

Linköping University Medical Dissertation No. 1672

Biomarkers of Inflammation and Intestinal Mucosa Pathology in Celiac Disease

Hanna Gustafsson Bragde

Linköping University Medical Dissertation
No. 1672

Biomarkers of Inflammation and Intestinal Mucosa Pathology in Celiac Disease

Hanna Gustafsson Bragde



Laboratory medicine, Jönköping, Region Jönköping County

Linköping University
Department of Clinical and Experimental Medicine
SE-581 83 Linköping
Sweden

Linköping 2019

© Hanna Gustafsson Bragde, 2019

Published articles have been reprinted with the permission of the copyright holders

Printed in Sweden by LiU-Tryck, Linköping 2019

ISSN 0345-0082

ISBN 978-91-7685-105-0

”Den som planterat ett träd har inte levt förgäves”

Kinesiskt ordspråk

Abstract

Celiac disease (CD) is a chronic small intestinal immune-mediated enteropathy triggered by gluten. The only currently available treatment is complying with a lifelong gluten-free diet, which should not be commenced before a CD diagnosis has been established by diagnostic test results (including histopathologic assessment of small intestinal biopsies and CD-specific antibody levels). This makes diagnostic swiftness and accuracy important. In cases with low CD-specific antibody levels and/or low-grade intestinal injuries the diagnosis can be difficult to establish. The main objective of this thesis was to complement and improve CD diagnostics by identifying and implementing new biomarkers, mainly based on gene expression, in small intestinal biopsies and blood. In paper I, genes were selected to reflect villous height, crypt elongation, immune response, and epithelial integrity. The results showed that a subset of those genes could discriminate active CD mucosa from mucosa without CD-related changes and grade the intestinal injury. In paper III, an unbiased investigation of gene expression in CD mucosa was performed using transcriptome analysis. Active CD and non-CD mucosa showed differential expression in a subset of genes, and some were differentially expressed in CD mucosa before histopathologic assessment could confirm intestinal alterations compatible with a CD diagnosis. Gene set analysis revealed that there are many biological processes affected in CD mucosa, including those associated with immune response, microbial infection, phagocytosis, intestinal barrier function, metabolism, and transportation.

In parallel, gene expression was investigated in stabilised whole blood. Blood is a more accessible sampling material than intestinal biopsies, and stabilised blood is suitable for routine diagnostics since transcript levels are preserved at sampling. In paper II, expressions from a selection of genes were quantified in stabilised whole blood (RNA) and/or plasma (protein). Three genes with differential expression in CD were identified. Compared to the CD-specific autoantibodies against tissue transglutaminase (anti-TG2) alone, the addition of the information from the new potential markers resulted in a non-significant contribution to the diagnostic capacity of anti-TG2. An unbiased investigation using transcriptome analysis (paper IV) showed that gene level expression differences in stabilised whole blood were small between CD and non-CD. However, expression differences on a gene set level could potentially be used in CD diagnostics. CD-associated biological processes suggested

by the results included a pro-inflammatory response, negative regulation of viral replication, proliferation, differentiation, cell migration, cell survival, translation, and haemostasis.

Expression analysis using real-time polymerase chain reaction (PCR) is easy to perform, with instrumentation available at most clinical laboratories. Although select solitary biomarkers could be very useful in the diagnosis of CD, basing gene expression profiles on pathway information instead of single genes might also disclose disease heterogeneity between patients and add stability to a diagnostic method based on gene expressions.

In conclusion, the results of this work demonstrate that analysing the expression of a few small intestinal genes can complement CD diagnostics. The application of gene expression analysis in cases with minor small intestine histopathological changes shows promising results, but needs further investigations. Additionally, gene expressions in other inflammatory diseases of the small intestine need to be investigated and compared with CD to complete the picture. Moreover, the findings from this work give clues about the biological contexts in which CD resides, and the potential of gene expression in blood at a gene set level is of interest for further investigations.

Biomarkörer vid celiaki

Glutenintolerans (celiaki) är en kronisk sjukdom där gluten som finns i vete, råg och korn ger inflammation och skadar tunntarmen. Personer med celiaki bör undvika gluten för att tarmen ska läka, men prover som visar att det är celiaki måste tas innan start med glutenfri kost. Därför är det viktigt att provtagning och analys kan utföras snabbt och ge säkra svar.

I dagens celiakidiagnostik mäter man antikroppar i blodet och en patolog kan titta på prover från tunntarmsväggen för att hitta skador som vid celiaki. Ibland är tarmproverna svåra att bedöma eller antikroppsnivåerna låga, och då kan det vara svårt att ställa en säker diagnos. I ett försök att komplettera och förbättra diagnostiken vid celiaki undersökte vi om nivåer av proteiner och/eller ribonukleinsyra (RNA, ett mellansteg mellan DNA och protein) i tunntarmsprover och blod kunde användas som diagnostisk metod.

Resultat från studien

RNA-nivåer i tunntarmen kunde spegla viktiga delar av patologens bedömning av tunntarmsproverna, och nivåerna varierade med graden av skada i tarmen. Analys av tarmprover från individer med celiaki och aktiv sjukdom och de utan celiaki visade att det fanns många RNA med olika nivåer mellan dessa grupper. För några RNA kunde vi se en nivåskillnad redan innan de typiska celiakirelaterade tarmskadorna kunde fastställas av patologen. Det resultatet är dock baserat på mycket få individer (4 st) och måste undersökas i en större grupp innan några slutsatser kan dras. Analysens utformning gav även möjlighet att titta på vilka biologiska processer i tunntarmen som påverkades vid celiaki, och resultaten inkluderade bl. a. inflammation, infektion, tarmens barriärfunktion och metabolism.

Vi undersökte även RNA och proteinnivåer i blod. Blod innebär en enklare provtagning än en tunntarmsbiopsi vilket gör det lättare att använda analysen både för diagnos och för uppföljning av celiaki. Resultaten från analysen visade att inga enskilda RNA eller proteiner som vi mätte kunde matcha förmågan hos de antikroppar som idag används för celiakidiagnostik. Resultaten visade dock att biologiska processer som t ex inflammation, negativ kontroll av virusförökning, blodstillning, samt blodcellers tillväxt, förflyttning, specialisering och överlevnad kan

vara påverkade vid celiaki. Att analysera för skillnader i biologiska processer istället för enstaka RNA kan ge ökad stabilitet till en diagnostisk metod.

Slutsats

Resultaten från studien visar att analys av ett fåtal RNA i tunntarmsprover kan användas som ett tillägg till dagens diagnostiska metoder för celiaki. Vi har dock ännu inte undersökt om RNA-nivåerna i tunntarm hos individer med celiaki är lika eller skiljer sig från individer med andra sjukdomar i tarmen. Detta behöver undersökas vidare, liksom potentialen för analys av RNA i blod baserat på biologiska processer.

List of publications

- I. Bragde H, Jansson U, Jarlsfelt I, Soderman J. Gene expression profiling of duodenal biopsies discriminates celiac disease mucosa from normal mucosa. *Pediatric research* 2011;69(6):530-7.
- II. Bragde H, Jansson U, Fredrikson M, Grodzinsky E, Soderman J. Potential blood-based markers of celiac disease. *BMC gastroenterology* 2014;14:176.
- III. Bragde H, Jansson U, Fredrikson M, Grodzinsky E, Soderman J. Celiac disease biomarkers identified by transcriptome analysis of small intestinal biopsies. *Cell Mol Life Sci* 2018;75(23):4385-401.
- IV. Bragde H, Jansson U, Fredrikson M, Grodzinsky E, Soderman J. Characterization of gene and pathway expression in stabilized blood from children with celiac disease. *Manuscript*.



Abbreviations

ANOVA	analysis of variance
anti-DG	antibodies against deamidated gliadin
anti-TG2	autoantibodies against tissue transglutaminase
APOC3	apolipoprotein C3
AUC	area under the curve
bp	base pairs
CD	celiac disease
CD163	CD163 molecule
CDKN1B	cyclin dependent kinase inhibitor 1B
cDNA	complementary DNA
cpm	counts per million
C _q	quantification cycle
CTLA4	cytotoxic T-lymphocyte associated protein 4
CXCL10	C-X-C motif chemokine ligand 10
CXCL11	C-X-C motif chemokine ligand 11
CYP3A4	cytochrome P450 family 3 subfamily A member 4
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
EIF2B1	eukaryotic translation initiation factor 2B subunit alpha
ELISA	enzyme-linked immunosorbent assay
EMA	endomysium
eQTL	expression quantitative trait loci
ESPGHAN	the European Society for Paediatric Gastroenterology Hepatology and Nutrition
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
GBP5	guanylate binding protein 5
GD	gluten-containing diet
GFD	gluten-free diet
GO	Gene Ontology
GSEA	gene set enrichment analysis
GWAS	genome-wide association study

HDEG	highly differentially expressed gene
HLA	human leukocyte antigen
HPRT1	hypoxanthine phosphoribosyltransferase 1
IBD	inflammatory bowel disease
IEL	intraepithelial lymphocyte
IFI27	interferon alpha inducible protein 27
IFN	interferon
IFNG	interferon gamma
Ig	immunoglobulin
IL	interleukin
IL17A	interleukin 17A
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAD2L1	MAD2 mitotic arrest deficient 2 like 1
MHC	major histocompatibility region
MKI67	marker of proliferation Ki-67
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
OCLN	occludin
PBMC	peripheral blood lymphocyte
PCA	principal component analysis
PCR	polymerase chain reaction
PGK1	phosphoglycerate kinase 1
RNA	ribonucleic acid
ROC	receiver operating characteristic
RPKM	reads per kilobase per million mapped reads
SNP	single nucleotide polymorphism
TE	Tris-EDTA
TG2	tissue transglutaminase
TNFRSF9	TNF receptor superfamily member 9
TNFSF13B	TNF superfamily member 13b
Tris	tris(hydroxymethyl)aminomethane
UBD	ubiquitin D
ULN	upper limit of normal
ZFR	zinc finger RNA binding protein

Contents

Introduction	- 1 -
Celiac disease	- 3 -
Epidemiology	- 3 -
Genetics	- 4 -
Environmental factors	- 6 -
Effects of celiac disease in the small intestine.....	- 6 -
<i>Intestinal mucosa</i>	- 7 -
<i>Gluten</i>	- 8 -
<i>Pathogenesis</i>	- 9 -
Effects of celiac disease in blood.....	- 10 -
Extra-intestinal manifestations of celiac disease	- 11 -
Celiac disease diagnostics, treatment, and follow-up	- 11 -
<i>Histopathology</i>	- 11 -
<i>Antibody levels</i>	- 12 -
<i>HLA typing</i>	- 13 -
<i>Current recommended diagnostic flow</i>	- 13 -
<i>Celiac disease treatment</i>	- 14 -
<i>Celiac disease follow-up</i>	- 14 -
<i>Conditions with intestinal symptoms similar to celiac disease</i>	- 14 -
Gene expression	- 15 -
Biomarkers for celiac disease	- 16 -
Methods — general overview	- 21 -
RNA and DNA isolation	- 21 -
Real-time PCR and TaqMan technology	- 21 -
<i>Relative quantitative real-time PCR</i>	- 22 -
Massive parallel sequencing	- 23 -
<i>RNA sequencing</i>	- 25 -
Enzyme-linked immunosorbent assay	- 26 -
Luminex xMAP technology.....	- 26 -
Genotyping	- 27 -

<i>Genotyping using a sequence-specific primer PCR method</i>	- 27 -
<i>Genotyping using real-time PCR and TaqMan probes</i>	- 27 -
Thesis aims	- 29 -
Methods – study specific	- 31 -
Study subjects and groups	- 31 -
Biopsy and blood sampling.....	- 33 -
Statistical methods and tools.....	- 34 -
Paper I	- 35 -
<i>Sample preparations</i>	- 36 -
<i>Gene expression analysis</i>	- 36 -
<i>Statistical analysis</i>	- 36 -
Paper II.....	- 37 -
<i>Sample preparations</i>	- 37 -
<i>Gene expression analysis</i>	- 38 -
<i>Protein quantification</i>	- 38 -
<i>Genotyping</i>	- 38 -
<i>Statistical analysis</i>	- 38 -
Paper III	- 39 -
<i>Sample preparations</i>	- 39 -
<i>Gene expression analysis</i>	- 39 -
<i>Genotyping</i>	- 40 -
<i>Statistical analysis</i>	- 40 -
Paper IV.....	- 41 -
<i>Sample preparations</i>	- 41 -
<i>Gene expression analysis</i>	- 41 -
<i>Genotyping</i>	- 41 -
<i>Statistical analysis</i>	- 41 -
Results and analysis	- 43 -
Papers I and III – Gene expression analysis in small intestinal biopsies....	- 43 -
Papers II and IV – RNA and protein levels in blood.....	- 46 -
Discussion	- 47 -
Conclusions	- 51 -

Thesis contribution and future perspectives..... - 53 -
Acknowledgements..... - 55 -
References..... - 57 -



Introduction

Celiac disease (CD) is an immune-mediated systemic disorder with gluten as the trigger, and is associated with various combinations of clinical manifestations, small intestinal enteropathy with reduced villous height, crypt hyperplasia, infiltration of the epithelial layer by intraepithelial lymphocytes (IELs), and CD-specific antibodies¹. CD diagnostics include detection of CD-specific antibodies in the blood, histopathologic assessment of small intestine characteristics, and typing of genetic predisposing factors (human leukocyte antigen [HLA]-DQ2 and DQ8). Although patients may be asymptomatic, characteristic CD symptoms include *e.g.*, diarrhoea, chronic abdominal pain, malabsorption, iron-deficiency anaemia, and stunted growth¹.

