majority of these phenotype-associated variants (~93%) map to non-coding regions of the genome, including long non-coding RNAs (lncRNAs). To investigate the low frequency variants not present on the GWAS platforms, we performed variant discovery based on amplicon sequencing of four lncRNAs known to be associated with celiac disease (CeD) loci. For this deep sequencing, we used a cohort of 511 Dutch celiac patients. We found three variants (all intronic) in the UBE2E3 locus on chromosome 2 that were significantly associated with CeD. We also identified four exonic SNPs that affect the secondary structure of two lncRNAs (XLOC_005478 and XLOC_013967). These SNPs appear to exert an effect on the secondary structure of the lncRNAs, and might therefore affect the structure-dependent interactions with their partner molecules, and thereby the lncRNA function. Moreover, we explored the potential function of the detected variants by annotating them with features indicative of biochemical functions. This analysis revealed that 83% of the exonic variants described in the sequenced long intergenic non-coding RNA (lincRNA) genes intersect with enhancer histone marks, highlighting their possible importance in gene regulation. We identified other regulatory elements based on ENCODE data. Finally, we investigated cis-eQTL (expression-quantitative trait loci) effects of lncRNA variants associated with CeD loci on protein-coding and non-coding RNA genes in peripheral blood. This suggested the functional mechanism of the genetic variants. Together this information provides a first glimpse of the potential effects of genetic variation on lncRNAs in CeD associated-loci, which might be a mechanism involved in CeD pathophysiology.
INTRODUCTION

The biggest surprise coming out of genome-wide association studies (GWAS) is that the vast majority of disease/trait-associated single nucleotide polymorphisms (SNPs) reside in non-coding regions. Maurano et al. showed that ~93% of GWAS-variants associated with diseases or quantitative traits are concentrated in regulatory parts of DNA that most probably act in a cell-specific and/or developmental-stage manner [1]. These regions can harbor gene regulatory sequences (promoters, enhancers) or regulatory non-coding RNAs, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) [1]. When only looking at the autoimmune disease-associated SNPs in the GWAS catalogue [2], this percentage does not change dramatically [3], suggesting that in autoimmune diseases the majority of disease-associated SNPs also affect gene expression rather than protein structure. This can either be achieved directly, by disruption of promoter/enhancer (transcription factor binding sites) sequences, or indirectly, by affecting the function or expression of regulatory RNAs.

Previously, we have performed GWAS and Immunochip studies on celiac disease (CeD), [4–6]. These studies discovered 57 independent CeD-associated signals (SNPs) in 39 non-HLA loci. Despite these exciting results, we realize that these loci still need to be fine-mapped, as in most cases they still represent very large DNA sequences and can encompass multiple genes. In CeD, 54 of the 57 variants are located in non-coding genes or in regulatory regions (5’-UTR, 3’UTR, intronic or intergenic regions). In fact, in silico analysis of the 39 CeD-associated non-HLA loci showed that these loci contain many non-coding RNA sequences, and in particular long intergenic non-coding RNAs (lncRNAs, n=43) from the Human lincRNA catalogue [7]. By looking more carefully into the CeD loci we identified some in which the associated SNPs do not seem to affect protein-coding genes but rather lncRNAs.

As there is no data on how autoimmune SNPs affecting lncRNAs contribute to disease phenotypes, we decided to perform an explorative study focusing on three loci. We hypothesized that the SNPs associated with lncRNAs in CeD-associated loci might affect the lncRNAs by altering their structure, and/or function, or their expression. We performed variant discovery based on amplicon sequencing in four potentially interesting CeD-associated lncRNAs. This was done in 511 Dutch CeD patients in order to capture low frequency/rare variants, which we tested for association with CeD. Next, we interrogated the variants detected in those lncRNAs in terms of their potential structural and functional consequences. We identified SNPs predicted to disrupt structure and found that some of them intersect with regulatory elements in the genome. Finally, we found that some of the discovered variants exert cis-eQTL effects in peripheral blood cells. Together these results suggest the functional potential of the genetic variants found by our deep-sequencing of CeD patients.

MATERIALS AND METHODS

CeD loci-associated lncRNA prioritization

To select candidate lncRNA genes for our study, we zoomed into the linkage disequilibrium (LD) block defined in our CeD Immunochip study [6], plotted the association signals for all 39 non-HLA loci from this study [6] in R-studio [8], and annotated them with gene coordinates from GENCODE release 10 and from the Human lincRNA catalogue [7, 9]. We prioritized three CeD loci, in which the top Immunochip association signals seem to affect lincRNAs instead of protein-coding genes, for variant discovery based on amplicon sequencing (Figure 1).
Celiac patient dataset sequencing and variant discovery

DNA samples had previously been isolated from whole blood samples of 511 unrelat-
ed celiac patients who had been used in the Immunochip study [6]. Written informed
consent was obtained from all subjects after we obtained approval from the ethics com-
mittees and institutional review boards of all the participating institutes. All subjects were
Dutch with European ancestry. Individuals with celiac disease were diagnosed according to standard criteria [10].

Identification of variants and sequencing of the selected lncRNA genes was performed
using Sanger sequencing. Primer sequences and conditions for PCR are given in the
Supplementary materials (Supp. Table 1). PCR amplifications were performed in 96-
well plates, in a total volume of 18 ul. PCR products were purified with 1x ExoSap
(12.5 ul Shrimp Alkaline Phosphatase (1 U/ul), 1.25 ul Exonuclease I (20 U/ul), (both
Thermo Scientific, Waltham, MA, USA), 250 ul Glycerol, (Sigma-Aldrich, St. Louis, MO,
USA), 236.25 ul Milli-Q water). Samples were directly sequenced using the Applied
Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit and analyzed using the
Applied Biosystems 3730xl DNA Analyzer (both Thermo Scientific) and Mutation
Surveyor V3.23 DNA Variant Analysis Software (SoftGenetics, PA, USA). Only ge-
etic variants detected in at least two individuals were followed up.

Genotypes of controls

Dutch population-based controls were gen-
otyped using the Illumina HiSeq 2000 plat-
form (Illumina Inc., San Diego, CA, USA) at
the Beijing Genomics Institute (BGI, China).
The genotype calling algorithms and quality
control parameters have been described pre-
viously [11]. We selected genetic variants
discovered by Sanger sequencing of the CeD
cohort in at least two individuals and extract-
ed the genotypes from the GoNL website
(release four) based on their chromosome
positions (http://www.nlgenome.nl/search/).

Variant annotation and association study

The lncRNA sequences were analyzed for intronic, exonic, and intergenic variants
(both novel as well as known SNPs). All variants were annotated by base pair posi-
tion in human genome 19 (hg19) and with respect to their localization in the lncRNA
genes. Standard case-control single-variant association analysis was performed using
a two-tailed Fisher’s exact test to examine the differences in allele frequency.

