CHAPTER 1

General introduction
General Introduction

Celiac disease (CeD) is a common immune-mediated disorder triggered by intake of grain-derived gluten proteins that affects 1-2% of individuals in the western world. Although the precise cause of CeD is unknown, multiple environmental and genetic factors have been found to contribute to the development of this complex disease.

In most CeD patients gluten ingestion triggers a strong immune response that provokes the activation of gluten-specific CD4\(^+\) T cells, production of anti-transglutaminase 2 (TG2) antibodies by B cells and lymphocyte infiltration in the small intestine, and all these processes contribute to the villous atrophy and crypt hyperplasia characteristic of the disease. As a result, CeD patients have a variable degree of small intestinal inflammation and a broad range of manifestations that include diarrhea, abdominal pain and malabsorption. In addition to the classical gastrointestinal symptoms, CeD patients can also present extra-intestinal manifestations such as anemia, osteoporosis, dermatitis herpetiformis and neurological disorders. Because of this wide range of symptoms, it is thought that many cases of CeD go undiagnosed.

Currently, a lifelong gluten-free diet (GFD) is the only available treatment for CeD. Avoidance of gluten contributes to the recovery of the intestinal mucosa, the reduction of levels of anti-TG2 antibodies and eventually to reduction of symptoms. In spite of its benefits, a GFD can be challenging to maintain due to multiple factors, including social restrictions, nutrient deficiencies and high costs. In addition, a small sub-group of patients (1-5%) fail to respond to a GFD, and these individuals are at risk of developing a severe condition called refractory celiac disease (RCD) that is characterized by a remarkable infiltration of intraepithelial cytotoxic T cells (IE-CTLs) with abnormal phenotype. In RCD these intestinal abnormalities persist and can eventually contribute to the development of enteropathy-associated T cell lymphoma.

Although gluten exposure is the most important environmental factor in CeD pathogenesis, recent evidence implicates other environmental factors in disease development. Intestinal viral infections and bacterial microbiota have been linked to CeD as possible environmental triggers. In a recent study Bouziat et al. suggested that reovirus infection induces a proinflammatory response with a concomitant loss of oral tolerance to gluten, while the involvement of the bacterial microbiome has been suggested by studies reporting gut microbiome dysbiosis in CeD patients as compared to healthy individuals. Interestingly, these CeD-associated changes in microbiota composition have been shown to affect the processing of gluten peptides, which may affect gluten presentation to gluten-specific CD4\(^+\) T cells and thereby increasing the inflammatory response. Thus, in addition to gluten, environmental factors such as the gut microbiome and virome may contribute to the environmental component of the risk of developing CeD, although their respective contributions
to CeD development are still unclear\textsuperscript{1,11,13}. In contrast, genetic risk factors have been estimated to contribute approximately 50% of CeD risk, making them the major predisposing factors currently known for CeD\textsuperscript{1,17}.

**Genetic risk factors**

The strongest genetic factor associated to CeD risk is the human leukocyte antigen (HLA) region. More than 90% of CeD patients carry either the HLA-DQ2 (HLA-DQ2.5) or the HLA-DQ8 allele, and these alleles appear to account for up to 40% of the genetic risk of developing the disease\textsuperscript{18}. However, although the absence of these alleles in individuals means they will not get the disease, their presence alone cannot predict who will develop CeD because these alleles are present in approximately 30-40% of the general population. This suggests that, while the HLA-DQ2 and -DQ8 alleles are necessary for CeD development, additional genetic factors are required\textsuperscript{19}. To date, genome-wide association studies (GWAS) have identified 42 non-HLA genetic variants to be associated with CeD (Fig. 1)\textsuperscript{17,20}. Due to the modest effect of each non-HLA variant on overall disease risk, these together account for approximately 15% of heritability\textsuperscript{20}. Interestingly, most of these genetic variants are also shared with other

![Manhattan plot showing the results of association for 39 of 42 non-HLA CeD risk loci. Known loci (black), novel loci (blue) and risk loci with multiple signals (underlined) are depicted. The vertical line represents the genome-wide significant threshold (p value 5x10^{-8}). Adapted from Trynka, G. et al, 2012\textsuperscript{20}.](image-url)
immune-mediated disorders such as rheumatoid arthritis\textsuperscript{21} (RA) and type I diabetes\textsuperscript{22} (T1D), indicating the presence of a common etiological component in immune-mediated complex diseases.