When this project was initiated in 2009, the diagnostic procedure for CD always included a histopathologic assessment of intestinal biopsies performed by a pathologist². Assessments performed on the same small intestinal biopsy specimens by different pathologists do not always agree^{3,4}, and they can be hindered by technical difficulties such as suboptimal orientation of biopsies^{5,6}. Moreover, minor intestinal alterations that can be associated with other pathologies⁷ can complicate CD diagnostics. Between 1973 and 2009, there was an average annual increase in CD incidence of 4% in Swedish children⁸. Everything considered we saw a need for a more accessible analysis of CD characteristics in intestinal biopsies that could be performed using equipment available at most clinical laboratories, reflecting both important characteristics of the histopathologic assessment and other aspects of CD pathogenesis. Therefore, we wanted to investigate the utility of measuring gene expression levels in intestinal biopsies as a CD diagnostic tool. Gene expression profiling shows potential as a robust test for classification purposes with high interlaboratory reproducibility^{9,10}, and gene expression analysis is currently used in clinical diagnostics¹¹. The initial idea was to reflect important elements in the histopathologic assessment using the expression of a selection of genes. In search of a method to reduce the need for an intestinal biopsy in CD diagnostics, we also investigated whether gene expression (RNA/protein) in blood could be used as a complement in the diagnostic procedure for CD and in the follow-up of patients on a gluten-free diet (GFD). CD-specific blood-based antibodies do not strongly correlate with mucosal healing, possibly due to their long half-life and the fact that they reflect the immune response rather than direct intestinal damage¹². However, the analysis of

a few genes gives limited information compared to an unbiased investigation of gene expression using ribonucleic acid (RNA) sequencing. As the project progressed, characterisation of all purified RNA from CD mucosa and blood was performed to investigate the involvement of CD in different biological processes and find potential CD biomarkers. Because the CD population is genetically heterogenic and under the influence of gluten and possibly other environmental factors¹³, it could be difficult to reflect the disease spectrum using a few biomarkers. The characterisation of gene expressions with an unbiased method like RNA sequencing could offer insights into CD pathogenesis, provide information to guide the development of CD biomarker panels, and by extension suggest pharmacological targets in the treatment of CD.

Celiac disease

CD can be defined as a 'chronic small intestinal immune-mediated enteropathy precipitated by exposure to dietary gluten in genetically predisposed individuals'¹⁴. It manifests with symptoms such as diarrhoea, abdominal pain, and malabsorption, but also by more atypical symptoms such as iron deficiency, osteoporosis, short stature, and infertility¹⁵. Sometimes there are no recognisable symptoms. CD was described as early as 1888 by Samuel Gee¹⁶, and the possibility to treat the disease by excluding wheat was suggested by Dicke *et al.* in 1953¹⁷. The associations of active CD with antibodies directed against gliadin¹⁸ and endomysium (EMA)¹⁹, and autoantibodies against tissue transglutaminase (anti-TG2)²⁰ were established in the second half of the 20th century.

In CD-affected individuals, gluten ingestion unleashes an inflammatory response. Events associated with active CD usually include production of antibodies against deamidated gliadin (anti-DG), EMA, and anti-TG2; an inflamed small intestine with reduced villous height, crypt hyperplasia, and IEL infiltration of the epithelial layer; and a diverse range of symptoms¹. CD diagnostics are based on histopathologic assessment of small intestine characteristics, detection of CD-specific antibodies in the blood, and genotyping of the predisposing HLA-DQ2 and DQ8. The genetic profile is important in the risk of developing CD, as indicated by the concordance rate in monozygotic twins, which was estimated at 0.49 in a recent study including 107,000 twins²¹. The concordance rate for dizygotic twins was estimated at 0.10.

Epidemiology

In Europe the prevalence of CD is ~1% (with large inter-country variability)²², but can reach as high as 3%, as reported in a cohort of 12-year-old Swedish children born in 1993²³. This birth year is included in the 'Swedish CD epidemic', which happened during 1984 to 1994, when the CD incidence in Sweden increased from an average of 65 cases per 100,000 person years to 198 cases per 100,000 person years in children younger than 2²⁴. In a study performed in the Swedish county of Östergötland, the incidence rate in children younger than 2 years old was 301 new cases per 100,000 person years in 1994²⁵. After this period, the CD incidence in Sweden decreased to approximately the same levels as before the epidemic²⁴. Additionally, a prevalence as high as 1/256 could be calculated for healthy adult blood donors based on a 1992 screening study by Grodzinsky *et al.*²⁶. The higher CD incidence in the 'Swedish CD

epidemic' was partly attributed to a change in national recommendations regarding gluten introduction and consumption²⁴. Recent studies indicate that neither delayed introduction of gluten nor breastfeeding has any effect on CD risk among at-risk infants^{27,28}. However, a later introduction of gluten delays disease onset²⁷. Recommendations from ESPGHAN (the European Society for Paediatric Gastroenterology Hepatology and Nutrition) regarding gluten introduction and risk for CD in children were published in 2016. They state that gluten may be introduced anytime between 4 and 12 months of age, but large amounts of gluten should be avoided in the first weeks after gluten introduction and during infancy, since amount of ingested gluten may be associated with CD risk²⁹.

The average CD incidence has increased by about 4% every year in Swedish children between 0 and 14.9 years of age from 1973 to 2009, from ~10 cases per 100,000 person years in 1973 to 42 cases per 100,000 person years in 2009, and the median age at diagnosis has increased from 1.0 year in the 1970s to 6.8 years in 2009⁸. CD is an increasingly common diagnosis, and establishment of a CD diagnosis is more common for females than males³⁰. However, the difference seems less pronounced in screening-based studies^{31,32}. More recent and local data on incidence rate were presented in a study including children in the Swedish county of Östergötland²⁵. The 2013 rate was 50 new cases per 100,000 person years for children and adolescents below 18 years of age. CD prevalence is higher in children with conditions such as type I diabetes, Down syndrome, and autoimmune thyroid disease³³.

Genetics

All vertebrates possess a major histocompatibility region (MHC), a large multigenic region containing the genes for the MHC class I and II molecules that have considerable allelic polymorphism³⁴. There are three classic MHC class I molecules (HLA-A, -B, and -C) and three MHC class II molecules (HLA-DR, -DQ, and -DP). Both MHC class I and II molecules present protein-derived peptides: MHC class I molecules to CD8⁺ cytotoxic T cells and MHC class II molecules to CD4⁺ T helper cells.

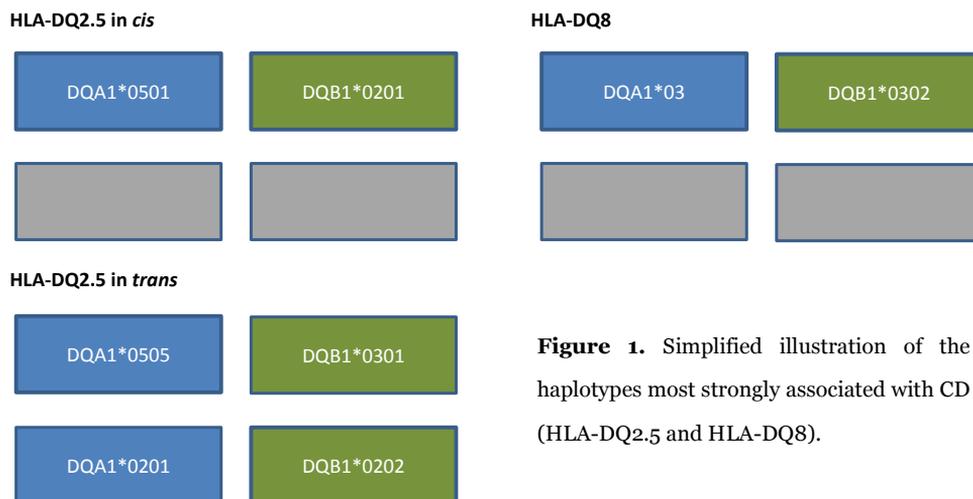


Figure 1. Simplified illustration of the haplotypes most strongly associated with CD (HLA-DQ2.5 and HLA-DQ8).

MHC class II molecules HLA-DQ2 and HLA-DQ8 are necessary, but not sufficient, for CD development^{35,36}. Most CD patients express HLA-DQ2.5, either on the same haplotype (*cis*), as DQA1*0501/DQB1*0201, or in *trans*, with DQA1*0505 on one haplotype and DQB1*0202 on the other³⁷ (Fig. 1). Some express HLA-DQ8 (DQA1*0301/DQB1*0302) (Fig. 1). Although both HLA-DQ2 and -DQ8 are associated with CD, the risk is higher for individuals with the DQA1*05/DQB1*02 haplotype. CD risk gradients based on HLA type have been calculated^{38,39}.

In Western Europe, the allele frequencies of HLA-DQ2 and -DQ8 are ~5–20% and 5–10%, respectively⁴⁰. In a screening study of Swedish 12-year-olds, all children eventually diagnosed with CD carried HLA-DQ2/8 compared with 53% of the non-CD controls³¹. Despite this high general prevalence of HLA-DQ2/8 and the fact that most people are exposed to gluten, not all (~1%) develop CD³⁰.

The HLA locus is estimated to account for close to 40% of CD heritability⁴¹. Genome-wide association studies (GWAS) and extensive follow-up studies^{42–48} have identified a total of 42 non-HLA loci⁴⁹ that contribute to the heritability of CD. Despite these efforts to identify factors responsible for heritability, only ~55% of the genetic risk factors have been identified (HLA and non-HLA loci)⁵⁰. Many of the candidate genes present in non-HLA loci are associated with immune response, and some with intestinal barrier function⁵¹. Most of the identified CD-associated single nucleotide polymorphisms (SNPs) are located in non-exon, intergenic regions⁵¹. The CD risk loci are enriched for SNPs that affect gene expression (expression

quantitative trait loci, eQTL), and could in that way influence CD susceptibility⁴⁵. This is further discussed in the chapter on gene expression (pp. 15–16).

Genetic studies of autoimmune and inflammatory diseases, including CD, indicate that these diseases have common genetic risk factors^{37,52}. Diseases that share risk loci with CD include rheumatoid arthritis, type 1 diabetes, systemic lupus erythematosus, multiple sclerosis, psoriasis, ulcerative colitis, and Crohn's disease³⁷. Some risk loci are shared by many of the diseases, whereas others are more specific for CD⁴⁹.

Environmental factors

Although CD has a strong genetic component, it is a multifactorial disease¹³. Gluten is a known environmental trigger, and it seems that there are individual variations in what amount of gluten can be tolerated by CD-affected individuals⁵³. Gluten is found in food sources such as wheat, rye, and barley⁵⁴. The major wheat protein fractions are gliadin and glutenin, where the proline- and glutamine-rich gliadins are the principal toxic components for CD-affected individuals. The counterparts in rye and barley are called secalins and hordeins, respectively.

The intestinal microbiota seems to play a role in CD pathogenesis⁵⁵, and the interplay between viral or bacterial infections and CD has also been studied⁵⁶. These factors, together with genetics, might influence why some people develop CD in childhood, and some as adults. It has been proposed that a number of factors, which considered separately do not cause CD, when combined in different ways lead to the development of disease^{37,57}. This concept of a spectrum of disease susceptibility considers what HLA genes and non-HLA genetic susceptibility markers the individual carries, as well as exposure to environmental stressors³⁷. Based on these factors, the introduction (and amount) of dietary gluten could provoke CD in infant years, later in life, or not at all.

Effects of celiac disease in the small intestine

Gluten ingestion induces inflammation and alters small intestine mucosa in individuals with CD. The small intestine consists of three sections: the duodenum, jejunum, and ileum⁵⁸. Although CD might affect all parts of the small intestine, CD-associated alterations are mainly found in the first part, the duodenum, and the adjacent jejunum⁵⁹. The duodenal bulb is the most proximal end of the first part of the duodenum⁶⁰. In this first part of the duodenum, the mucosa is flat and smooth,

whereas the more distal parts display Kerckring folds, which are circular or semicircular folds of the small intestine wall.

The small intestine wall consists of five layers: mucosa, submucosa, circular muscularis, longitudinal muscularis, and serosa⁶¹. Kerckring folds increase the small intestine absorption surface, together with projections from the surface termed villi. The first mucosa layer is epithelial cells situated above the lamina propria, which is a loose coat of predominantly reticular connective tissue, within which thin fibres of smooth muscle radiate from the muscularis mucosae (circular muscularis and longitudinal muscularis) to the villi tips. The lamina propria also contains terminal branches of blood vessels, central lacteal or lymph vessels of the villi, and nerve fibres. In the submucosa, which is made up of collagen connective tissue, there are numerous blood and lymphatic vessels. The outermost components are the two muscle layers and the serosa of the small intestine.

In CD diagnostics, mucosa is assessed by a pathologist using sections of intestinal biopsies. The most common way of extracting biopsies is by pinch biopsy, using biopsy forceps during upper endoscopy, although suction capsule biopsies are occasionally performed⁶². Pinch biopsies are about 4 to 8 mm in length and generally include the layers down to the muscularis mucosae.