Secondary structure-disruptive SNPs
with a local effect on CeD lncRNAs

We used RNAsnp [12, 13] to investigate in sil-
ico whether the lncRNA-associated variants
could affect the local secondary structure of the
lncRNAs. This algorithm identifies local
sub-regions exhibiting structural differences
in the presence of wild-type versus mutant
(SNP) alleles and quantifies these changes by
empirical P-values. The effect of SNPs on lo-
cal RNA secondary structure can be detected
for short sequences (<1000 nt) by RNAfold
[14, 15]; for large sequences (>1000 nt) by
RNAplfold [14–16]; and for general screening
by combination of RNAfold and RNAplfold.
The differences between wild-type and mu-
tant (SNP) base pair probabilities for the local
regions were measured using the Euclidean
distance with the corresponding P-values. It
was suggested by the developers of RNAsnp
that a P-value less than 0.2 (P<0.2) represents
a significant structural change of a local re-
region. The outputs were visualized in the
RNAsnp webserver [12].

Functional annotation of CeD loci
associated SNPs

In order to interpret the functional con-
sequences of the SNPs discovered, the
observed variants were annotated with regulatory chromosomal features extracted from RegulomeDB and HaploReg V3 databases [17, 18]. RegulomeDB includes information from various datasets from the ENCODE project [9] and the NCBI Sequence Read Archive (e.g. transcription factor (TF) ChIP-seq data, histone ChIPSeq results, FAIRE experiments, DNAseq hypersensitive site determination, eQTL mappings, etc.), [17]. It also provides algorithms to predict TF and protein-binding sites and a scoring system that allots a score to each variant based on the number and type of its overlapping functional properties. HaploReg provides a similar systematic annotation of non-coding SNPs using SNP identification numbers (rs ID) based on chromatin state (promotor/enhancer histone marks, DNAseq hypersensitive sites), conserved regions (calculated by two mammalian conservation algorithms: GERP [19], and SiPhy [20]), eQTL, gene annotation (RefSeq, GENCODE) and predicts the effect of SNPs on regulatory motifs for TF/protein-binding [21].

Expression quantitative trait loci analysis in PBMCs in a Dutch cohort (eQTLs)
Cis-eQTL mapping for all exonic SNPs and variants significantly associated (P<0.05) with CeD was performed in 629 peripheral blood samples from the LifeLines-deep cohort [22]. The detailed normalization, correction of expression data based on principal component analysis, and eQTL mapping method has been described previously [23, 24]. Briefly, cis-eQTL analysis was performed on SNP-gene combinations for which the distance from the center of the transcript to the genomic location of the SNP was ≤250 kb. All SNP-gene expression associations outside this 500 kb window were considered to be trans-eQTL. As it is known that cis-eQTLs might be driven by LD, rather than by a direct effect of the “disease gene SNP” (gSNP, in this case the CeD variant), we performed conditional analysis on the gSNP-gene pairs and the “top eQTL SNP” (eSNP; SNP with maximal cis-eQTL effect on the same gene probe as gSNP)-gene pairs [Ricaño-Ponce and Wijmenga, manuscript in preparation]. All associations were tested by non-parametric Spearman’s rank correlation test and corrected for multiple testing using a false discovery rate (FDR<0.05) based on the number of SNPs.

Unfortunately, the eQTL pipeline established for LifeLines data only focused on GENCODE as a reference database. Therefore, we limited our analysis to the GENCODE lncRNAs when some of them overlapped with our studied lncRNAs from the Human lincRNA catalogue [7].

Expression profiles of eQTL genes in seven immune cell types
The cell-type-specific expression profiles of the lncRNAs and protein-coding genes associated with independent eQTL effects were extracted from an RNA-sequencing dataset obtained from seven peripheral blood leukocyte populations (granulocytes, monocytes, natural killer (NK) cells, B-cells, memory T-cells, naive CD4+ and naive CD8+ T-cells) [25]. The visualization of the relative expression profiles was performed using Gene-E software [26].

RESULTS

Immunochip CeD study and LncRNA gene selection
We prioritized three CeD-associated loci in which the peak of association signals from the Immunochip study [6] maximized in one or a cluster of multiple lincRNA genes from the Human lncRNA catalogue [7], (Figure 1). These loci, located on chromosomes 2, 6 and 21 (Figure 1) were previously named after three protein-coding genes they contained (UBE2E3, TNFAIP3, and ICOSLG, respectively). In total, we observed 18 CeD loci-
Figure 1: Association of selected CeD-associated loci. Plots of the CeD Immunochip study data annotated with lincRNAs (red) and other genes (blue). All lincRNAs are represented by a red bar and indicated by an ID (“lincRNA” followed by its “XLOC” number; e.g. lincRNA XLOC_001776 is visualized as “lincRNA_001776”). The UBE2E3 locus (chr2) contains one protein-coding gene UBE2E3 (+), one microRNA 4437 (-) and three lincRNAs XLOC_001776 (+), XLOC_002417 (-), XLOC_001775 (+). The TNFAIP3 locus (chr6) accommodates, besides one microRNA and one miscRNA, one protein-coding gene TNFAIP3 (+) and seven lincRNAs: XLOC_005847 (-), XLOC_005478 (+), XLOC_005848 (-), XLOC_005479 (+), XLOC_005849 (-), XLOC_005480 (+), XLOC_005851 (-). The last locus (ICOSLG) encompasses three protein-coding genes: ICOSLG (-), DNMT3L (-), AIRE (+); one previously predicted mRNA now annotated to be antisense IncRNA APO1059.5 (+), and eight lincRNAs: XLOC_013966 (+), XLOC_013967 (+), XLOC_013968 (+), XLOC_014111 (+), XLOC_014112 (+), XLOC_014113 (-), XLOC_014114 (-), XLOC_014115 (-). All genomic positions are given as NCBI Build 36 (hg18) coordinates. Signs between brackets (+,-) represent the plus or minus DNA strand. The lincRNAs highlighted by red circles were sequenced in our study.
Figure 2: Regions of CeD-associated loci lincRNAs selected for deep-sequencing. Each region is represented by a UCSC genome browser view showing the location of the top CeD-associated SNP (in red square). The proxy SNPs ($r^2 \geq 0.8$) are shown as pink stripes and the region of high LD as a pink bar. LincRNA transcripts from the Cabili catalogue [7] with their transcript names (TCONS_number) are shown in red with a black frame and the lincRNA term with the number from their XLOC gene name (e.g. lincRNA named XLOC_001776 is visualized as “lincRNA_001776”). The black squares capture the ten sequenced amplicons associated lincRNAs across all three regions (LD blocks). For the sequencing we focused on 4 transcripts (associated with the ‘peak’ of the association signal), which were subdivided into 10 amplicons to be sequenced (Figure 2).

Sequencing results and CeD loci annotated lincRNA variants

We selected 511 Dutch CeD patients and performed Sanger sequencing, with an extensive quality control using manual corrections. We discovered 100 variants across all the amplicons except one (3rd amplicon, of the ICOSLG locus). Of these, 37 variants were located in lincRNA exons; 46 in introns.
Table 1: Variants types in non-coding regions of CeD-loci associated lncRNAs.