Although GWAS have been very successful in associating genomic loci with disease, there is a need for follow-up studies aimed at pinpointing causal genetic variants (single nucleotide polymorphisms, SNPs) and genes in these loci\textsuperscript{23}. There are still some limitations that prevent successful identification of causal SNPs\textsuperscript{24}. First, due to linkage disequilibrium, many adjacent SNPs are co-inherited and are likely to have a similar association or correlation, which complicates prioritisation\textsuperscript{23}. Second, in only four of the CeD-associated loci – \textit{MMEL1}, \textit{SH2B3}, \textit{NCF2} and \textit{IRAK} – do the SNPs affect protein encoding regions\textsuperscript{25–27}. Most of the risk SNPs associated with CeD (and complex immune-mediated diseases in general) fall in non-coding regions of the genome, including intergenic regions, and their consequence is not understood\textsuperscript{28,29}. In complex diseases it has consistently been found that these variants are enriched in regulatory domains controlling gene expression, including promoter and enhancer elements characterized by regions of open chromatin and specific histone modifications\textsuperscript{30}. This indicates that, rather than altering protein function, risk SNPs associated with immune diseases control the expression of genes encoding proteins or non-coding RNAs.

To overcome the difficulties discussed above, complementary methods are applied to move from SNP associations to the downstream consequences on gene expression and the regulation of biological pathways. Zooming in on GWAS loci via fine-mapping, for instance, can identify smaller regions that encompass smaller groups of variants with the highest probability of causality\textsuperscript{24}. The functional impact of fine-mapped SNPs can then be tested by assessing the correlation between expression and genotype (quantitative trait locus (eQTL) analysis)\textsuperscript{31}; the interaction between GWAS SNPs and other genomic regions (chromatin interaction conformation assays (3C,4C))\textsuperscript{32}; the enrichment-overlap with functional elements such as enhancers and promoters (using data generated by the ENCODE\textsuperscript{33} or Epigenome roadmap projects)\textsuperscript{33} and the potential alterations of transcription factor binding sites\textsuperscript{23}. These approaches have confirmed that CeD-associated SNPs affect gene expression rather than change the amino acid sequence of proteins\textsuperscript{17}. Additionally, application of computational approaches such as gene set enrichment analysis and gene network analysis to genes differentially expressed in CeD can elucidate the biological pathways and tissues where these risk genes play a role in disease pathophysiology\textsuperscript{34}, and the prioritized set of candidate genes that results can then be further validated by \textit{in vitro} and \textit{in vivo} assays\textsuperscript{24}.

\textbf{From SNPs to disease mechanisms}

It has been hypothesized that the intestinal barrier is compromised in CeD\textsuperscript{35}, thereby facilitating the passage of gluten peptides into the lamina propria, where they are deaminated by tissue TG2. This deamination process strongly increases the binding affinity of gluten peptides to HLA-DQ2 or -DQ8 molecules on the surface of antigen presenting cells (APCs) such as
dendritic cells, B cells and macrophages. The gluten peptides presented in the context of HLA-DQ2 or HLA-DQ8 molecules on the surface of APC (dendritic cells, B cells, macrophages). Presentation to gluten-specific CD4+ T cells results in activation and proliferation of these cells, which then release IFN-γ and IL-21. These cytokines provide signals that enhance the cytolytic properties of IE-CTLs and promote the differentiation of B cells towards plasma cells that produce anti-gluten and anti-TG2 antibodies. Some of the environmental factors (gluten, microbiome, infections) that can influence the disease onset are indicated. Cytokines/inflammatory molecules expressed by intestinal epithelial cells are depicted (IL-15, IFN-1), as are intestinal epithelial cells (IEC), intraepithelial cytotoxic lymphocytes (IE-CTLs), antigen presenting cells (APC).

A recently performed genetic study integrated different layers of genomic information, including eQTL analysis, cell-type-specific enhancer enrichment, functional annotation of GWAS SNPs and co-expression analyses. In addition to confirming what was already known about the involvement of the immune system in CeD, this study provided genetic evidence for
the involvement of the adaptive immune system via the IFNγ signaling pathway (although the IFN locus itself has not been associated with CeD) and a role for B cells in CeD. Moreover, the same study prioritized several genes (LPP, C1orf106) in CeD-associated loci that might contribute to decreased intestinal barrier function. Disturbance of intestinal permeability could not only facilitate the passage of gluten into the lamina propria but also that of infectious agents. This would boost the presentation of gluten peptides to CD4+ T cells, resulting in the release of pro-inflammatory signals, and these processes would contribute to a stronger immune response. Although this evidence suggests that barrier dysfunction can contribute to CeD onset, it is still unclear whether this is a primary defect and a cause that contributes directly to disease onset or of it is a consequence of the inflammatory environment in the gut of CeD individuals.