Intestinal mucosa

The intestinal epithelium is organised into crypts and villi⁶³ (Fig. 2). The crypts of Lieberkühn contain proliferating cells (transit-amplifying cells) that originate from stem cells at the bottom of the crypt. These cells migrate upwards while proliferating and then terminally differentiate. There are four types of differentiated cells in the mucosal epithelium: absorptive enterocytes, mucous-secreting goblet cells, hormone-secreting enteroendocrine cells, and innate immunity-related Paneth cells. The first three cell types cover the villi, while Paneth cells reside at the crypt bottom. Terminally differentiated cells (except for Paneth cells) continue the upward migration, and undergo spontaneous apoptosis as they are shed into the lumen when they reach the villus tip. Protrusions called microvilli are found on the apical side of the fully differentiated enterocyte⁶⁴. These are quite uniform in height with a highly ordered packing and can increase small intestine surface area by 9–16 fold.

A normal small intestinal mucosa has villi with intermittent crypts (villous height to crypt depth ratio 3:1 to 5:1) and only a few IELs (<25 per 100 enterocytes)⁶⁵, but in CD these characteristics are altered. CD-related changes that occur in the small

intestine were graded by Marsh in 1992⁶⁶, and later the modified Marsh scale was proposed by Rostami *et al.*⁶⁷ and Oberhuber *et al.*⁶⁸. The modified scale includes the categories Marsh 0 (normal mucosa); Marsh 1 (IEL infiltration); Marsh 2 (IEL infiltration and crypt elongation); and Marsh 3A, 3B, and 3C (mild, marked, and total villous blunting, respectively, together with IEL infiltration and crypt elongation) (Fig. 2). The lamina propria normally contains a mixture of plasma cells, lymphocytes, and occasional eosinophils and macrophages⁶⁵, but in CD there is a massive influx of inflammatory cells consisting largely of plasma cells and lymphocytes⁶⁹.

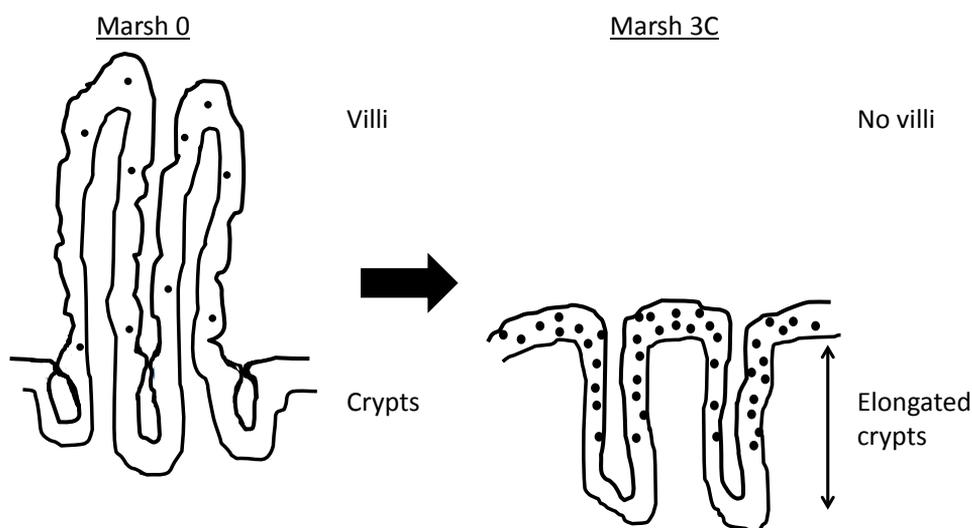


Figure 2. A simplified illustration of the characteristics of Marsh grade 0 (normal mucosa) and Marsh grade 3C mucosa, based on the modified Marsh scale. In Marsh grade 0, there should be villi and crypts with a ratio of 3:1 to 5:1. In Marsh grade 3C, the villi are blunted, and the crypts are elongated. A few IELs (less than 25 per 100 enterocytes) are present in Marsh grade 0, while in Marsh grade 3C there is IEL infiltration.

Gluten

Some ingested peptides are more resistant to hydrolysis by intestinal peptidases, including the proline-rich gliadins in gluten⁷⁰. These peptides have been shown to be suitable substrates for tissue transglutaminase (TG2) deamidation⁷⁰. TG2 is an enzyme found in the normal small intestinal mucosa; it has many functions including stabilisation and remodelling of the extracellular matrix and regulation of cell death, differentiation, and apoptosis⁷¹. It requires calcium binding for activation and

functions by catalysing covalent protein crosslinking and transamidation or deamidation processes. The cross-linked products created with the help of TG2 are highly resistant to mechanical challenge and proteolytic degradation.

Deamidation by TG2 converts non-charged glutamine residues in the gliadin peptides into negatively charged glutamate residues (Fig. 3)⁷². This increases the binding affinity for positively charged pockets in HLA-DQ2 and DQ8 on antigen-presenting cells such as dendritic cells. HLA-DQ2.5 recognises more gluten epitopes than DQ8, which could explain the higher CD risk for individuals with HLA-DQ2.5⁷².

Pathogenesis

The key steps of CD pathogenesis are outlined in Fig. 3. The first line of defence in the duodenum consists of mucus layers, with an underlying single layer of epithelial cells forming the luminal lining⁷³. The epithelial cells are connected by tight junctions and adherens junctions that control the paracellular route over the epithelium⁷³. In CD, the immune response-activating gliadin peptides most likely travel through the epithelial cells by endocytosis⁷⁴ but might also travel the paracellular way⁷⁵. In a study by Lundin *et al.*, HLA-DQ2-restricted gluten-specific CD4⁺ T cells could be induced by gluten stimulation of CD intestinal biopsies⁷⁶. When gliadin peptides are taken up and presented by antigen-presenting cells to gluten-specific CD4⁺ T cells in the context of HLA-DQ, the T cells become pro-inflammatory and home to the small intestinal lamina propria where they serve as effector T cells³⁶. These cells produce pro-inflammatory cytokines that can promote cytotoxic IEL activation.

IELs are made up of a mixture of CD8⁺ $\alpha\beta$ T cells and non-MHC restricted $\alpha\beta$ or $\gamma\delta$ T cells⁷³. The numbers of IELs in the epithelium are increased in active CD⁷⁷, and interleukin (IL)-15 upregulation has been identified in the epithelium and lamina propria of CD patients⁷⁸. Under the influence of IL-15, the IELs express high levels of activating natural killer cell receptors like NKG2D and CD94/NKG2C⁵⁷. The ligands for these receptors, MIC and HLA-E, are upregulated on intestinal epithelial cells in active CD⁵⁷. Interactions between these receptors and ligands lead to IEL-mediated lysis of epithelial cells. Active CD is also characterised by a massive influx of plasma cells into the lamina propria, and anti-TG2 and anti-gluten antibodies are produced³⁷. This response is most likely assisted by the gluten-specific CD4⁺ T cells, but the mechanism is not fully elucidated.

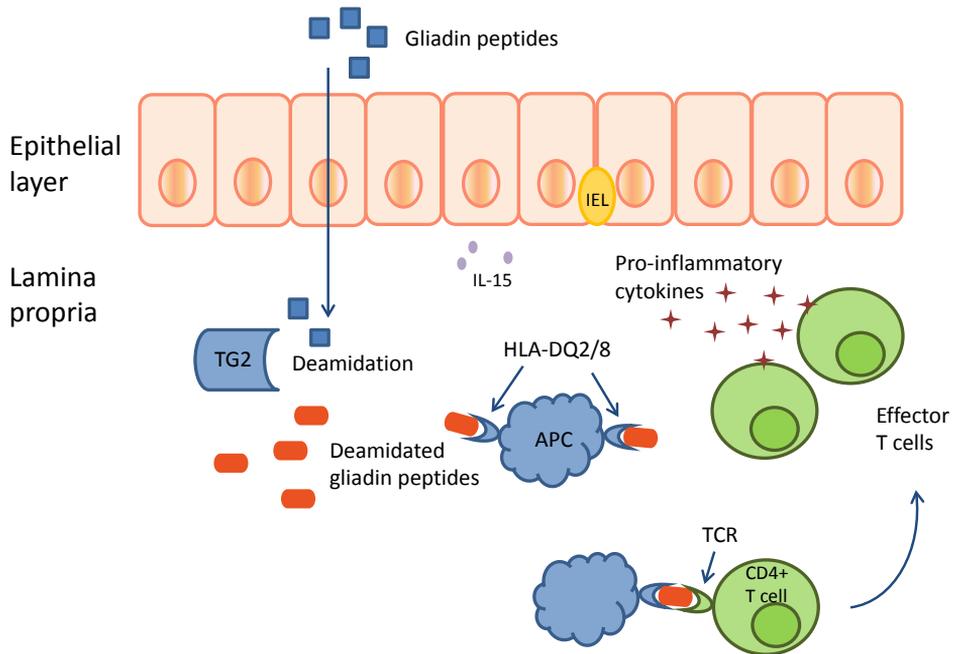


Figure 3. A simplified illustration of key steps in CD pathogenesis.

APC, antigen-presenting cell; IEL, intraepithelial lymphocyte; HLA, human leukocyte antigen; TCR, T cell receptor; TG2, tissue transglutaminase.

Effects of celiac disease in blood

Blood is composed of erythrocytes, platelets, and leukocytes suspended in plasma⁷⁹. The main leukocyte types are neutrophils, lymphocytes, monocytes, eosinophils, and basophils. The ratios between different types of immune cells in the blood can vary between individuals, depending on age, and in disease⁸⁰.

The presence of anti-TG2 in the blood is indicative of CD⁸¹. CD-specific antibodies include anti-TG2, anti-DG, and EMA¹. In 1997, the autoantigen of EMA was identified as TG2⁸².

Gluten challenge of CD patients on a GFD induces gluten-specific HLA-DQ2 restricted CD4⁺ T cells in peripheral blood⁸³.

Haematological manifestations of CD include anaemia, thrombocytosis and thrombocythemia, leukopenia, thromboembolism, and increased bleeding tendency⁸⁴.

Extra-intestinal manifestations of celiac disease

There are many different manifestations of CD in children, and they are not confined to the small intestine⁸⁵. Examples include short stature, delayed puberty, osteoporosis, liver and biliary disease, headaches, behavioural changes, and psychiatric disorders. Additionally, *dermatitis herpetiformis* presents as a skin rash with deposits of immunoglobulin (Ig) A found in the skin⁸⁵. The rash disappears with a GFD. This condition is however rare in children⁸⁵.

Celiac disease diagnostics, treatment, and follow-up

Histopathology

The changes that occur in CD can be evaluated by a pathologist who examines small intestinal biopsy sections and often grades the injury according to the modified Marsh scale, which is recommended by ESPGHAN for grading of histopathologic alterations associated with CD¹. Simplified classifications have been proposed, mainly classifying intestinal biopsies based on IELs and the presence/absence of villous shortening, with a separate group for total absence of villi structures^{3,86}. This approach seems to improve the interobserver reproducibility³. The original Marsh scale did not include a subdivision of Marsh 3⁶⁶, and Marsh *et al.* do not agree with the subdivision of Marsh 3 into A, B, and C⁸⁷.

The ability to make a correct histopathologic assessment can be hindered if too few biopsies are available for interpretation, as this decreases the chances of catching a patchy enteropathy but also lowers the likelihood of finding assessable areas in the biopsies (at least 3 or 4 consecutive villous-crypt units visualised in their entirety and arranged in parallel)⁶⁵. Other factors include poor biopsy specimen orientation, making it difficult to assess villus and crypt status, and Brunner's artefact, where Brunner glands cause artefactual villi distortion. Brunner glands are found in the duodenal submucosa, going through the muscularis mucosae and opening into the crypts⁶⁰. They are more common and denser in the proximal parts of the duodenum, decreasing in size and number toward the distal section.

Interobserver variation between pathologists assessing the same intestinal biopsy specimens has been investigated and showed fair to substantial agreement^{3,4}. Most disagreements were found for Marsh 1–3B, but some also involved normal mucosa³.

Before 1990, three intestinal biopsy samplings were recommended by ESPGAN (now termed ESPGHAN) to confirm gluten sensitivity. First a primary biopsy on

gluten-containing diet (GD), followed by a biopsy taken during GFD to confirm mucosal recovery when gluten had been withdrawn, and finally, a third biopsy on a gluten challenge to confirm that gluten was the factor causing the mucosal injury². In 1990, the criteria were revised and only required one biopsy if the samples showed characteristic CD histopathology and clinical improvement was seen within a few weeks of a strict GFD². The response to a GFD cannot be assessed in asymptomatic patients, so a control biopsy was still considered necessary in those cases.

Biopsies (≥ 4) for the histopathologic assessment in CD diagnostics should be sampled, preferably during upper endoscopy, from the second or third portion of duodenum, and it is recommended to include at least one biopsy from the bulb¹.

Antibody levels

Anti-gliadin antibodies are associated with CD and have been used in CD diagnostics². The sensitivity and specificity of assays for anti-gliadin antibodies, however, were not sufficient to exclude the intestinal biopsy from the diagnostic flow⁸⁸, and anti-gliadin antibodies are not included among those that are currently considered CD specific (anti-TG2, anti-DG, and EMA)¹. Today, detection of IgA anti-TG2 is usually the first step toward a CD diagnosis¹. For studies including only patients with Marsh 3 histopathology, the sensitivity of IgA anti-TG2 is high (94.6%), but if patients with Marsh 1–2 histopathology are included, the sensitivity decreases (88.4%)⁸¹. The anti-TG2 specificities in studies including Marsh 3, or Marsh 1–3 histopathology, are comparable (96.6% and 94.9%, respectively). Raising the cut-off for anti-TG2 improves specificity. The positive predictive value of anti-TG2 levels >10 times the upper limit of normal (ULN) in EMA-positive symptomatic children is very close to 100%, and reaches 100% in children for whom malabsorption is one of the symptoms⁸⁹. The sensitivity of anti-TG2 is decreased in children younger than 18 months⁹⁰. In this age group, anti-gliadin antibody levels could aid in the identification of active CD cases^{90,91}. Selective IgA deficiency is associated with CD⁹². There are, however, IgG-based tests available for the detection of anti-TG2, as well as for anti-DG⁸¹.