<table>
<thead>
<tr>
<th>Variant type</th>
<th>All variants</th>
<th>Variant type</th>
<th>All variants</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>exonic</td>
<td>37</td>
<td>exonic</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>intronic</td>
<td>46</td>
<td>intronic</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>intergenic</td>
<td>17</td>
<td>intergenic</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total variants</td>
<td>100</td>
<td>Total variants</td>
<td>42</td>
<td>15</td>
</tr>
</tbody>
</table>

Singleton (occurrence =1x) 58
Doubleton (occurrence=2x) 6
Multiton (occurrence >2) 42

Novel=not seen in dbSNP (141) nor in 1000Genome Project nor GoNL (version 4)

Figure 3: Work flow of selecting loci and sequencing results. All 39 non-HLA loci were annotated with the Human lincRNA catalogue and inspected for association peaks from our Immunochip study into celiac disease. This led to three loci being selected. Ten amplicons containing CeD loci-associated lncRNAs were sequenced, resulting in 100 variants. Of these, 42 variants were explored further.

of lncRNAs and 17 variants were outside both coding and non-coding genes (Table 1). 58 variants were only seen in one (singleton) and six variants were only seen in two individuals (doubletons). For further analysis we selected variants which were present in two or more individuals (n=42) and 18 of these (43%) were in exons, while 23 (55%) were located in introns of sequenced lncRNAs. Only one of the original 42 variants was located intergenically. Of all 42 variants, 15 (35.6%) are novel compared with published datasets (dbSNP141, Genome Project, GoNL), [27–29]. To investigate differences in allele frequencies between CeD patients and controls at the 42 variants, we extracted genotype data of population-based controls from GoNL [29] and performed a single-variant association analysis. This revealed three intronic variants, all in the same region in the UBE2E3 locus, with significantly different allele frequencies between celiac patients and controls (P-value<0.05; Table 2, Supp. Table 2).

SNPs with a local effect on the secondary structure of CeD lncRNAs

Although the mechanism of action for most lncRNAs is unknown, it has been reported that the secondary structure of a lncRNA determines its function [30]. It was also proposed that genetic variants that affect this secondary structure lead to the altered function or activity of the lncRNA and could possibly contribute to disease phenotype in this way [31]. We used RNAsnp [12, 13] to study the effects of our 17 exonic variants on local secondary structure. We found four local structure-disruptive SNPs that could potentially cause significant changes to the secondary structure (Figure 4, Supp. Table 3). Two different transcripts (TCONS_00012607 and TCONS_00011381) of lincRNA XLOC_005478 in the UBE2E3 locus appeared to be affected structurally by two different SNPs (chr6:137994564 (P=0.106), rs625124 (P=0.1578)) located 105 bp away from each other (Supp. Table 3). In the ICOSLG locus, two SNPs (rs113526534. (P=0.1567) and rs188507544 (P=0.1991)) are separated by 279 bp and affect the secondary structure of TCONS_00028855, the only transcript of lincRNA XLOC_013967 (Supp. Table 3)
Table 2: Summary of results for the 42 variants selected for further analysis. The variants were annotated with results from the single-variant association study (P<0.05), secondary structure prediction (RNAsnp), functional element annotations (RegulomeDB, HaploRegV3) and eQTL-mapping. Results are color-coded; yes (red), no (blue), not available in the dataset or not tested (white). RegulomeDB classification scores are color-coded: category 2 “likely to affect binding” (red), category 3 “less likely to affect binding” (orange), lower categories (4-6) with “minimal binding evidence” (blue).

<table>
<thead>
<tr>
<th>Locus</th>
<th>Amplicons</th>
<th>SNP_ID</th>
<th>Doubleton / Multiton</th>
<th>Exonic</th>
<th>P&lt;0.05</th>
<th>Structure</th>
<th>RegulomeDB</th>
<th>GERP cons.</th>
<th>SiPhy cons.</th>
<th>Promoter</th>
<th>Enhancer</th>
<th>DNAse</th>
<th>Protein Motif</th>
<th>eQTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-1</td>
<td>rs7588217</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs7496727</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs7593597</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs62180198</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs34700861</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs35110674</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs7580110</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr2:182091077</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs148553735</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs667158</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR-2</td>
<td>chr:6:137994564</td>
<td>D</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs625124</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137994916</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs605777</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137995009</td>
<td>D</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR-3</td>
<td>chr:6:13799300</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137995402</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs78986857</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137995650</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137995651</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:137995652</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:13800895</td>
<td>D</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1518048</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:6:138008255</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs75525553</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR-4</td>
<td>rs113526534</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:21:45621257</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs62228210</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:21:45621413</td>
<td>D</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs188507544</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs62228211</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs144086965</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2329709</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:21:45621779</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6518350</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR-7</td>
<td>rs75578787</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs56299324</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs56132007</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs55965762</td>
<td>M</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR-8</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PCR-9</td>
<td>chr:21:45626120</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>chr:21:45626789</td>
<td>M</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: The effects of structure-disruptive SNPs on IncRNAs. These structures are based on the minimum free energy of wild-type (green) and mutant (red) sequences. The figure visualizes the structure-disruptive effect of four SNPs located in XLOC_005478 in the UBE2E3 locus (A,B) and in XLOC_013967 in the ICOSLG locus (C,D).
THE ROLE OF CELIAC DISEASE LOCI-ASSOCIATED lncRNA GENES

C) rs113526534 (P=0.1567)
   wild-type
   mutant

D) rs188507544 (P=0.1991)
   wild-type
   mutant
Functional annotation of SNPs associated with CeD loci

Because the minor allele frequency (MAF) for more than half (24/42) of our candidate SNPs was below 5%, they were often not represented in the databases (RegulomeDB, HaploReg) we used for the functional annotation of variants (Supp. Table 2). We used both databases to gain the most information. RegulomeDB allows us to search for variants based on the chromosome position; therefore observed variants without a reference SNP identification number (rs ID) can be included. HaploReg, on the other hand, utilizes more recent data of ENCODE [32], but variants without an rs ID are excluded from the analysis. RegulomeDB provided data for 28 SNPs (67%). Four of our candidate variants (chr21:45621413; rs56132007, rs55965762, and chr21:45626789), all located in the ICOSLG locus on chromosome 21, were scored as “a functional category likely to affect binding” (RegulomeDB score 2b). Category 2 scores indicate direct functional evidence (ChIP-seq data, DNaseI hypersensitivity analysis, with either a matched position weight matrices (PWM) to the ChIP-seq factor or a DNase footprint). An additional intronic SNP (rs74796727) from the TNFAIP3 region was scored as “3a” by RegulomeDB, which is “less likely to affect binding” (SNPs in this category intersect with ChIP-seq evidence and either have a motif that matches the ChIP-seq data, but not with DNaseI evidence, or have DNaseI evidence simultaneously with any other regulatory motif). The remaining 23 variants were placed in even lower categories (4-6) with “minimal binding evidence”.

Using HaploReg V3 we were able to annotate 24 variants (57%), (see Supp. Table 2 for overlap). Next, we focused on the possible overlap between functional elements and extracted information for three regulatory properties: promoter and enhancer histone marks and DNaseI hypersensitivity sites. Six variants out of 24 shared all three functional annotations (promoter, enhancer and DNaseI marks). The most common functional annotation intersected by our candidate variants was enhancer histone marks (20/28 SNPs). We also analyzed the overlap between all three functional elements on the organ/tissue level. All three elements were discovered to be functional in eight tissues (gastro-intestinal tract, placenta, muscle, blood, heart, skin, breast, and thymus), of which the gastro-intestinal tract, blood, and thymus can be linked to CeD etiology.