In addition to dysregulation of the adaptive immune system and adaptive cytokines, innate cytokines such as IL-15 and IFN type I (IFN-1) are upregulated in intestinal epithelial cells of CeD patients. These cytokines enhance the cytolytic and proinflammatory properties of CTLs and dendritic cells, respectively. Simultaneously, an induction of stress-induced non-classical major histocompatibility complex class I molecules is observed on the surface of epithelial cells, which are recognized by natural killer receptors expressed on IE-CTLs. This interaction licenses IE-CTLs to kill intestinal epithelial cells, thus contributing to villous atrophy. These findings demonstrate the involvement of both adaptive cytokines (IFNγ and IL-21) and tissue-derived cytokines (IL-15 and IFN-1) in activation of CTLs. To date, little is known about the signaling processes elicited in IE-CTLs by these cytokines.

In the work described in this thesis, we have studied how CeD-associated genetic variation can be translated to molecular culprits (genes, pathways and relevant cell types). One of the most exciting discoveries of genomics research using next generation sequencing methodologies has been the existence of a novel class of genes that are transcribed but not translated: long non-coding RNAs (lncRNAs). Our group found genetic evidence that, in loci associated with

<table>
<thead>
<tr>
<th>Cell types and culture systems</th>
<th>Gene and environmental perturbations</th>
<th>Conventional and novel lab methods</th>
<th>In silico Methods and cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>IELs</td>
<td>SiRNA shRNA</td>
<td>WB qPCR</td>
<td>Bioinformatic analysis</td>
</tr>
<tr>
<td>Biopsy-derived intestinal cells</td>
<td>Cytokine stimulation</td>
<td>Olink</td>
<td>Co-expression Analysis</td>
</tr>
<tr>
<td>T cells</td>
<td>Rescue</td>
<td>Bulk RNA-seq</td>
<td>Available datasets</td>
</tr>
<tr>
<td>Biopsy-derived intestinal cells</td>
<td>Anti CD3/CD28 Stimulation</td>
<td>Single cell RNA-seq</td>
<td>Population Cohorts</td>
</tr>
<tr>
<td>2D/3D Cultures</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>PBMCs</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>Cell lines</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>Biopsy-derived intestinal cells</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>2D/3D Cultures</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>PBMCs</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>Cell lines</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>Biopsy-derived intestinal cells</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
<tr>
<td>2D/3D Cultures</td>
<td></td>
<td>Bulk RNA-seq</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Overview of the cell types and methods used in this thesis. A) Human-derived primary and immortalized cell lines and culture systems (2D and 3D). B) Tools to target genes and mimic inflammation. C) Lab approaches. D) Data analysis strategies, in silico tools, datasets and population cohorts employed to conduct this study.
complex diseases including CeD, lncRNAs are affected by disease-associated SNPs\textsuperscript{48}. Thus, it is not only non-coding regulatory elements, but also lncRNAs, that may contribute to the deregulation caused by CeD-associated variants.

In the research leading to this thesis I made use of \textit{in silico} approaches in combination with \textit{in vitro} experiments in human-derived cells exposed to conditions that resemble the inflammatory environment in the intestine of CeD patients in order to prioritize genes and lncRNAs potentially involved in CeD and to further understand their role in disease onset and pathogenesis (Fig. 3).

\textbf{Aim and outline of the thesis}

The aim of this thesis is to gain further insights into the function of some of the CeD candidate genes and the molecular pathways and mechanisms that play a role in CeD. The candidate genes have been selected by \textit{in silico} and omics approaches in cell types known to play a role in CeD pathophysiology.

As was described above, it has been hypothesized that CeD patients have an inherent barrier function defect that could facilitate gluten transport into the lamina propria. Previously, \textit{LPP} has been identified as one of multiple CeD-associated genes that may be involved in cell-cell interaction and intestinal barrier homeostasis. In CeD biopsies, \textit{LPP} expression is reduced when compared to normal biopsies. In \textbf{Chapter 2}, we examined whether a reduction in \textit{LPP} does lead to decreased barrier homeostasis using the Caco-2 cell line, which is widely used in barrier function and pharmacology research, as an \textit{in vitro} model. We generated a stable \textit{LPP} knockdown cell line of the parental Caco-2 cell line and evaluated the effect on proliferative capacity, permeability and lumen formation in 2D and 3D culture environments. Moreover, we evaluated the transcriptional response in this cell line under standard culture conditions and upon challenge with IFN\textsubscript{r}, a cytokine known to be involved in the pathogenesis of CeD.