Among patients with infection, transient anti-TG2 positivity can be induced independently of gluten⁹³. This can also occur in patients with newly diagnosed type 1 diabetes⁹⁴.

HLA typing

Since HLA-DQ2 or DQ8 are present in almost all CD patients³⁵, but also in many non-CD individuals, HLA-DQ2/DQ8 typing is mainly utilised to exclude CD as a diagnosis¹.

Current recommended diagnostic flow

In the most recent ESGPHAN recommendations, there are two different approaches for the diagnosis of CD: one for symptomatic children and one for asymptomatic children at high risk for CD¹. For the symptomatic children, the diagnosis of CD without a small intestinal biopsy is suggested for those with anti-TG2 levels >10 times ULN since the likelihood of a Marsh grade 3 histopathology is high in these cases. For a CD diagnosis to be established without an intestinal biopsy, the presence of HLA-DQ2 or DQ8 and antibodies against EMA should be verified. Symptoms should also improve on a GFD, further enforcing the diagnosis. However, recent results have shown that analysis of HLA-DQ2/DQ8 presence might not be necessary to establish a non-biopsy proven diagnosis if the other requirements are fulfilled⁸⁹. In symptomatic children with anti-TG2 1–10 times ULN, intestinal biopsies should be sampled and show Marsh grade 2 or 3 to establish a CD diagnosis¹. Marsh 0 or 1 results are classified as unclear cases and warrant further investigation. In HLA-DQ2/DQ8-positive asymptomatic children with positive anti-TG2, the CD diagnosis should always be verified by intestinal biopsies, where Marsh grade 2 or 3 justifies a CD diagnosis and Marsh 0 or 1 constitutes unclear cases. Negative anti-TG2 does not warrant further investigations into a possible CD diagnosis except in cases with *e.g.*, selective IgA deficiency, age below 2 years, low gluten intake, or severe symptoms. In cases with IgA deficiency, levels of IgG anti-TG2, IgG anti-DG, or IgG EMA should be determined.

Since symptoms, antibodies, and CD-associated enteropathy should disappear after starting a GFD, it is of utmost importance not to start a GFD before blood samples, and if needed, intestinal biopsies have been retrieved and analysed and a CD diagnosis has been confirmed¹.

As symptoms of CD can be varying¹⁵, or even absent^{14,26}, the initial assessment can be complex. Furthermore, histopathologic assessment can indicate low-grade intestinal injuries, a state which can be caused by factors other than CD⁷. Additionally, the diagnosis is sometimes hard to determine due to a patchy lesion distribution⁹⁵ or presence of Brunner glands in the small intestinal biopsies, which

may cause artefactual distortion of the villi⁶⁵. On a more technical basis, other difficulties include poor biopsy quality or suboptimal orientation of biopsies prepared for histopathologic assessment^{5,6}. Furthermore, the amount of gluten an individual with CD consumes together with the amount of gluten they can tolerate⁵³ can affect enteropathy and CD-specific antibody levels. Increased levels of CD-specific antibodies can be found without enteropathy¹⁴.

Celiac disease treatment

Currently, the only effective treatment for the majority of CD patients is a strict GFD¹. A few adult-onset CD patients do not respond to a GFD (refractory CD)⁵⁷. Other potential CD treatments are being investigated, and include genetic modification of grains to decrease immunogenicity, co-polymeric binders that keep gluten in the intestinal lumen, hydrolysis of gluten peptides to generate smaller peptides with less immunogenicity, and vaccines to induce tolerance to gluten⁹⁶.

Celiac disease follow-up

The tools used for diagnosis of CD (CD-specific antibodies and intestinal biopsies) are less useful for measuring GFD compliance¹². Using biopsies to monitor diet compliance over time would be impractical, expensive, and invasive, but CD-specific antibodies are not very well correlated with mucosal healing, possibly due to their long half-life and that they reflect the immune response rather than direct intestinal damage¹². A meta-analysis of anti-TG2 and EMA in the follow-up of CD patients on a GFD indicates that the tests have low sensitivity for detecting persistent villous atrophy⁹⁷. Additionally, small or infrequent exposures to gluten are not reflected by serological tests¹². Other potential ways to monitor diet compliance include faecal calprotectin or serum/plasma levels of intestinal fatty acid binding protein or citrulline¹². Measuring gluten immunogenic peptides in faeces or urine is another potential tool that has the added potential to detect occasional gluten exposure¹².

Conditions with intestinal symptoms similar to celiac disease

Conditions other than CD that could cause villous atrophy include autoimmune enteropathy, common variable immunodeficiency, graft-versus-host disease, inflammatory bowel disease (IBD), and tropical sprue, sometimes in combination with increased IEL counts⁶⁵. Causes of increased IEL counts with intact villi include

Helicobacter pylori-associated gastroduodenitis, medications, infections, and immune dysregulation⁶⁵.

There is also a condition termed non-celiac gluten sensitivity where symptoms (intestinal and extra-intestinal) arise if gluten is ingested and cease when gluten is withdrawn, but without CD-typical enteropathy or serology⁹⁸.

Gene expression

Gene expression varies considerably depending on time and place in the body as well as sex, and can be altered in disease⁹⁹. The same regulatory mechanisms are available to almost all cells in the human body, and utilisation of these can create differential expression. The expression of a messenger RNA (mRNA) can be influenced at the level of initiation of transcription but also by posttranscriptional processing⁹⁹. The initiation of transcription involves many factors such as promoters, which are deoxyribonucleic acid (DNA) sequences that guide the RNA polymerase to the correct starting point, and transcription factors, which are proteins that bind the promoter and are necessary for the RNA polymerase to recognise a promoter as a starting point. Transcription initiation involves many other enhancing and suppressing factors to enable fine tuning of gene expression. After transcription, non-coding sequences (introns) present in the produced pre-mRNA are removed in a process termed RNA splicing. This process can yield transcripts with different numbers of exons (alternative splicing), which affects the subsequent protein sequence.

At any given position in the genome, a base can be substituted for another base in a portion of a population. If the frequency of such a single base exchange in a given population is $\geq 1\%$, it qualifies as a SNP¹⁰⁰. The presence of SNPs or other genetic variations in protein-coding sequences can in a few cases lead to differences in protein function, which might induce disease¹⁰⁰. Most of the genetic variation, however, is found outside of the protein-coding sequences¹⁰¹, for example in transcription-regulating sequences such as promoters. As previously mentioned, the identified risk loci in CD are enriched in SNPs that affect gene expression (eQTLs) and are mostly found in non-protein-coding regions. Withoff *et al.*⁴⁹ compiled a list of 112 eQTL genes in 32 non-HLA loci from eQTL studies in CD based on peripheral blood lymphocytes (PBMCs), thymus tissue, monocytes, dendritic cells, and small intestinal biopsies. They advise that interpretation of eQTL data derived from tissues composed of several different cell types should be done with care since eQTLs can exhibit cell- and tissue-specific effects.

Studies of thousands of gene expressions are often performed by microarray or, more recently, by RNA sequencing. Microarrays have been used to characterise the gene expression in CD compared with controls both in whole biopsies¹⁰²⁻¹⁰⁴ and isolated epithelial cells¹⁰⁵. RNA sequencing has been utilised to characterise expression in CD4⁺ T cells from CD patients compared with controls¹⁰⁶. CD-associated pathways identified in studies of the transcriptional changes were compiled in a recent review¹⁰⁷, and the main findings were summarised as an aggravation of the immune response and dysregulation of signalling and cell cycle pathways.

A study of protein expression in CD mucosa using sections of formalin-fixed paraffin-embedded (FFPE) small intestinal biopsies and liquid chromatography-mass spectrometry with label-free protein quantitation¹⁰⁸, comparing the same patients at diagnosis and after a period of GFD, showed enrichment of several immune response processes and a response to endoplasmic reticulum stress in active CD. Enrichment for multiple processes related to nutrient metabolism and enterocyte function were found after treatment with a GFD.

Gene expression is currently used as a clinical diagnostics tool¹¹. RNA sequencing has great potential and could be used to detect gene fusions, differential expression of known disease-causing transcripts, and a diversity of RNA species including regulatory non-coding RNAs¹¹.

Since gene expression varies with cell type, cellular compositions could affect gene expression analysis results in tissue samples such as intestinal biopsies and blood. Cellular compositions could also be altered in disease, for example as a result of inflammatory cell influx into the small intestine of CD patients. This could make it more difficult to understand biological context based on test results, but at the same time might facilitate the separation of samples based on disease/not disease.

Biomarkers for celiac disease

The MeSH terms ‘celiac disease’ and ‘biomarkers’ were queried on PubMed to identify studies of biomarkers in CD over the last 10 years (2009–2018). The studies of biomarkers in blood and small intestinal biopsies listed in Table 1 investigated biomarker levels in CD patients compared to control groups. The purpose was not always to identify a specific marker for CD; some authors were searching for markers for enteropathy, or to follow the effects of GFD in a non-invasive manner. In individuals included in the control groups CD was always excluded as a diagnosis, but

many experienced gastrointestinal symptoms. Other groups used in the comparison with CD patients included *e.g.*, those with autoimmune diseases, Crohn's disease, and infectious enteritis. CD patients on a GFD were often included in the studies for comparison with active CD. The purpose of Table 1 is to provide an overview of the work associated with biomarker discovery in CD over the last 10 years.

Potential biomarkers for CD or for diet monitoring have also been identified in other materials such as urine¹⁰⁹ and faeces¹¹⁰.

Table 1. Examples of studies of potential biomarkers in the small intestine and blood for the diagnosis or follow-up of celiac disease and/or identification of enteropathy. The table includes studies from 2009 to 2018.

Biomarker/s ^a	Analyte	Method of detection ^b	Reference	Group ^c	Children/adults
CTLA-4	Protein in serum	ELISA	Simone <i>et al.</i> ¹¹¹ Italy (2009)	Active CD	Children and adults
T-bet, pSTAT1	Markers on T cells, B cells, and monocytes	Flow cytometry	Frisullo <i>et al.</i> ¹¹² Italy (2009)	Active CD	Adults
I-FABP	Protein in serum	ELISA	Derikx <i>et al.</i> ¹¹³ The Netherlands (2009)	Active CD	Children and adults
Th1-, Th2-, and APC-derived cytokines	Protein in serum	Sandwich immunoassay	Manavalan <i>et al.</i> ¹¹⁴ USA (2009)	Active CD	Adults
Adenosine deaminase	Protein in serum	Enzymatic spectrophotometric method	Cakal <i>et al.</i> ¹¹⁵ Turkey (2010)	Active CD	Adults
GEE: APOC3, CYP3A4, OCLN, MAD2L1, MKI67, CXCL11, IL17A, CTLA4	mRNA in duodenal tissue	Real-time PCR	Bragde <i>et al.</i> ¹¹⁶ Sweden (2011)	Active CD	Children
Reg1 α	Protein in serum	ELISA	Planas <i>et al.</i> ¹¹⁷ Spain (2011)	Active CD	Children and adults
I-FABP	Protein in serum	ELISA	Adriaanse <i>et al.</i> ¹¹⁸ The Netherlands (2013)	Active CD	Adults
Simvastatin metabolism	CYP3A4-mediated metabolism of simvastatin in serum	LC-MS/MS	Morón <i>et al.</i> ¹¹⁹ Schweiz, India, Finland, USA (2013)	Active CD	Adults
PARK7	mRNA and protein in duodenal tissue	Real-time PCR, Western blot	Vörös <i>et al.</i> ¹²⁰ Hungary, Slovakia (2013)	Active CD	Children

Table 1. Continued.

Biomarker/s ^a	Analyte	Method of detection ^b	Reference	Group ^c	Children/adults
Resistin	Protein in serum	ELISA	Russo <i>et al.</i> ¹²¹ Italy (2013)	Active CD	Adults
GEP: LPP, c-REL, KIAA1109, TNFAIP3	mRNA in peripheral blood monocytes	Real-time PCR	Galatola <i>et al.</i> ¹²² Italy (2013)	Active CD	Children
GEP: LPP, c-REL, TNFAIP3, IL-21, RGS1	mRNA in duodenal tissue	Real-time PCR	Galatola <i>et al.</i> ¹²² Italy (2013)	Active CD	Children
CXCL10	Protein in serum	ELISA	Bondar <i>et al.</i> ¹²³ Spain (2014)	Active CD	Children and adults
Citrulline	Amino acid in serum	RF-HPLC	Basso <i>et al.</i> ¹²⁴ Italy (2014)	Active CD	Children
ASCA IgA, IgG	Antibodies in serum	EIA	Vitasalo <i>et al.</i> ¹²⁵ Finland (2014)	Early CD, active CD	Children and adults
Reg3 α	Protein in serum	ELISA	Marafini <i>et al.</i> ¹²⁶ Italy (2014)	Intestinal damage (incl. active CD)	Adults
CXCL11, TNFSF13B	Protein in plasma/ mRNA in blood	Luminex technology, Real-time PCR	Bragde <i>et al.</i> ¹²⁷ Sweden (2014)	Active CD	Children
TNFRSF9	mRNA in blood	Real-time PCR	Bragde <i>et al.</i> ¹²⁷ Sweden (2014)	Normalised CD on GFD	Children
Pronurotensin	Protein in plasma	Chemiluminometric sandwich immunoassay	Montén <i>et al.</i> ¹²⁸ Sweden (2016)	Active CD	Children
Soluble Syndecan-1	Protein in serum	ELISA	Yabecovitch <i>et al.</i> ¹²⁹ Israel, Germany (2017)	Active CD	Children

Table 1. Continued.