Expression quantitative trait loci analysis

To gain a deeper understanding of the function of the variants in the CeD loci-associated lincRNAs, we performed eQTL mapping: firstly, on four Immunochip top variants, and secondly, on 21 sequenced variants: i) 18 exonic SNPs and ii) three SNPs with significantly differently presence (P<0.05) in Dutch CeD patients and population-based controls from the GoNL project [11, 29]. We identified a significant cis-eQTL (P<0.05; FDR<0.05) for one top CeD Immunochip variant: rs1018326 (chr2) on the UBE2E3 protein-coding gene (P=0.033, Z-score=-2.13; Table 2, Supp. Table 4). Subsequently, we analyzed the 21 variants and observed a cis-eQTL (P<0.05; FDR<0.05) for one variant in the UBE2E3 locus, for one variant in the TNFAIP3 locus, and for six variants in the ICOSLG locus (Table 2). In the latter, we observed eQTL effects on protein-coding as well as on non-coding genes (lincRNA, antisense IncRNA). Interestingly, we saw no eQTL-effect on TNFAIP3, although we did observe a significant effect of variant rs78986857 on gene RP11-356I2.2 (corresponding to lincRNA XLOC_005848) in the TNFAIP3 locus. (Supp. Table 4). In summary, we observed significant eQTL effects on four genes: on the UBE2E3 protein-coding gene in the UBE2E3 locus (chr. 2), on the
Table 3: Protein-coding and non-coding RNA genes showing significant eQTL effects after conditional analysis to the SNP (eSNP) with the maximal eQTL effect on the same gene-probe as the tested gSNP. SNP rs10188326 is a top-association signal from Immunochip; the remaining variants are out of 42 SNPs described in this study. Additional detailed eQTL analysis results are given in Supp. Table 4.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Gene with eQTL</th>
<th>Gene type</th>
<th>eQTL effect direction</th>
<th>gSNP P-value</th>
<th>gSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>UBE2E3</td>
<td>protein-coding</td>
<td>↓</td>
<td>0.0331</td>
<td>rs1018326 (iCHIP top SNP)</td>
</tr>
<tr>
<td>6</td>
<td>RP11-356I2.2</td>
<td>lincRNA (XLOC_005848)</td>
<td>↑</td>
<td>0.0007</td>
<td>rs7896857</td>
</tr>
<tr>
<td>21</td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
<td>↑</td>
<td>0.0197</td>
<td>rs113526534</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑</td>
<td>0.0210</td>
<td>rs62228210</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑</td>
<td>0.0184</td>
<td>rs62228211</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑</td>
<td>0.0273</td>
<td>rs6518350</td>
</tr>
<tr>
<td></td>
<td>ICOSLG</td>
<td>protein-coding</td>
<td>↓</td>
<td>0.0240</td>
<td>rs188507544</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↓</td>
<td>0.0013</td>
<td>rs2329709</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑</td>
<td>0.0210</td>
<td>rs62228211</td>
</tr>
</tbody>
</table>

RP11-356I2.2 gene (lincRNA XLOC_005848, chr. 6) in the TNFAIP3-locus, and finally on the antisense lncRNA AP001059.5 and the ICOSLG protein-coding gene (both in the ICOSLG-locus on chr. 21) (Table 2). The eQTL-genes were further followed-up below.

Expression profiles of cis-eQTL genes
The cell-type-specific expression profiles of the discovered cis-eQTL genes are presented in the heatmap in Figure 5. Two genes, UBE2E3 and lincRNA RP11-356I2.2 (XLOC_005848), are highly expressed in natural killer cells and another two genes (ICOSLG and lncRNA AP001059.5) in B-cells (Figure 5). These data suggest that NK cells and B-cells are the most interesting candidate cell types in which to study the mechanism of action of the two selected lncRNAs.

DISCUSSION
GWAS and Immunochip studies have identified common variants of modest effect in 39 non-HLA loci. The vast majority (approx. 95%) of these CeD-risk variants have been found outside regions coding the amino acid sequence of proteins, but located within the regulatory regions (such as three prime untranslated regions (3’UTRs), 5’UTRs, introns, intergenic regions), [33]. In this explorative study, we prioritized and deep-sequenced four lncRNA genes associated to CeD loci. This discovered 42 variants that were detected in at least two CeD patients.

Almost half (47%) of the variants discovered here exhibited a minor allele frequency (MAF) below 2% in sequenced
celiac patients, and we also detected eight rare variants (MAF ≤ 0.5%). Although our study is underpowered, even with this relatively small number of samples (511 cases, 498 controls) we detected significant differences between CeD and control individuals (P < 0.05) for three SNPs, suggesting the association of these three variants must be strong. All three are located in an intron of lncRNA XLOC_001776 (UBE2E3 locus) and all cause a change in binding motifs for multiple chromatin-modifying proteins and transcription factors. Two of the variants (rs74796727, rs62180198) also intersect with enhancer histone marks. Moreover, SNP rs74796727 is located in a conserved region (according to GERP score) annotated with enhancer/promoter histone marks and DNAseI hypersensitivity. As these results were obtained with a small number of subjects, it is possible that increasing the sample size would improve the significance of the three associations. Moreover, this increase would also open up the possibility of finding more associations.

Secondary structures dictate the function of lncRNAs and their interplay with interaction partners via their RNA-, DNA- or protein-binding domains [34]. Hence, a mutation of a single nucleotide could change the local secondary structure dramatically and affect lncRNA function in such a way that it leads to pathophysiological consequences. We investigated potential structural changes induced by the exonic variants (n=18) in our dataset and our results suggest that four SNPs induce significant structural changes in two lncRNAs. One of these SNPs (chr6:137994564 in the TNFAIP3 region) unfortunately has no rs ID and therefore no other data is available on it. A second SNP (rs625124) in the same locus intersects with DNAseI marks and modifies a region that changes the binding motif for a transcriptional repressor PLZF (promyelocytic leukemia zinc finger, Zbtb16) a key transcription factor involved in the development of natural killer T-cells [35, 36]. Interestingly, one of the main cell types involved in CeD are intraepithelial CD8+ T-lymphocytes (IELs) and these kill the intestinal epithelial cells (enterocytes) leading to intestinal barrier dysfunction. It is thought that these IELs represent a lymphocyte subset that is losing T-cell characteristics and acquiring NK cell properties [37]. In this sense, an interaction of lncRNA and PLZF in CeD-associated IELs could contribute to CeD pathology. The two remaining SNPs (rs113526534, rs188507544) are both located in the ICOSLG region, both intersect with enhancer histone marks and confer a cis-eQTL effect. Variant rs113526534 exhibits a cis-eQTL effect on antisense lncRNA AP001059.5. The other, rs188507544, shows a cis-eQTL effect on two protein-coding genes C21orf2 and ICOSLG (two proteins highly expressed in B-cells according to co-expression analysis using genenetwork.nl) in combination with a DNAseI mark and a binding motif change for multiple proteins (Supp. Table 2). Also, the involvement of B-cell proteins makes sense in the context of CeD, as B-cells specific for the gluten antigen in the small intestine may amplify the immune response signal for CD4+ T-cells by presenting gluten antigens on their surface [38].