During the course of my thesis research, population cohort studies became available that could be used for eQTL analysis. Additionally, novel statistical genetics approaches were developed in our lab or published by others that could be applied to the data generated in the cohort studies to prioritize culprit SNPs and genes in disease-associated loci. In \textbf{Chapter 3} we applied a systematic approach to integrate eQTL data from the BIOS cohort (total RNA transcriptomics from whole blood of 4000 participants from general population cohorts)\textsuperscript{49} with CeD association data derived from the most recent CeD GWAS meta-analysis\textsuperscript{50}. We applied four different \textit{in silico} approaches (LD-overlap, Bayesian co-localization, Mendelian randomization and DEPICT) to prioritize potential causal genes, resulting in the identification of 126 positional and functional candidate genes. Co-expression and pathway analysis were applied to prioritize the main cell types and biological pathways in which these genes are most likely to play a role. \textit{TRAFD1}, one of the prioritized genes, was selected for functional follow-up.
The pro-inflammatory response elicited by gluten in CeD patients ultimately converges on IE-CTLs that are consequently licensed to kill epithelial cells. Although several key CeD-associated cytokines are known to affect IE-CTLs, little is known about the transcriptional programs triggered by these cytokines. In Chapter 4 we generated TCRαβ⁺ CD8⁺ cytotoxic T cell lines from human small intestinal epithelium to study the dynamics of the transcriptional response and of genome-wide H3K27 acetylation (H3K27ac) changes in response to stimulation with tissue-derived cytokines, also known as alarmins (IFNβ and IL-15) or a T cell–derived cytokine (IL-21). These three cytokines have not only been associated with tissue destruction in CeD, but also with other autoimmune disorders with tissue-specificities such as RA, inflammatory bowel disease (IBD) and T1D. The data we generated was analyzed in depth to describe the biological pathways that are triggered in IE-CTLs in response to tissue-derived cytokines versus T cell–derived cytokines. We further studied the relation between gene expression and epigenomic (H3K27ac) profiles to get a better understanding of the potential mechanism of gene regulation. Finally, we tested the potential enrichment of genes responding to cytokine stimulation in risk loci associated with autoimmune diseases (GWAS data) to identify genes that might contribute to immune deregulation in IE-CTLs.

The next generation sequencing and genomics revolution has led to the discovery of novel classes of genes and novel insights into disease biology. A significant portion of the SNPs associated with complex immune-mediated diseases have been shown to overlap with DNA motifs that control the expression or binding sites of micro-RNAs (miRNAs) or to intersect gene regulatory motifs of IncRNAs. Chapter 5 is an overview of the general features of these two classes of non-coding RNAs in terms of synthesis, structure and the potential mechanism by which they modulate gene expression or other processes in the cell. This review focusses on the role of these genes in different cells of the adaptive and the innate immune system, and on their involvement in immune mediated disorders such as CeD, IBD and multiple sclerosis. Chapter 6 focuses on a single IncRNA, IncRNA RP11-291B21.2, that is strongly modulated in response to TCR activation in CD8⁺ T cells, including IE-CTLs. We describe the expression pattern of this IncRNA in different CD8⁺ T cell populations derived from blood and infer its potential biological function using single cell RNA-seq data and co-expression network analysis. Finally, knockdown experiments performed in IE-CTL cell lines and RNA sequencing data were interrogated to provide additional clues for pinpointing this IncRNA’s function in IE-CTLs, which are the effector cell type in CeD.

Chapter 7 summarizes the major findings of this thesis research project and sets them in a broader perspective by discussing their implications in the context of CeD and immune-mediated disorders in general. The main drawbacks and limitations of our experimental approaches are described, as are directions and suggestions for how to dig deeper into the biological contributions of the risk genes in the pathogenesis of complex diseases.
References


31. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. Plos Genet. 6, e1000888 (2010).


1Department of Genetics, University Medical Center Groningen, University of Groningen, the Netherlands.
2Department of Cell Biology, University Medical Center Groningen, University of Groningen, the Netherlands.
3Depts. of Gastroenterology, Infectious Diseases and Rheumatology, Charité–Universitätsmedizin, Berlin, Germany.
4Institute of Anatomy, University of Jena, Jena, Germany.