Biomarker/s ^a	Analyte	Method of detection ^b	Reference	Group ^c	Children/adults
HMGB1	Protein in serum	ELISA	Manti <i>et al.</i> ¹³⁰ Italy (2017)	Active CD	Children
Alkylresorcinol	Lipid in serum	LC-MS/MS	Choung <i>et al.</i> ¹³¹ USA (2017)	Dietary gluten exposure	Adults
BECN1	mRNA in blood	Real-time PCR	Comincini <i>et al.</i> ¹³² Italy (2017)	Active CD	Children
ATG7, BECN1, miR-17, miR-30a	mRNA and miRNA in duodenal tissue and blood	Real-time PCR	Comincini <i>et al.</i> ¹³² Italy (2017)	Active CD	Children
I-FABP	Protein in plasma	ELISA	Adriaanse <i>et al.</i> ¹³³ The Netherlands (2017)	Active CD	Children
KIAA1109, TAGAP, SH2B3, TNFSF14, RGS1	mRNA in peripheral blood monocytes	Real-time PCR	Galatola <i>et al.</i> ¹³⁴ Italy (2017)	Pre-symptomatic CD	Children
HSP-70, HIF-1 α	mRNA in duodenal tissue	PCR and agarose gel electrophoresis	Piatek-Guziewicz <i>et al.</i> ¹³⁵ Poland (2017)	CD, active CD	Adults

^aAPC, antigen-presenting cell; APOC3, apolipoprotein C-III; ASCA IgA or IgG, anti-Saccharomyces cerevisiae antibodies of type immunoglobulin (Ig) A or IgG; ATG7, autophagy-related 7; BECN1, beclin 1; c-REL, REL proto-oncogene, NF- κ B subunit; CTLA, cytotoxic T-lymphocyte associated protein; CXCL, C-X-C motif chemokine ligand; CYP3A4, cytochrome P450, family 3, subfamily A, polypeptide 4; GEP, gene expression profile; HIF-1 α , hypoxia-inducible factor 1 α ; HMGB1, high mobility group box 1; HSP-70, heat-shock protein 70; I-FABP, intestinal fatty acid binding protein; IL, interleukin; KIAA1109, uncharacterised protein KIAA1109; LPP, LIM domain-containing preferred translocation partner in lipoma; MAD2L1, mitotic arrest deficient 2 like 1; miR, microRNA; MKI67, antigen identified by monoclonal antibody Ki-67; OCLN, occludin; PARK7, Parkinsonism associated deglycase; pSTAT1, phosphorylated signal transducers and activators of transcription 1; Reg, regenerating family member; RGS1, regulator of G-protein signalling 1; SH2B3, SH2B Adaptor Protein 3; TAGAP, T cell activation RhoGTPase activating protein; Th, T helper; TNFAIP3, tumour necrosis factor alpha-induced protein 3; TNFRSF, tumor necrosis factor (TNF) receptor superfamily; TNFSF, TNF superfamily.

^bEIA, Enzyme immunoassay kit; ELISA, enzyme-linked immunosorbent assay; LC-MS/MS, liquid chromatography–tandem mass spectrometry; PCR, polymerase chain reaction; RF-HPLC, reverse-phase high-performance liquid chromatography.

^cGroup with differential levels of the specified analyte compared to controls. CD, celiac disease; GFD, gluten-free diet.

Methods — general overview

RNA and DNA isolation

RNA and DNA isolation can be performed in multiple ways; two common examples are spin filter-based methods and magnetic bead-based methods¹³⁶. Both utilise a work flow of lysis, binding, washing, and eluting. Cells are lysed, and the free RNA/DNA is bound to silica-containing spin filters or silica-coated magnetic beads in the presence of chaotropic salts or alcohols at high concentrations and low pH. The mechanism by which the RNA/DNA is bound involves hydrogen bonds between the RNA/DNA and silica. The RNA/DNA is released (eluted) when salt or alcohol is removed. Both methods are quick and easy to perform. Spin columns usually generate very clean eluates but can clog if the sample is too thick. Magnetic bead-based isolation is suitable for automation. Other methods, less frequently used, are liquid-phase extraction and solid-phase ion exchange.

What method is used can vary, for example depending on the starting material or available instrumentation. Automation is attractive with regard to implementation in routine diagnostics.

Isolated DNA can be refrigerated for months or frozen for years in a buffer such as Tris-EDTA (TE) buffer¹³⁷. RNA can be stored in a similar buffered solution or RNase-free water but is less stable and should be stored at -70°C . As shown by Duale *et al.*¹³⁸, RNA in samples can be preserved by solutions that stabilise mRNA concentrations for expression analysis.

Real-time PCR and TaqMan technology

The process of the polymerase chain reaction (PCR) consists of controlled temperature cycles that make small oligonucleotides (primers) complementary to a specific DNA sequence bind and a DNA polymerase create a copy of the DNA using deoxynucleotide triphosphates (dNTPs)¹³⁹. The number of copies grows exponentially, since each new copy can function as a template in the next temperature cycle. The generated copies can be detected in multiple ways, *e.g.*, by using a dye such as SYBR Green that fluoresces when it binds to double-stranded DNA, or by using more specific probes designed to bind to a sequence within the PCR product and fluoresce when their target sequence is present in the reaction¹⁴⁰. Hydrolysis probes (TaqMan probes) are commonly used in real-time PCR¹⁴⁰. These

probes are constructed with a reporter dye in the 5' end and a quencher in the 3' end, with the purpose to quench the signal from the reporter dye (Fig. 4). As the probe binds to the template during the temperature cycles, the Taq polymerase decomposes the probe, physically separating the reporter and the quencher. As the quencher no longer can stop the reporter signal, detectable fluorescence is emitted. The signal increases exponentially with the number of cycles, and more template in the initial reaction allows faster signal detection. This means that the number of cycles for a detectable signal to reach a designated fluorescence threshold (quantification cycle [C_q]-value) is negatively correlated with the starting number of template copies.

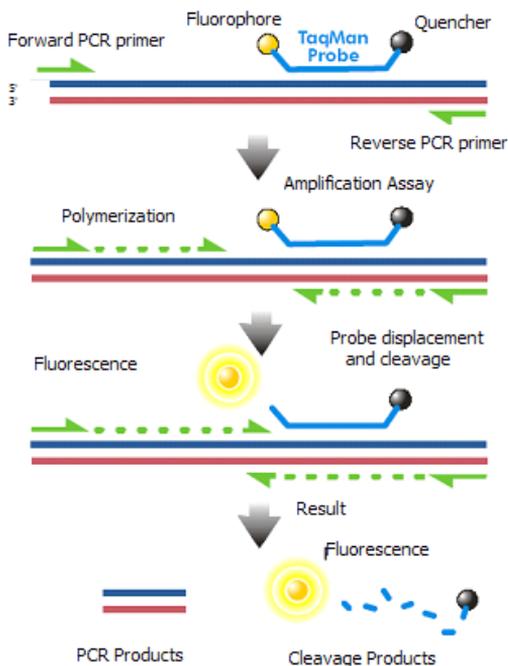


Figure 4. Real-time PCR using TaqMan (hydrolysis) probes.

TaqMan® probe chemistry mechanism' by Braindamaged [Public domain] via Wikimedia Commons.

Relative quantitative real-time PCR

For RNA to function in real-time PCR, it first needs to be converted into complementary DNA (cDNA) using reverse transcription with the enzyme reverse transcriptase⁴⁰. Different starting points can be used for the enzyme, *e.g.*, random primers (short oligonucleotides with random sequences that bind different parts of the RNA) or oligo-dT primers that bind to the poly-A tail of mRNA. Gene-specific primers can also be used. What method is used for reverse transcription can affect

gene expression analysis results¹⁴⁰. For example, oligo-dT primers are convenient to investigate only mRNAs, but if the RNA is somewhat degraded, cDNAs of different length will be created from a certain transcript during reverse transcription. They will all contain the RNA sequence closest to the poly-A tail, but only some include the sequence distal to the poly-A tail¹⁴¹.

In relative quantitative real-time PCR, the levels of transcripts are not quantified in an absolute sense, but rather a relationship between samples in terms of transcript levels is determined¹⁴⁰. Reference genes are used to avoid misinterpreting sample variations introduced during sample preparation and analysis for higher/lower expression of the target genes¹⁴². Genes that are stably expressed in the sample material under the specified conditions are identified and detected in the sample, and the results are subtracted from the target transcript to account for such influences. The evaluation of suitable reference genes is crucial to prevent introducing differences between samples that are merely an effect of a poor choice of reference gene/s rather than a true difference. More than one reference gene is recommended to provide a stable reference C_q -value¹⁴².

After C_q -values have been derived from real-time PCR results, the mean of the C_q -values for the reference genes is subtracted from each target gene C_q -value, and the relative gene expression between samples can be calculated using the Delta-delta C_q method of quantification¹⁴³, which gives a good overview of the data spread and differences between samples and groups.

Real-time PCR is fast, sensitive, and convenient for detecting pre-selected targets. At the same time, it only quantifies what you have pre-defined that you want to measure. Real-time PCR analysis of multiple targets is labour intensive. For a more unbiased approach, for example to investigate potential new biomarkers, RNA sequencing would be more suitable.

Massive parallel sequencing

Massive parallel sequencing, also called next-generation sequencing, became cost-competitive with the much used Sanger sequencing in 2005, and the cost continues to decrease¹⁴⁴. The method is used in clinical practice, for example in cancer diagnostics, and shows great potential as a diagnostic tool in many areas¹⁴⁵.

In massive parallel sequencing, DNA or cDNA is fragmented, and adapters are ligated to the ends of each fragment¹³⁹ before they are sequenced. On the Illumina platform (Illumina, San Diego, CA), bridge amplification and sequencing-by-

synthesis are utilised to sequence the fragments¹⁴⁴. The fragments are loaded onto a flow cell covered in oligos complementary to the adapters. Each fragment is amplified by bridge amplification (Fig. 5), generating clusters of clones from that fragment¹³⁹. Sequencing is then performed, which is a sequencing-by-synthesis procedure that incorporates reversible terminator dNTPs, and each incorporated base is detected by fluorophore excitation¹³⁹. After excitation the dNTP is enzymatically cleaved to remove the terminator, the next dNTP can be added, and the cycle is repeated. To simultaneously sequence multiple samples and still know which sequence was generated from which sample, the fragments can be labelled with different indexes, which enables separation of the sequences by sample in the bioinformatics analysis¹³⁹. The resulting sequenced fragments need first to be divided by sample and then mapped against a reference genome.

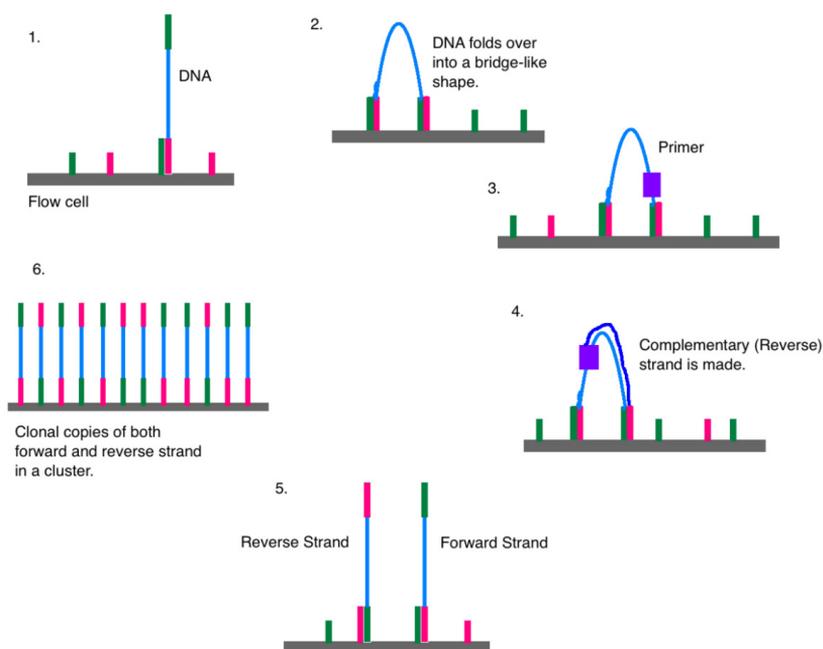


Figure 5. Cluster generation using bridge amplification in massive parallel sequencing.

'Cluster Generation' by DMLapato [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)] via Wikimedia Commons.

RNA sequencing

For RNA sequencing (also called transcriptome sequencing or RNA-seq), mRNA is selected by poly(A) or RNA is depleted of ribosomal RNA¹⁴⁶. Ribosomal RNA makes up more than 90% of total cellular RNA, and therefore would significantly influence sequencing results if not removed¹⁴⁶. The main component of whole blood is erythrocytes, and globin mRNA from progenitor erythrocytes constitutes ~70% of the total mRNA population¹⁴⁷. Globin mRNA and ribosomal RNA depletion must be performed before converting RNA into cDNA for massive parallel sequencing.