Many of the variants we discovered by deep-sequencing can be classified as less common (<5%) or rare (<0.5%) variants. Many of them do not have an rs identification number (given by dbSNP) and this makes it difficult to search for their potential effects or functional annotation in the normal databases. We observed that 83% (n=20) of 24 SNPs for which data was available in HaploReg V3 intersected with regions marked with enhancer histone marks. Six SNPs (rs75576871, rs56299324, rs74796727, rs55965762, rs56132007, rs7588217) intersected with three functional properties simultaneously (enhancer/promoter histone marks and DNAseI marks) and 10 SNPs (50%) annotated only with enhancer marks.
The six SNPs described here in more detail are all known variants recorded in dbSNP, are all intronic, and they were all also annotated to bind protein. Altogether this suggests the SNPs affect expression of another gene. These data suggest that the majority of the SNPs that we identified affect regions that have gene regulatory functions.

Correlating the RNA gene expression levels with SNPs and identifying eQTL may help to functionally annotate SNPs and pinpoint which genes they are likely to regulate [39]. Therefore, we first mapped eQTLs to the four top associated SNPs from the Immunochip study that were located in our three selected loci and performed conditional analysis. Only the Immunochip CeD-top association signal on chromosome 2 displayed an independent eQTL effect, which was directed on to the protein-coding gene UBE2E3. Next, we mapped eQTLs to all exonic variants (n=18) and also to the three significantly associated genetic variants from our single-variant association study (P<0.05). After conditional analysis, we observed a strong independent effect of exonic variant rs78986857 (P=6.8x10^(-4)) located in lncRNA XLOC_005478 in the TNFAIP3 locus; this increased expression of a second lncRNA (RP11-356I2.2=XLOC_005848) located in the same LD block but outside the association peak as identified by the Immunochip study (Figure 1). This exonic variant was not predicted to have a significant effect on the secondary structure of lncRNA XLOC_005478 and was functionally annotated with enhancer histone marks only in blood and a motif change for TATA-binding protein (TBP) using HaploReg (this variant was not found in RegulomeDB). TBP is a general transcription factor binding to the TATA-box, a region approximately 30 bp upstream from the transcription start site in many promoters.

Another interesting eQTL observation was that in PCR-region 7 (XLOC_013967; ICOSLG locus) six out of ten variants displayed an eQTL effect on either the protein-coding gene ICOSLG or on antisense lncRNA AP001059.5. All ten variants in this sequenced region were exonic, but only two (rs113526534, rs188507544) affected the secondary structure of their lncRNA XLOC_013967 markedly. It is possible that a combination of two strong, secondary structure disruptive-SNPs together with other variants with a modest disruptive effect may together dramatically alter lncRNA XLOC_013967 and its function. All ten SNPs were present in RegulomeDB but only one, chr21:45621413, was categorized as “likely to affect binding” with a score “2b”; the remaining nine variants were labeled with lower scores (4 and 5) representing “minimum binding evidence”. Seven variants were present in HaploReg, all displaying enhancer histone marks in muscles, placenta and/or gastrointestinal tract, and some of them also had DNaseI hypersensitivity sites with/without motif changes. These enhancer-associated features suggest that in the relevant tissue this lncRNA may function as an active enhancer (enhancer-derived RNA), binding multiple transcription factors and interacting with nearby genes [40, 41].

One of the main characteristics of lncRNAs is their functionality in the correct tissue/cell type. Using eQTL mapping we can predict which gene might be affected by a specific variant, but a detailed analysis of expression data from a range of different cell types and tissues might pinpoint a relevant environment in which the selected lncRNA is functional. In our study, we discovered seven variants with a cis-eQTL effect on coding as well as non-coding genes. For these variants we extracted their relative expression across seven immune cell types. The results show that our eQTL genes are most highly expressed in B-cells (lncRNA AP001059.5, protein ICOSLG (both chr. 21)) and NK cells (protein UBE2E3 (chr. 2), lncRNA RP11-356I2.2=XLOC_005848 (chr. 6)) followed by...
monocytes (protein UBE2E3 (chr. 2)). We described how these cell types have been implicated in CeD. For experimental follow-up on any of the candidate SNPs or genes implicated in this study, both NK cells and B-cells should also be considered, although CeD is still often regarded as a “T-cell disease”.

To our knowledge, this study is only the second one to consider the annotation of GWAS-associated variants and their effect on lncRNAs. In the first study, the authors focused on association signals found by GWAS and Immunochip in type 1 diabetes and inflammatory bowel disease and on common variants from these platforms present in the disease loci [42]. In contrast, our study involves amplicon-sequencing-based variant discovery and was specifically designed to reveal less common novel variants.

Understanding the functional mechanism underlying the GWAS associations with disease phenotype is challenging. Here, we have presented the results of an exploratory study using an integrated approach based on deep-sequencing and annotation of SNPs that appear to affect lncRNA genes in celiac disease loci. This type of analysis may help to improve our understanding of the molecular effects of non-coding variants associated with celiac disease and their resulting physiologic and pathophysiologic consequences. We observed evidence of functional effects at multiple levels (on chromatin as well as on transcript level), but functional experiments are necessary to validate our findings.

ACKNOWLEDGMENTS

The authors thank Jackie Senior and Kate McIntyre for editing this text. The senior authors received funding from the European Research Council under the EC Seventh Framework Program FP/2007-2013 (ERC Grant Agreement no. 322698 to CW), the Netherlands Organization for Scientific Research (NWO-VICI 918.66.620 to CW), the Dutch Digestive Diseases Foundation (MLDS, WO11-30 to CW), the Sigrid Juselius Foundation to CW, and the Dutch Multiple Sclerosis Foundation (11-752 to SW).

REFERENCES


## Supplementary table 1

<table>
<thead>
<tr>
<th>Locus</th>
<th>Amplicon</th>
<th>Primer</th>
<th>Primer sequence</th>
<th>Length</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBE2E3</td>
<td>PCR-1</td>
<td>Forward</td>
<td>gggatgcgtataatggttcc</td>
<td>819bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>gaatgcctgcgtgtaataaagg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-2</td>
<td>Forward</td>
<td>tttgatgtacccctttggttga</td>
<td>784bp</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>tttcacctgtctctcagcttc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-3</td>
<td>Forward</td>
<td>gtcccattaacccaaaagata</td>
<td>772bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>tgtctcaattccatcattacca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>PCR-4</td>
<td>Forward</td>
<td>ctaaggtgtgctgggagctga</td>
<td>822bp</td>
<td>61.4°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>gcaggttagctagtttttaatgaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-5</td>
<td>Forward</td>
<td>caggtaggggagggactgaga</td>
<td>762bp</td>
<td>59.3°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>ctaaacctgtgacaaaaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-6</td>
<td>Forward</td>
<td>atgtgagctgttgctgtt</td>
<td>708bp</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>caatggagacagggagaaaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOSLG</td>
<td>PCR-7</td>
<td>Forward</td>
<td>tggccagagctacaatctcc</td>
<td>817bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>aggcagaaaaacgtaagg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-8</td>
<td>Forward</td>
<td>ggtccctttctctgctctgat</td>
<td>701bp</td>
<td>60.4°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>cctgcaacatttactgacg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-9</td>
<td>Forward</td>
<td>ggcagccccagctctcacttc</td>
<td>726bp</td>
<td>60.4°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>gcagcttttttttttaacgaaag</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR-10</td>
<td>Forward</td>
<td>cttgctctctattgctggt</td>
<td>842bp</td>
<td>60.4°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>ggacacctgaattccgacacta</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary table 3.**