The design of an RNA sequencing experiment includes choices of whether to sequence shorter or longer reads, paired- or single-end reads, and whether to retain information regarding strand orientation¹⁴⁶. Single-end short reads are usually sufficient in well-annotated organisms for the purpose of differential gene expression analysis. Strand orientation information is preferable to identify what DNA strand is actually expressed, making it easier to quantify antisense or overlapping transcripts. If the purpose is to identify differential expression of medium or highly expressed genes, fewer reads are necessary as compared to characterising the entire transcriptome¹⁴⁸. Statistical power to detect differential expression varies with effect size (desired fold change), sequencing depth, and replicate number¹⁴⁸. At least three biological replicates in each group should be included to capture natural biological variation, but increasing the number of replicates improves the power of the study more than increasing the number of reads or sequencing depth¹⁴⁸. The expected number of differentially expressed genes also influences the power to detect differential expression¹⁴⁹. Quality control of generated data from RNA sequencing includes examining the sequences to detect low-quality bases, overrepresented sequences *etc*¹⁴⁸. Williams *et al.*¹⁵⁰ stated that trimming of reads (removing low-quality bases) should be done with caution, since no or modest trimming seems to give the most biologically accurate gene expression estimates.

The sequenced fragments can be aligned to the genome using splice-aware aligners such as STAR¹⁵¹, TopHat2¹⁵², or Subread¹⁵³. There is also an option to align to the transcriptome, but with the risk of missing uncharacterised transcripts¹⁴⁸.

Quantification of mapped reads can be performed on a transcript level or gene level, and the resulting reads should be normalised to the total number of counts accumulated for each sample in the RNA sequencing experiment to ensure that differences in library sizes are not interpreted as differences in expression¹⁴⁶. There is

also the option of correcting for gene length, since longer genes accumulate more reads. This is not required when comparing changes in expression within the same gene across samples, such as in the analysis of differential expression between samples and groups. Normalisation should be performed to adjust for differences in composition of the sequencing libraries, for example the effects on library composition that occur when some genes are being transcribed in only one of the groups, using *e.g.*, the trimmed mean of M-values¹⁵⁴. Read counts are either assigned or not assigned to a transcript, so differential expression can be calculated using discrete probability distributions such as the negative binomial distribution (used for example in edgeR and DESeq)¹⁴⁶. Differential expression can also be calculated using linear modelling with the limma package if the inherent mean-variance trend in RNA-seq data is adjusted for. Both methods perform well, but there are fewer false positives as the number of samples increases for the limma analysis pipeline¹⁵⁵.

Although still costly, a great advantage of RNA sequencing is that it is unbiased in the sense that it does not include a pre-selection of targets, *i.e.*, all RNA sequences in the sample have the potential to be detected.

Enzyme-linked immunosorbent assay

Enzyme-linked immunosorbent assays (ELISAs) can be performed in microplates with wells coated with an antibody specific for a target antigen¹⁵⁶. When the sample has been added and the antigen is bound, an enzyme-labelled secondary antibody is attached, followed by an enzyme substrate. The amount of generated product is proportional to the amount of antigen in the sample. Specific antibodies in a sample can be quantified in a similar way. In this case, antigens instead of antibodies are used to coat the wells, and the second reagent is an enzyme-labelled antibody specific for the antibody to be detected in the sample¹⁵⁶.

ELISA is a suitable method to detect one or maybe a few proteins/antibodies. To analyse multiple proteins using limited samples, a better choice might be a multiplex strategy, such as the one used in Luminex xMAP technology¹⁵⁷.

Luminex xMAP technology

Luminex xMAP technology is based on microspheres that are differentiable by fluorescence¹⁵⁷. The different coloured beads can be coated with various agents, *e.g.*, antibodies specific for a particular protein. The beads are mixed with the sample, and the protein is bound to the beads. An antibody labelled with phycoerythrin is then

attached to the bound protein. In the instrument, phycoerythrin and the dye mixture in the beads are excited by laser, and the fluorescence detected. The bead colour identifies the target protein, and the fluorescence from phycoerythrin is proportional to the amount of that protein in the sample.

Multiplex application of Luminex xMAP technology has major advantages including reduction of sample, time, and reagent consumption, but might reduce sensitivity compared with the same assays run as single reactions¹⁵⁷.

Genotyping

Characterisation of genetic variation can be performed in many different ways; the two used in this thesis are described below.

Genotyping using a sequence-specific primer PCR method

In this method primers designed to be specific for different alleles, marked with a fluorescent dye specific for each locus, only create a PCR product if the match is perfect¹⁵⁸. The products are analysed using capillary electrophoresis, where the fluorescent dye determines the locus, and the product size determines what allele was present in the sample. Since HLA regions contain extensive sequence variation³⁴, this method is suitable for HLA typing.

Genotyping using real-time PCR and TaqMan probes

This method is based on the presence of two TaqMan probes designed to match the two variants of a target polymorphism¹⁵⁹. The probes are labelled with different fluorophores. A homozygote for one variant will mainly generate PCR products to which one of the probes binds and fluoresces as it is cleaved by the polymerase, while a homozygote for the other variant mainly generates fluorescence from the other probe. A heterozygote will generate a mix of the two signals.

This method is fast and convenient, especially if pre-designed assays from a manufacturer are used. However, multiplex analysis using real-time PCR and TaqMan probes is difficult, so a more suitable multiplexing method should be used if several targets are to be genotyped in multiple samples.



Thesis aims

The main objective of this thesis was to complement and improve CD diagnostics by identifying and implementing new biomarkers, mainly based on gene expression (RNA and protein) in small intestinal biopsies and blood. The more specific aims were to:

- ✦ Correlate expression of genes (RNA, protein, biological functions) in small intestinal biopsies and/or blood with:
 - ✦ Presence/absence of CD
 - ✦ Small intestinal alterations (Marsh scale or similar)
 - ✦ Recovery during GFD

- ✦ Find ways to identify individuals affected by CD among those with low-grade intestinal injury and/or CD-specific antibody levels increased to different extents.

- ✦ Investigate biological processes involved in CD.



Methods — study specific

Study subjects and groups

Paediatric patients (0–18 years) investigated for suspected CD or at follow-up on a GFD or a GD at Ryhov County Hospital, Jönköping, were included in the studies after obtaining written consent. For all included study subjects, data on levels of CD-specific antibodies and histopathologic assessment were obtained. The study subjects that agreed to participate in the studies during the years 2007–2016 are characterised in Table 2. The subjects from this study population included in papers I–IV are further characterised in Table 3.

There were more females than males among the study subjects (female:male ratio 1.74), but the exact ratios varied between the groups and studies (Tables 2 and 3).

The study groups in papers I–IV partly overlap: 80 subjects are unique to 1 study, 40 subjects are included in 2 studies, 12 subjects in 3 studies, and 9 subjects in all 4 studies. In total, 141 study subjects were included in papers I–IV.

This study was approved by the Regional Ethical Review Board in Linköping (paper I: M16-06, M16-06 T99-08; paper II-IV: M16-06, M16-06 T99-08, 2011/239-31).

Table 2. Characteristics of subjects included in the study over the years 2007–2016.

Group ^b	n (F) ^c	Diagnosis	Diet	Age ^d	Histopathology ^a						Serology
					NAS	Mo	M1	M2	M3A/B/C	anti-TG2 ^e	
Active CD	110 (64)	CD	GD	8.4 (0.67–18)	0	0	0	1	30/33/46	0.02, 0.30, 0.68	
Active CD GFD	5 (4)	CD	GFD	15 (11–17)	0	0	1	0	4/0/0	0.60, 0.20, 0.20	
Not CD	37 (19)	Not CD	GD	8.6 (1.5–17)	0	36	1	0	0	0.91, 0.09, 0	
Not CD later	17 (12)	Not CD	GD	9.0 (1.1–17)	2	13	1	1	0	0.47, 0.53, 0	
CD later	10 (8)	CD	GD	11 (3.1–16)	0	7	3	0	0	0.11, 0.56, 0.33	
Normalised CD	39 (31)	CD	GFD	9.5 (3.0–19)	0	39	0	0	0	0.74, 0.26, 0	
Unclear	20 (12)	Unclear	GD	9.6 (3.1–16)	4	9	6	1	0	0, 0.63, 0.37	
Other ^f	3 (3)	CD	GD/GFD	12 (6.8–16)	0	2	1	0	0	0, 0.67, 0.33	
TOTAL	241 (153)			9.0 (0.67–19)	6	102	12	2	113	0.32, 0.32, 0.36	

^aIf a range of Marsh grades was provided by the pathologist, the highest Marsh grade is accounted for (Marsh [M]0, M1, M2, M3A, M3B, M3C). NAS ('Non-assessable'); Biopsies for histopathologic assessment were suboptimal and could not be used as grounds for a Marsh grade.

^bStudy subjects in groups CD later and Not CD later did not receive a CD diagnosis at the time of the biopsy sampling, but were later diagnosed as CD and not CD, respectively. Study subjects in group Unclear are under continued observation but have not yet received a CD diagnosis. Four subjects in group Active CD had a patchy distribution of the intestinal lesions.

^cThe same study subject can be included in multiple groups, with separate biopsy samplings at different time points. Data for each sampling is included in the appropriate group: Six study subjects are included in group Active CD, with follow-up biopsies in group Normalised CD. One study subject is included in Active CD GFD with follow-up biopsy in Normalised CD. One study subject in group Unclear has a follow-up biopsy in group Other. Two study subjects in group CD later have later biopsy samplings in group Active CD, of which one subject has two separate CD later biopsy samplings. In group Not CD later, three study subjects are included with two samplings each. In group Unclear, two study subjects are included with two samplings each. F = female.

^dMean age in years (min–max).
^eFractions of <1 times upper limit of normal (ULN), 1–10 times ULN, and >10 times ULN for autoantibodies against tissue transglutaminase (anti-TG2) in each group. There were missing values for two study subjects each in groups Active CD, CD later, and Not CD, and for one subject in group Unclear.

^fThe group Other contains two study subjects with a confirmed CD diagnosis, and a control biopsy on a gluten-containing diet (GD) after several years on a gluten-free diet (GFD). A biopsy was sampled from the third study subject in this group on a GFD after a previous Marsh 1 biopsy on GD.

Table 3. Characteristics of the study subjects included in papers I–IV.

	Paper I	Paper II	Paper III	Paper IV
Total (n)	67 ^a	52 ^b	83	30
Age ^c	7.4 (1.0–17)	9.4 (0.67–17)	8.8 (0.67–18)	8.4 (1.6–17)
Female/male	45/22	36/16	48/35	19/11
Exploratory/follow-up (n)	40/27	52/0	40/51 ^d	30/0
Reference on GD (n)	17	14	26	10
Marsh 0 (n)	17	12	26	10
Marsh 1–2 (n)	0	2	0	0
Anti-TG2 ^e	0.65, 0.35, 0	0.86, 0.14, 0	1, 0, 0	1, 0, 0
Follow-up on GFD (n)	15	14	5	10
Marsh 0 (n)	12	11	5	10
Marsh 1–2 (n)	3	3	0	0
Marsh 3 (n)	0	0	0	0
Anti-TG2 ^e	0.73, 0.20, 0.07	0.57, 0.43, 0	1, 0, 0	0.90, 0.10, 0
Active CD on GD (n)	31	18	42	10
Marsh 0 (n)	0	0	0	0
Marsh 1–2 (n)	3	0	0	0
Marsh 3 (n)	28	18	42	10
Anti-TG2 ^e	0, 0.27, 0.73	0.15, 0.20, 0.65	0, 0.21, 0.79	0, 0.40, 0.60
Other (n)	4	6	10	0
Marsh 0 (n)	2	2	2	0
Marsh 1–2 (n)	1	3	4	0
Marsh 3 (n)	1	1	4	0
Anti-TG2 ^e	0.25, 0.25, 0.50	0.33, 0.50, 0.17	0.30, 0.50, 0.20	-

^aThree out of 70 study subjects were excluded based on poor biopsy quality.

^bThree out of 56 study subjects were excluded due to low sample amount and one subject based on poor biopsy quality.

^cMean age in years (min–max).

^dEight study subjects from the exploratory study group were included in the follow-up study to validate the RNA sequencing results.

^eFractions of <1 times upper limit of normal (ULN), 1–10 times ULN, and >10 times ULN for autoantibodies against tissue transglutaminase (anti-TG2).

Biopsy and blood sampling

Between one and three small intestine biopsies were sampled, placed in pre-chilled RNAlater RNA Stabilization Reagent (Qiagen, Hilden, Germany), and stored at 4 °C overnight and then at -20 °C until RNA isolation. In papers I–IV, no more than one

biopsy from each sampling has been included. Sampling was performed immediately after routine samples were collected to ensure that diagnostic biopsies were prioritised. EDTA blood for DNA and protein studies was collected at the same time as blood sampling for routine diagnostics and did not require additional venous puncture. An aliquot of blood was removed for DNA isolation, and then plasma was isolated by centrifugation ($1500 \times g$ for 10 minutes). The plasma was stored in aliquots at -80°C . Blood was also collected for RNA studies, where a Tempus tube (Life Technologies, Carlsbad, CA) with stabilised blood was stored at 4°C overnight and then at -20°C until RNA isolation.

For routine diagnostics, biopsies were collected for histopathologic assessment and blood was collected for detection of IgA anti-TG2 (or in case of IgA deficiency, IgG anti-TG2), IgA antibodies against gliadin, and IgG anti-DG in serum. Antibody detection was performed using EliA kits on Phadia250 (Thermo Fisher Scientific, Waltham, MA) as described by the manufacturer. Biopsies for histopathologic assessment were formalin-fixated and paraffin-embedded, and hematoxylin-eosin stained.

Statistical methods and tools

Statistica (Statsoft, Tulsa, OK) was used for a large part of the statistical analyses in papers I–III: version 9.1 for paper I, version 10 for paper II, and version 13 for paper III.

To investigate gene expression in an unsupervised manner, hierarchical clustering and principal component analysis (PCA) were performed in Statistica (papers I and III) or Partek Genomics Suite version 6.6 (Partek Incorporated, St. Louis, MO) (paper III).

Significant differences in RNA/protein levels between groups were identified using non-parametric tests for significance in Statistica, including the Mann-Whitney U test (papers I–III), Kruskal–Wallis one-way analysis of variance (ANOVA) by ranks (paper II), and Pearson’s chi-squared test of independence (paper I). Parametric tests for the same purpose included ANOVA (paper III) in Partek Genomics Suite.

Correlations between different variables were identified using Spearman rank correlation in Statistica (papers I–III). For paper III, Spearman rank correlation was additionally performed in Partek Genomics Suite, and product-moment correlation was performed in Statistica.

Two methods were used to adjust for the risk of identifying false-positive results as a consequence of performing multiple tests for significance: false discovery rate¹⁶⁰ (FDR) (papers II–IV) and Bonferroni adjustment (papers II and III). Bonferroni was used for post-hoc comparisons in paper II and in power calculations in paper III. This conservative method of adjusting for multiple testing could miss true-positive associations¹⁶¹. FDR corrects for an expected proportion of false-positives under the null hypothesis¹⁶¹.

In paper II, logistic regression models with combinations of assays were calculated, and assay performances were evaluated using receiver operating characteristic (ROC) curves in MedCalc Statistical Software version 13.1.2 (MedCalc Software, Ostend, Belgium).

Group predictions based on gene expressions were performed using discriminant analysis in Statistica (paper I).

Pathway ANOVA (Partek Incorporated) was used to investigate differential gene expression on a pathway level without pre-selection of differentially expressed genes (paper III).

Overrepresentation of differentially expressed genes in gene groups corresponding to Gene Ontology (GO) terms was investigated using Partek Genomics Suite (paper III) and Fisher's test.

Gene set enrichment analysis (GSEA) was performed to identify gene groups and pathways of interest in CD using the GSEA software version 3.0 from Broad institute (paper IV).

Information regarding gene inclusion in gene groups or pathways was extracted from GO (papers III and IV), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (paper III), and Reactome (paper IV).

Additionally, multiple tools for clustering, visualisation, and annotation were applied to data in Cytoscape¹⁶² versions 3.4.0 (paper III) and 3.6.1 (paper IV), including EnrichmentMap¹⁶³, WordCloud¹⁶⁴, and AutoAnnotate¹⁶⁵.

Paper I

In paper I¹⁶, expression of 109 RNAs in small intestinal biopsies from individuals with and without a CD diagnosis were quantified using real-time PCR and analysed based on differential expression between Marsh 0 and 3 with the purposes of finding biomarkers reflecting the histopathologic assessment and developing a discriminant analysis based on those markers. Gene selection was based on previous knowledge

derived from the literature and focused on the hypothesis that important aspects of the histopathologic assessment could be reflected by a gene expression profile. Genes were primarily selected to reflect villous height, crypt elongation, immune response, and epithelial integrity.

Sample preparations

Biopsies were homogenised and then RNA was isolated using a spin column-based method. The resulting RNA was eluted in a 0.1×TE-buffer (1 mM Tris-HCl and 0.1 mM EDTA) with RNase inhibitor. RNA purity and concentration were determined spectrophotometrically. RNA integrity was assessed using chip-based gel electrophoresis, measured by RNA Integrity Number¹⁶⁶. Random primers were used for reverse transcription of the RNA into cDNA.

Gene expression analysis

Relative quantitative real-time PCR was performed on cDNA from each sample using TaqMan probes as the detection format (pre-designed gene expression assays from Life technologies). C_q-values were established with a baseline calculated using data from cycles 3 to 15, and thresholds were adjusted to the log-linear range and set to the same level for all samples within a single assay. The results were normalised using reference genes *HPRT1* and *PGK1*, chosen from 11 potential reference genes based on low sample-to-sample variation, calculated using the GeNorm¹⁶⁷ and NormFinder¹⁶⁸ algorithms in the Genex software (MultiD Analyses, Göteborg, Sweden). Reference gene calculations were based on two Marsh 0 samples, and four samples with Marsh grades between 2 and 3C.

Statistical analysis

In a subset of study subjects from the exploratory study group, genes with differential expression between Marsh 0 (n = 4) and Marsh 3 mucosa (n = 10) were identified without adjusting for multiple testing. A number of filtrations were subsequently made, and it would have been unwise to remove genes from the selection prematurely. A total of 50 differentially expressed genes were used in subsequent analyses based on the exploratory study group and sample relations based on these expressions were investigated by PCA and hierarchical clustering. Additionally, correlations between Marsh grade and gene expressions were investigated. The genes that best reflected Marsh grade were selected using a method that did not rely on a

linear relation in regard to gene expression between the Marsh grades; the range of values for each gene was segmented into 10 categories, and then Pearson's chi-squared test of independence was performed to find the 16 best predictors for a 3-group scenario (Marsh 0, Marsh 2–3B, Marsh 3C). The same method was also used for a 2-group scenario (Marsh 0, Marsh 2–3C), and the combination of these lists resulted in 24 genes which were used as input in a general discriminant analysis model. The gene subsets that were best at correctly classifying Marsh 0 and Marsh 2–3C were identified. Additionally, the gene subsets best at correctly classifying Marsh 0, Marsh 2–3B, and Marsh 3C were identified. The final selection of genes to use in the discriminant analysis was determined based on Wilk's lambda, the number of included predictors, and representation (villi/crypt, epithelial permeability, and immune response). The developed discriminant analyses were tested by stepwise omission of a single sample at a time, creating an algorithm based on the other samples, and then testing what group the omitted sample was placed in. The classification potential of the discriminant analysis from the 3-group scenario was further evaluated using an independent set of 27 biopsies.

A p-value <0.05 was considered significant.

Paper II

In paper II²⁷, the expression of RNAs and proteins in blood from individuals with and without a CD diagnosis were quantified using real-time PCR and Luminex xMAP technology/ELISA. Gene selection was based on information from the literature and on functional context. RNAs and proteins were analysed for differential expression between active CD (Marsh 2-3C), non-CD (Marsh 0-1), and normalised CD (Marsh 0-1), with the purpose of identifying blood-based CD biomarkers.

Sample preparations

RNA from stabilised blood was pelleted using centrifugation and then purified using a spin column-based method. For two samples, RNA was instead isolated from EDTA blood using a spin column-based method. The resulting RNA was eluted in 0.1×TE-buffer with RNase inhibitor. RNA quality assessment and reverse transcription were performed as in paper I.

DNA was purified from EDTA blood using an automated magnetic bead-based method, and concentration was determined spectrophotometrically.

Gene expression analysis

Gene expressions (RNA) were analysed using relative quantitative real-time PCR and cDNA from each sample, with TaqMan probes as the detection format (pre-designed assays, Life technologies). C_q-values were established as in paper I. Three potential reference genes were selected using a similar approach as in paper I, except that 32 instead of 11 reference genes were evaluated. The reference genes were evaluated based on three Marsh 0 samples, and six samples with Marsh grades between 2 and 3C. After evaluation of these three genes based on the entire dataset, *CDKN1B* was selected as the most stable reference gene to use for normalisation.

Protein quantification

Multiplex detection of proteins in plasma was performed using Luminex xMAP technology (Milliplex kits, Millipore, Billerica, MA). Standard curves were included and optimised using the Bio-Plex Manager (Bio-Rad, Hercules, CA).

No multiplexing option was available for the CD163 soluble protein, so it was detected using an ELISA, and the results were obtained and analysed using a microplate absorbance reader. A standard curve was included and optimised using the Akima method for curve fitting in the Magellan version 7.0 software package (Tecan Group Ltd, Männedorf, Switzerland).

Genotyping

HLA-DQ2 or DQ8 presence was determined using a sequence-specific primer PCR method and capillary electrophoresis^{169,170}. The data were used to calculate relative CD risks³⁸.

Statistical analysis

Differential expression was investigated between active CD, normalised CD, and non-CD subjects. For significant findings, post-hoc comparisons of mean ranks of all pairs of groups were performed to determine which group comparisons produced significant findings. Differences in gene expression were also investigated in study subjects with one *vs.* two DQB1*02 alleles.

Protein/mRNA levels were correlated with histopathology and CD risk gradient based on HLA type.

The diagnostic performances of individual assays and logistic regression models of assay combinations were evaluated.

A p-value <0.05 was considered significant. For all statistical analyses, except the post-hoc comparisons, an FDR was used and set to 5%.

Paper III

In paper III¹⁷¹, RNAs isolated from small intestinal biopsies from individuals with a CD diagnosis and active CD (Marsh 3) and individuals without a CD diagnosis (Marsh 0) were quantified using RNA sequencing to identify potential CD biomarkers and investigate biological processes involved in CD. A follow-up study was performed on a separate study group where a select number of genes' expressions were quantified using real-time PCR to identify CD biomarkers.

Sample preparations

RNA isolation and quality assessments were performed as for paper I, but with additional concentration measurements using a fluorometric method.

Gene expression analysis

Sequencing libraries were prepared from small intestinal biopsy RNA using the kit TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat (Illumina), which removes ribosomal RNA. RNA sequencing was performed on a HiSeq 2500 (Illumina) using a read length of 1 × 50 base pairs (bp). The resulting fragments were aligned to genome build hg19 using the STAR algorithm¹⁵¹, and gene expressions were quantified using the Partek E/M algorithm (Partek Flow), with annotation source Refseq transcripts release 71 from NCBI (National Center for Biotechnology Information).

Follow-up of selected genes' expressions was performed using relative quantitative real-time PCR and cDNA from each sample, with TaqMan probes as the detection format (pre-designed assays, Life technologies). C_q-values were established as in paper I. Two reference genes were selected for normalisation; *EIF2B1* and *ZFR*. *EIF2B1* was selected based on unpublished data from a study on reference genes performed in association with paper I, and *ZFR* was selected based on stable expression and low variation between samples in the RNA sequencing results. No group differences (active CD vs. non-CD) were found for the two genes, neither in the RNA sequencing results nor in the real-time PCR results.

Genotyping

The presence of HLA-DQ2.5 in *cis* was investigated using a SNP in linkage disequilibrium (rs2187668). Genotyping was performed using real-time PCR and TaqMan probes (pre-designed assay, Life technologies).

Statistical analysis

To reduce the influence of different total numbers of reads for each sample and different gene length, reads per kilobase per million mapped reads (RPKM) were used in most calculations. The influences of age and antibody levels on general small intestinal gene expression were investigated by correlating the coordinates along principal components from a PCA based on all detected gene expressions with these variables. Similarly, the influences of gender and Marsh grade were investigated with ANOVA.

Differentially expressed genes were identified using one-way ANOVA based on RPKM values. For the identification of highly differentially expressed genes (HDEGs) an additional method was used. A mean-variance trend was modelled, and the variance of each gene was then shrunk toward the mean variance of the entire dataset ('limma trend'¹⁷²), and differential expression was analysed using linear regression. The lists of HDEGs from this method and from ANOVA based on RPKM values were combined and used as a base for the selection of genes for the follow-up study. More stringent criteria were used for the HDEGs as compared to the differentially expressed genes (a total number of reads >1000, or a mean expression >1 RPKM in one or both study groups, together with a p-value <0.000001 and fold change >4 or <-4 between active CD and non-CD study subjects).

To find pathways and gene groups associated with CD, expressions of genes involved in different pathways (KEGG) and gene groups (GO) were analysed on a pathway/gene group level and compared between active CD and non-CD subjects.

Unsupervised investigation (PCA) of sample relations based on gene expressions was performed in the follow-up study, and differential expression was identified between groups. Correlations between Marsh grade and gene expressions were investigated.

Unless otherwise specified, all p-values were adjusted for multiple testing using FDR, and FDR-adjusted p-values <0.05 were considered significant. A lower cut-off was used for some analyses.

Paper IV

In paper IV, RNAs from stabilised blood from individuals with a CD diagnosis and active CD (Marsh 3), with normalised CD on a GFD (Marsh 0), or without a CD diagnosis (Marsh 0), were quantified using RNA sequencing to identify potential CD biomarkers and investigate biological processes involved in CD.

Sample preparations

RNA purification from stabilised blood was performed as in paper II, and quality assessments as in paper III.

Gene expression analysis

RNA sequencing libraries were prepared from stabilised blood using TruSeq Stranded Total RNA with Ribo-Zero Globin (Illumina), which removes ribosomal and globin RNA. RNA sequencing was performed on a HiSeq 2500 with a read length of 1×50 bp as in paper III.

The resulting reads were aligned to genome build hg38, and the number of reads that mapped to each gene was quantified with RSubread¹⁵³. Ensembl version 85 was used as an annotation source.

Genotyping

Genotyping of the SNP rs2187668 was performed as in paper III.