<table>
<thead>
<tr>
<th>A locus</th>
<th>SNP</th>
<th>SNP ID</th>
<th>position</th>
<th>CHR</th>
<th>BP [hg19], GoNL</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBE2L3</td>
<td>rs148553735</td>
<td>rs148553735</td>
<td>EXONIC</td>
<td>2</td>
<td>182091213</td>
<td>0.4482</td>
<td>0.796</td>
<td>0.7796</td>
<td>0.6167</td>
<td>0.7981</td>
<td></td>
</tr>
</tbody>
</table>

**B locus**

<table>
<thead>
<tr>
<th>SNP</th>
<th>SNP ID</th>
<th>position</th>
<th>CHR</th>
<th>BP [hg19], GoNL</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFAIP3_3-1_2_116</td>
<td>chr6:137994564</td>
<td>EXONIC</td>
<td>6</td>
<td>137994564</td>
<td>0.106</td>
<td>--</td>
<td>--</td>
<td>0.9217</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3-1_2_221</td>
<td>rs625124</td>
<td>EXONIC</td>
<td>6</td>
<td>137994669</td>
<td>0.2145</td>
<td>0.5335</td>
<td>0.1578</td>
<td>0.6475</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3-2_3_552</td>
<td>rs7896857</td>
<td>EXONIC</td>
<td>6</td>
<td>137995550</td>
<td>0.9986</td>
<td>--</td>
<td>0.9997</td>
<td>0.9896</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3-2_3_652</td>
<td>chr6:137995650</td>
<td>EXONIC</td>
<td>6</td>
<td>137995650</td>
<td>0.5023</td>
<td>--</td>
<td>0.5691</td>
<td>0.489</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3-2_3_653</td>
<td>chr6:137995651</td>
<td>EXONIC</td>
<td>6</td>
<td>137995651</td>
<td>0.3401</td>
<td>--</td>
<td>0.4046</td>
<td>0.2114</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3-2_3_654</td>
<td>chr6:137995652</td>
<td>EXONIC</td>
<td>6</td>
<td>137995652</td>
<td>0.3498</td>
<td>--</td>
<td>0.4219</td>
<td>0.2135</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3_3_3_89</td>
<td>chr6:138008255</td>
<td>EXONIC</td>
<td>6</td>
<td>138008255</td>
<td>--</td>
<td>0.8108</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

**C locus**

<table>
<thead>
<tr>
<th>SNP</th>
<th>SNP ID</th>
<th>position</th>
<th>CHR</th>
<th>BP [hg19], GoNL</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICOSLG_81_158</td>
<td>rs113526534</td>
<td>EXONIC</td>
<td>21</td>
<td>45621231</td>
<td>0.1567</td>
</tr>
<tr>
<td>ICOSLG_81_184</td>
<td>chr21:45621257</td>
<td>EXONIC</td>
<td>21</td>
<td>45621257</td>
<td>0.4465</td>
</tr>
<tr>
<td>ICOSLG_81_220</td>
<td>rs62228210</td>
<td>EXONIC</td>
<td>21</td>
<td>45621293</td>
<td>0.4747</td>
</tr>
<tr>
<td>ICOSLG_81_340</td>
<td>chr21:45621413</td>
<td>EXONIC</td>
<td>21</td>
<td>45621413</td>
<td>0.8102</td>
</tr>
<tr>
<td>ICOSLG_81_437</td>
<td>rs188507544</td>
<td>EXONIC</td>
<td>21</td>
<td>45621510</td>
<td>0.1991</td>
</tr>
<tr>
<td>ICOSLG_81_445</td>
<td>rs62228211</td>
<td>EXONIC</td>
<td>21</td>
<td>45621518</td>
<td>0.406</td>
</tr>
<tr>
<td>ICOSLG_81_447</td>
<td>rs14408965</td>
<td>EXONIC</td>
<td>21</td>
<td>45621520</td>
<td>0.5006</td>
</tr>
<tr>
<td>ICOSLG_81_488</td>
<td>rs23290709</td>
<td>EXONIC</td>
<td>21</td>
<td>45621561</td>
<td>0.8132</td>
</tr>
<tr>
<td>ICOSLG_81_706</td>
<td>chr21:45621779</td>
<td>EXONIC</td>
<td>21</td>
<td>45621779</td>
<td>0.4307</td>
</tr>
<tr>
<td>ICOSLG_81_745</td>
<td>rs6518350</td>
<td>EXONIC</td>
<td>21</td>
<td>45621818</td>
<td>0.6161</td>
</tr>
</tbody>
</table>

**Transcripts including exonic variants**

<table>
<thead>
<tr>
<th>Transcripts including exonic variants</th>
<th>TCONS_0003061 (686 bp)</th>
<th>TCONS_0002726 (1,765 bp)</th>
<th>Inc-UBE2E3-1-4 (1,771 bp)</th>
<th>TCONS_0003062 (511 bp)</th>
<th>TCONS_0003063 (342 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value</td>
<td>P-value</td>
<td>P-value</td>
<td>P-value</td>
<td>P-value</td>
<td>P-value</td>
</tr>
</tbody>
</table>

RNAsnp program was used to calculate the P-values using Mode1 for short sequences (< 1,000 bp) and Mode2 for long sequences (> 1,000 bp *).

SNPs causing significant local secondary structure changes in their associated lncRNA with empirical P-value < 0.2 are highlighted grey.
## Supplementary table 2

<table>
<thead>
<tr>
<th>SNP_ID</th>
<th>locus</th>
<th>SNP_ID</th>
<th>Position</th>
<th>Genotype</th>
<th>BP (bp)(% GST)</th>
<th>2-Tail p-value</th>
<th>Linkage SNP ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7588217</td>
<td>rs7588217</td>
<td>rs7588217</td>
<td>intronic</td>
<td>A</td>
<td>3.23 0.85 12.9</td>
<td>37.7 32.9 42.1</td>
<td>rs148553735</td>
</tr>
<tr>
<td>rs74765727</td>
<td>rs74765727</td>
<td>xxx</td>
<td>intronic</td>
<td>G</td>
<td>0.01 0.20 0.74</td>
<td>4.64 4.60 2.2</td>
<td>rs56299324</td>
</tr>
<tr>
<td>rs7933957</td>
<td>rs7933957</td>
<td>rs7933957</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>rs62180198</td>
<td>ns62180198</td>
<td>ns62180198</td>
<td>intronic</td>
<td>T</td>
<td>0.01 0.30 0.7</td>
<td>65.3 65.7 0.0</td>
<td>rs62228201</td>
</tr>
<tr>
<td>rs14700861</td>
<td>ns14700861</td>
<td>ns14700861</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>rs15110674</td>
<td>rs15110674</td>
<td>ns5110674</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs62228201</td>
</tr>
<tr>
<td>rs7580110</td>
<td>rs7580110</td>
<td>rs7580110</td>
<td>intronic</td>
<td>T</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs62228201</td>
</tr>
<tr>
<td>chr2:182091077</td>
<td>chr2:182091077</td>
<td>D</td>
<td>intronic</td>
<td>2</td>
<td>18201077 T T</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
</tr>
<tr>
<td>chr6:137949454</td>
<td>chr6:137949454</td>
<td>D</td>
<td>intronic</td>
<td>6</td>
<td>137949454 T</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
</tr>
<tr>
<td>rs562124</td>
<td>ns562124</td>
<td>rs562124</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>chr6:137949416</td>
<td>chr6:137949416</td>
<td>D</td>
<td>intronic</td>
<td>6</td>
<td>137949416 A T</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
</tr>
<tr>
<td>rs605777</td>
<td>rs605777</td>
<td>rs605777</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>chr6:13795009</td>
<td>chr6:13795009</td>
<td>D</td>
<td>intronic</td>
<td>6</td>
<td>13795009 A C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
</tr>
<tr>
<td>rs625124</td>
<td>rs625124</td>
<td>rs625124</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>rs62180198</td>
<td>ns62180198</td>
<td>ns62180198</td>
<td>intronic</td>
<td>T</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs62228201</td>
</tr>
<tr>
<td>rs7898657</td>
<td>rs7898657</td>
<td>rs7898657</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>rs74765727</td>
<td>rs74765727</td>
<td>xxx</td>
<td>intronic</td>
<td>G</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
<tr>
<td>rs562124</td>
<td>rs625124</td>
<td>ns625124</td>
<td>intronic</td>
<td>C</td>
<td>0.01 0.56 2.1</td>
<td>98.6 99.5 0.0</td>
<td>rs55965762</td>
</tr>
</tbody>
</table>

**Note:**
- SNP_ID: Single Nucleotide Polymorphism ID
- locus: Chromosomal locus
- Position: Chromosomal position
- Genotype: Genotypic distribution
- BP (bp): Base pairs from reference sequence
- 2-Tail p-value: Two-tailed p-value for association analysis
- Linkage SNP ID: ID of the nearest linkage SNP
<table>
<thead>
<tr>
<th>TF binding + DNase peak</th>
<th>Minimal binding + DNase peak</th>
<th>No</th>
<th>Data</th>
<th>TF binding + any motif + DNase peak</th>
<th>Minimal binding + any motif + DNase peak</th>
<th>Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLD, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td>ESDR, PLCT, BLD, FRE, CRVX, GI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RegulomeDB**

<table>
<thead>
<tr>
<th>TF binding + DNase peak</th>
<th>Minimal binding + DNase peak</th>
<th>No</th>
<th>Data</th>
<th>TF binding + any motif + DNase peak</th>
<th>Minimal binding + any motif + DNase peak</th>
<th>Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLD, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HaploReg_v3**

<table>
<thead>
<tr>
<th>TF binding + DNase peak</th>
<th>Minimal binding + DNase peak</th>
<th>No</th>
<th>Data</th>
<th>TF binding + any motif + DNase peak</th>
<th>Minimal binding + any motif + DNase peak</th>
<th>Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLD, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td>ESC, ESER, EPSG, FRM, BLD, BLD, MUS, GI, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESC, MUS, GI, THYMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**THE ROLE OF CELIAC DISEASE LOCI-ASSOCIATED lnc RNA GENES**

137
### Supplementary Table 4.

<table>
<thead>
<tr>
<th>A</th>
<th>Probe Chr</th>
<th>gSNP (Immunochip)</th>
<th>gSNP allele</th>
<th>gSNP (eQTL)</th>
<th>gSNP-PVal effect</th>
<th>cis-eQTL</th>
<th>gSNP-ZMeta</th>
<th>gSNP-FDR</th>
<th>Probe HUGO</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs1018326</td>
<td>C</td>
<td>4,24E-11</td>
<td>-6,60</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>UBE2E3</td>
<td>protein-coding</td>
</tr>
<tr>
<td>2</td>
<td>rs1018326</td>
<td>C</td>
<td>9,22E-08</td>
<td>5,34</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AC104820.2</td>
<td>lineRNA (XLOC_001776)</td>
</tr>
<tr>
<td>21</td>
<td>rs58911644</td>
<td>T</td>
<td>4,99E-05</td>
<td>4,06</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs58911644</td>
<td>T</td>
<td>8,26E-04</td>
<td>3,34</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>Probe Chr</th>
<th>gSNP (Immunochip)</th>
<th>gSNP allele</th>
<th>gSNP (eQTL)</th>
<th>gSNP-PVal effect</th>
<th>cis-eQTL</th>
<th>gSNP-ZMeta</th>
<th>gSNP-FDR</th>
<th>Probe HUGO</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs1018326</td>
<td>C</td>
<td>3,31E-02</td>
<td>-2,13</td>
<td>0,2</td>
<td></td>
<td></td>
<td></td>
<td>UBE2E3</td>
<td>protein-coding</td>
</tr>
<tr>
<td>2</td>
<td>rs1018326</td>
<td>C</td>
<td>9,40E-01</td>
<td>-0,08</td>
<td>0,975</td>
<td></td>
<td></td>
<td></td>
<td>AC104820.2</td>
<td>lineRNA (XLOC_001776)</td>
</tr>
<tr>
<td>21</td>
<td>rs58911644</td>
<td>T</td>
<td>5,68E-02</td>
<td>1,91</td>
<td>0,1</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs58911644</td>
<td>T</td>
<td>2,82E-01</td>
<td>1,08</td>
<td>0,466666667</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>Probe Chr</th>
<th>gSNP (Immunochip)</th>
<th>gSNP allele</th>
<th>gSNP (eQTL)</th>
<th>gSNP-PVal effect</th>
<th>cis-eQTL</th>
<th>gSNP-ZMeta</th>
<th>gSNP-FDR</th>
<th>Probe HUGO</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs7580110</td>
<td>A</td>
<td>2,86E-10</td>
<td>-6,31</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>UBE2E3</td>
<td>protein-coding</td>
</tr>
<tr>
<td>2</td>
<td>rs7580110</td>
<td>A</td>
<td>5,17E-08</td>
<td>-5,45</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AC104820.2</td>
<td>lineRNA (XLOC_001776)</td>
</tr>
<tr>
<td>6</td>
<td>rs78968657</td>
<td>C</td>
<td>1,19E-03</td>
<td>3,24</td>
<td>0,025</td>
<td></td>
<td></td>
<td></td>
<td>RP11-3566.2</td>
<td>lineRNA (XLOC_005848)</td>
</tr>
<tr>
<td>21</td>
<td>rs113526534</td>
<td>A</td>
<td>1,14E-05</td>
<td>4,39</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs113526534</td>
<td>A</td>
<td>1,08E-03</td>
<td>3,27</td>
<td>0,019642857</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs188507544</td>
<td>T</td>
<td>5,19E-04</td>
<td>-3,47</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>C21orf2</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs188507544</td>
<td>T</td>
<td>8,26E-04</td>
<td>-3,34</td>
<td>0,014285714</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs2329709</td>
<td>A</td>
<td>4,87E-07</td>
<td>-5,03</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs2329709</td>
<td>A</td>
<td>6,27E-04</td>
<td>-3,42</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>U6</td>
<td>pseudogene (RNU6-1150P)</td>
</tr>
<tr>
<td>21</td>
<td>rs62228210</td>
<td>C</td>
<td>1,28E-05</td>
<td>4,36</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs62228210</td>
<td>C</td>
<td>1,08E-03</td>
<td>3,27</td>
<td>0,019642857</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>3,40E-06</td>
<td>4,65</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>6,90E-05</td>
<td>3,98</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>1,59E-04</td>
<td>3,78</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs6518350</td>
<td>G</td>
<td>2,29E-05</td>
<td>4,24</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs6518350</td>
<td>G</td>
<td>1,41E-03</td>
<td>3,19</td>
<td>0,0375</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D</th>
<th>Probe Chr</th>
<th>gSNP (Immunochip)</th>
<th>gSNP allele</th>
<th>gSNP (eQTL)</th>
<th>gSNP-PVal effect</th>
<th>cis-eQTL</th>
<th>gSNP-ZMeta</th>
<th>gSNP-FDR</th>
<th>Probe HUGO</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs7580110</td>
<td>A</td>
<td>4,23E-02</td>
<td>-2,03</td>
<td>0,1</td>
<td></td>
<td></td>
<td></td>
<td>UBE2E3</td>
<td>protein-coding</td>
</tr>
<tr>
<td>2</td>
<td>rs7580110</td>
<td>A</td>
<td>9,20E-01</td>
<td>-0,10</td>
<td>0,9625</td>
<td></td>
<td></td>
<td></td>
<td>AC104820.2</td>
<td>lineRNA (XLOC_001776)</td>
</tr>
<tr>
<td>6</td>
<td>rs78968657</td>
<td>C</td>
<td>6,88E-04</td>
<td>3,39</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>RP11-3566.2</td>
<td>lineRNA (XLOC_005848)</td>
</tr>
<tr>
<td>21</td>
<td>rs113526534</td>
<td>A</td>
<td>1,97E-02</td>
<td>2,33</td>
<td>0,05</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs113526534</td>
<td>A</td>
<td>2,82E-01</td>
<td>1,08</td>
<td>0,141666667</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs188507544</td>
<td>T</td>
<td>5,06E-02</td>
<td>-1,96</td>
<td>0,08</td>
<td></td>
<td></td>
<td></td>
<td>C21orf2</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs188507544</td>
<td>T</td>
<td>2,40E-02</td>
<td>-2,26</td>
<td>0,05</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs2329709</td>
<td>A</td>
<td>1,30E-03</td>
<td>-3,22</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs2329709</td>
<td>A</td>
<td>8,22E-01</td>
<td>0,23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>U6</td>
<td>pseudogene (RNU6-1150P)</td>
</tr>
<tr>
<td>21</td>
<td>rs62228210</td>
<td>C</td>
<td>2,10E-02</td>
<td>2,31</td>
<td>0,05</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs62228210</td>
<td>C</td>
<td>2,82E-01</td>
<td>1,08</td>
<td>0,141666667</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>1,84E-02</td>
<td>2,36</td>
<td>0,0333333333</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>2,10E-02</td>
<td>2,31</td>
<td>0,05</td>
<td></td>
<td></td>
<td></td>
<td>ICSLG</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs62228211</td>
<td>T</td>
<td>8,38E-02</td>
<td>1,73</td>
<td>0,1333333333</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>21</td>
<td>rs6518350</td>
<td>G</td>
<td>2,73E-02</td>
<td>2,21</td>
<td>0,05</td>
<td></td>
<td></td>
<td></td>
<td>AP001059.5</td>
<td>antisense lncRNA</td>
</tr>
<tr>
<td>21</td>
<td>rs6518350</td>
<td>G</td>
<td>2,93E-01</td>
<td>1,05</td>
<td>0,466666667</td>
<td></td>
<td></td>
<td></td>
<td>DNMT3L</td>
<td>protein-coding</td>
</tr>
<tr>
<td>LD ((r^2)^{eSNP &amp; gSNP})</td>
<td>eSNP ((\text{variant with the strongest eQTL effect on the same gene-probe s gSNP}))</td>
<td>eSNP</td>
<td>eSNP-PVal</td>
<td>effect (\text{cis-eQTL})</td>
<td>eSNP-ZMeta</td>
<td>eSNP-FDR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.176969025</td>
<td>rs62180119</td>
<td>T</td>
<td>2.77E-23</td>
<td>9,940930826</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.938220056</td>
<td>rs4667118</td>
<td>T</td>
<td>2.48E-08</td>
<td>5,75041797</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.142120609</td>
<td>rs876200</td>
<td>T</td>
<td>1.35E-08</td>
<td>5,679035811</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.169926564</td>
<td>rs8129253</td>
<td>C</td>
<td>6.92E-08</td>
<td>5,393398285</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LD ((r^2)^{eSNP &amp; gSNP})</th>
<th>eSNP ((\text{variant with the strongest eQTL effect on the same gene-probe s gSNP}))</th>
<th>eSNP</th>
<th>eSNP-PVal</th>
<th>effect (\text{cis-eQTL})</th>
<th>eSNP-ZMeta</th>
<th>eSNP-FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.16890317</td>
<td>rs62180119</td>
<td>T</td>
<td>2.77E-23</td>
<td>9,940930826</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.980366933</td>
<td>rs4667118</td>
<td>T</td>
<td>2.48E-08</td>
<td>5,75041797</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.0760-05</td>
<td>rs4548024</td>
<td>C</td>
<td>1.44E-20</td>
<td>9,297484335</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.61306998</td>
<td>rs876200</td>
<td>T</td>
<td>1.35E-08</td>
<td>5,679035811</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.157847831</td>
<td>rs8129253</td>
<td>C</td>
<td>6.92E-08</td>
<td>5,393398285</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.028417795</td>
<td>rs2070573</td>
<td>A</td>
<td>1.20E-45</td>
<td>-14,18133304</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.010019292</td>
<td>rs2847224</td>
<td>A</td>
<td>2.42E-17</td>
<td>-8,471852688</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.045641227</td>
<td>rs2847224</td>
<td>A</td>
<td>2.42E-17</td>
<td>-8,471852688</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.926001499</td>
<td>chr21:45607996</td>
<td>T</td>
<td>1.77E-04</td>
<td>-3,750422899</td>
<td>0,0375</td>
<td></td>
</tr>
<tr>
<td>0.13206998</td>
<td>rs876200</td>
<td>T</td>
<td>1.35E-08</td>
<td>5,679035811</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.157847831</td>
<td>rs8129253</td>
<td>C</td>
<td>6.92E-08</td>
<td>5,393398285</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.166700528</td>
<td>rs876200</td>
<td>T</td>
<td>1.35E-08</td>
<td>5,679035811</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.045891083</td>
<td>rs2847224</td>
<td>A</td>
<td>2.42E-17</td>
<td>-8,471852688</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.136267616</td>
<td>rs8129253</td>
<td>C</td>
<td>6.92E-08</td>
<td>5,393398285</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.128437839</td>
<td>rs876200</td>
<td>T</td>
<td>1.35E-08</td>
<td>5,679035811</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.153403474</td>
<td>rs8129253</td>
<td>C</td>
<td>6.92E-08</td>
<td>5,393398285</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

THE ROLE OF CELIAC DISEASE LOCI-ASSOCIATED lncRNA GENES