Statistical analysis

Counts were normalised to the total number of reads per sample, expressed as counts per million (cpm), and normalisation factors for sample composition differences were calculated using the trimmed mean of M-values. The variance associated with mean log-cpm decreases with higher gene expression; precision weights that adjust for this were calculated using the voom function. The weights were used in subsequent steps to enable linear modelling of gene expression based on log-cpm values. The variance was further adjusted based on information about the pooled variance, and moderated t-statistics were calculated using empirical Bayes moderation with and without adjustment for multiple testing (FDR) and without a fold change requirement. A p-value < 0.05 after adjustment for multiple testing was considered significant.

GSEA based on a ranked list of genes (computed by multiplying direction [sign] of fold change and logarithm of p-value for each gene) was performed. Genes at the top

or bottom of the list are more likely to be involved in processes that differentiate the study groups, and so enriched genes (FDR-adjusted p-value <0.075) at the top or the bottom of the ranked list contained in GO Biological Process terms or Reactome pathways were identified.

Results and analysis

Papers I and III — Gene expression analysis in small intestinal biopsies

In paper I, genes were primarily selected based on their possible reflection of important elements in the histopathologic assessment, and RNA levels were determined in small intestinal biopsies from individuals with and without CD. From an exploratory set of 109 genes, the expressions of eight genes were identified as a good combination for separating Marsh 0, Marsh 2–3B, and Marsh 3C mucosa: *APOC3* and *CYP3A4* (villi markers), *OCNL* (epithelial integrity marker), *MAD2L1* and *MKI67* (crypt markers), *CXCL11*, *IL17A*, and membrane-bound *CTLA4* (immune response markers) (Fig. 6). The follow-up study including biopsies from a separate subset of individuals confirmed the results.

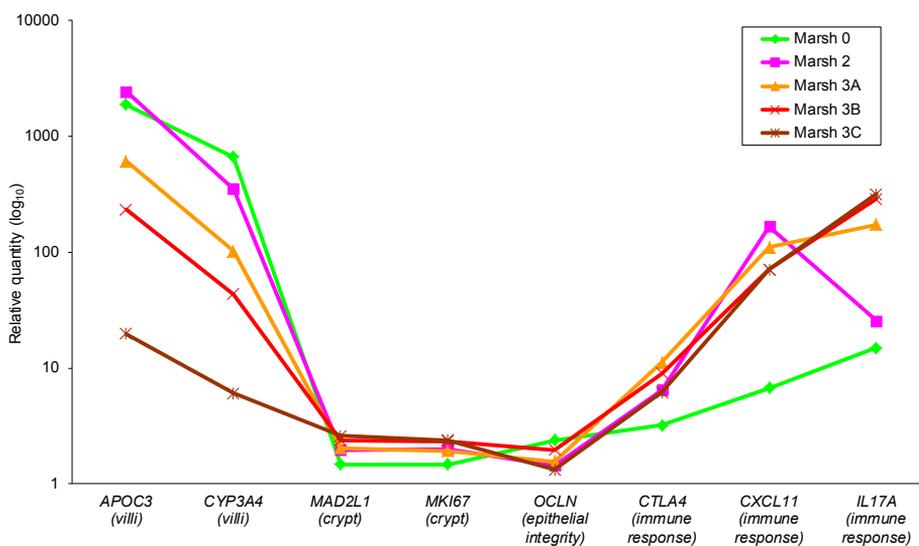


Figure 6. Median relative quantity per Marsh grade based on 39 study subjects in paper I for the eight genes included in the discriminant analysis. The results are presented on a logarithmic scale.

In paper III, RNA isolated from small intestinal biopsies from individuals with active CD and without CD was sequenced. Higher expression in active CD compared to non-CD was found for contexts such as phagocytosis, bacteria/virus, intestinal barrier function, immune response, and autoimmune conditions. Lower expression in active CD compared to non-CD was found for contexts such as metabolism and transportation. Other CD-associated contexts implicated in the analysis included cell cycle and signal transduction.

A total of 94 HDEGs were identified in paper III, of which 29 target genes were selected for analysis by real-time PCR. The selection was guided by results from the gene group (GO term) analysis. Two genes were later excluded due to lack of detection with real-time PCR. Along with two reference genes (*EIF2B1* and *ZFR*), the expression levels of the 27 genes were analysed in a follow-up study. A few samples were included from the RNA sequencing study groups, and the results showed that the real-time PCR and the RNA sequencing findings correlated well. Based on real-time PCR results from non-CD (Marsh 0) and active CD (Marsh 3A, 3B, and 3C) study subjects, gene expressions correlated with Marsh grade. A PCA was formed based on these gene expressions, and additional study subjects were projected onto the PCA based on their respective gene expressions. The gene expressions separated non-CD and active CD and showed a gradual transition from Marsh 0 through Marsh 3A and 3B to 3C (Fig. 7). CD subjects normalised on a GFD clustered with non-CD subjects. Marsh 0–2 subjects with positive anti-TG2 where a diagnosis could not be established until additional biopsy sampling/s had been performed ('CD later') were generally more similar to Marsh 3 subjects than Marsh 0 subjects (Fig. 7). Analyses of differential gene expression in this group compared to the non-CD subjects showed that five genes were significantly differentially expressed: *GBP5*, *CXCL10*, *IFI27*, *IFNG*, and *UBD* (Fig. 8). For these five genes, expression levels in CD later subjects were comparable to those in active CD (Fig. 8). A similar study subject that was determined not to have CD after 4 years of investigation was positioned within the Marsh 0 group ('Not CD later') (Fig. 7).

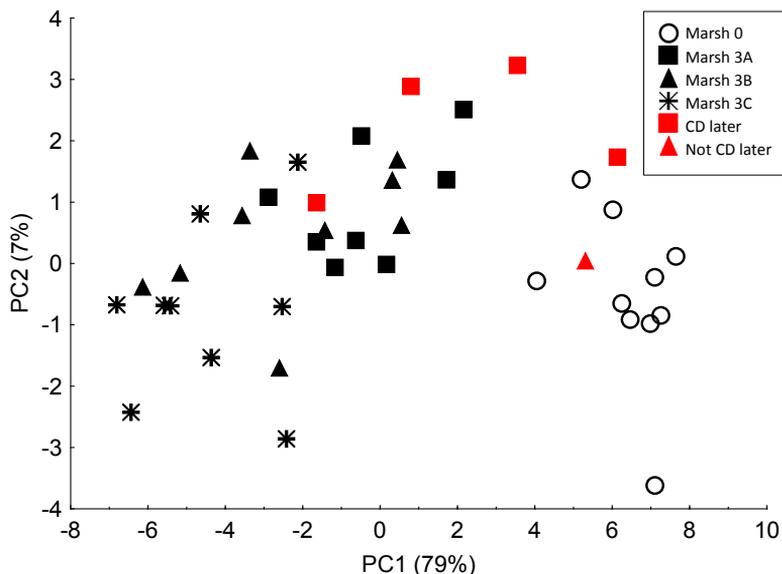


Figure 7. Coordinates of study subjects in a PCA based on gene expressions from 27 genes included in the follow-up study in paper III. Variation accounted for by each principal component (PC) is given in percent of the total variation in the dataset. *Figure simplified from Bragde et al.¹⁷¹*

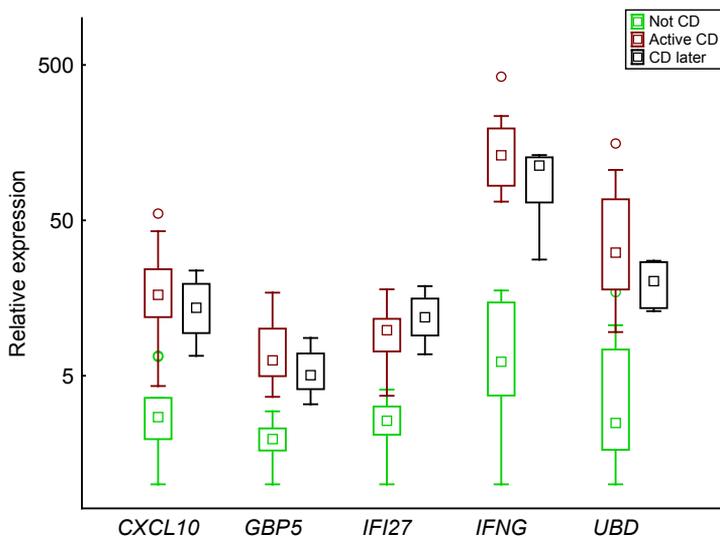


Figure 8. Box plot visualising the expression of five genes with differential expression in CD later subjects compared with non-CD subjects on a logarithmic scale. The box and the square within the box represent the 25–75% interquartile range and the median, respectively. The whiskers represent the non-outlier ranges. *Figure from Bragde et al.¹⁷¹*

Papers II and IV — RNA and protein levels in blood

Although a gene expression profile in biopsies can help determine a CD diagnosis, a gene expression profile (based on RNA and/or proteins) would be even more useful if it could be applied to a sample obtained using less invasive techniques, such as blood. In paper II, to investigate whether a blood-based gene expression profile could be used as a CD diagnostic tool, proteins and mRNAs were selected based on published studies on blood mRNA/protein expression in CD and by functional context. The results showed that *CXCL11* protein and *TNFSF13B* mRNA levels were higher in active CD than in non-CD study subjects, and *TNFSF13B* mRNA levels were also higher in normalised CD. *TNFRSF9* mRNA levels were higher in normalised CD subjects compared to non-CD and active CD subjects. No significant differences in RNA/protein levels were found based on number of DQB1*02 alleles and no correlations were found with HLA genotype risk for CD. ROC curve analysis of discrimination between subjects with active CD and without CD revealed an area under the curve (AUC) of 0.97 for anti-TG2, 0.81 for *CXCL11* protein, 0.85 for *TNFSF13B* mRNA, and 0.91 for a logistic regression model based on *CXCL11* protein and *TNFSF13B* mRNA. ROC curve analysis of discrimination between study subjects with active and normalised CD revealed AUCs of 0.90 for anti-TG2 and 0.78 for *TNFRSF9* mRNA. Compared to anti-TG2 alone, the addition of new potential markers resulted in a modest but non-significant contribution to the diagnostic capacity of anti-TG2.

To investigate gene expression unbiased by previous assumptions, RNA sequencing was performed on stabilised blood from active CD, normalised CD, and non-CD study subjects (paper IV), and the results showed that in whole blood no large differences in the expression of single genes could be found between the groups. However, by focusing on genes with expression differences on a gene set level, processes likely to be involved in CD pathogenesis were identified. The results indicated a mainly pro-inflammatory response in active CD. The results also pointed to negative regulation of viral replication being associated with active CD, together with positive regulation of nitric oxide metabolic processes, monocarboxylic acid transport, and impaired haemostasis regulation. Processes involving signal transduction, metabolic processes, proliferation, differentiation, cell migration, and cell survival could also be affected. In normalised CD, we observed lower activity in cell cycle and transport as compared to non-CD.

Discussion

Our results showed that Marsh grade from the histopathologic assessment can be reflected using the expression of a few genes in small intestinal biopsies from children (paper I). We also showed that processes such as immune response, microbial infection, phagocytosis, intestinal barrier function, metabolism, and transportation might be involved in CD and used this information in the selection of potential biomarkers (paper III). The expressions of the selected biomarkers correlated with Marsh grade, and for some of the markers (although based on a small number of samples), altered expression in CD mucosa was found before the histopathologic assessment could confirm intestinal alterations compatible with a CD diagnosis. Active CD study subjects with anti-TG2 ≤ 10 times ULN that require a histopathologic assessment of intestinal biopsies to confirm the CD diagnosis constituted 32% of the active CD subjects in our study population (Table 2). Furthermore, in the group with study subjects who received a CD diagnosis at a later time point compared to the biopsy sampling included in this study (group CD later), 67% had anti-TG2 levels ≤ 10 times ULN. This indicates that intestinal biopsies will still be performed to establish a CD diagnosis in many cases, and gene expression analysis could be used to complement the histopathologic assessment for these cases.

Differential expression in CD mucosa compared to normal mucosa was previously identified in several studies¹⁰⁷. There has also been work performed on the subject of pre-symptomatic CD, based on gene expression in PBMCs from nine individuals, where differential gene expression was detected before symptom onset and antibody production in CD¹³⁴. Several possible mRNA biomarkers for CD have been identified (Table 1). Results from us and others show that there is potential for gene expression analysis in CD diagnostics not only for clear cases, but also for unclear cases and early in the disease progression.

There is also potential for gene expression analysis in the follow-up of CD patients on a GFD. The ability of the identified intestinal markers to follow treatment (GFD) was indicated in paper I and in the follow-up study in paper III, where expressions in GFD-treated CD study subjects were similar to non-CD subjects and different from active CD subjects. However, blood is a more suitable material for follow-up, with less invasive sampling. In paper II, a blood marker was identified with some potential to discriminate CD subjects on a GFD from active CD subjects. A study by Sangineto *et al.* compared normalised CD study subjects on a GFD with non-CD subjects and

identified several genes with differential expression using microarray analysis in purified PBMCs¹⁷³. This could not be replicated in our study using RNA sequencing based on stabilised whole blood (paper IV), perhaps due to different study designs, but differential expression was identified on a gene set level.

Our results indicate that differences in gene expression in stabilised whole blood between active CD and non-CD are much smaller than in intestinal mucosa (papers II and IV). In fact, no single gene was found to be significantly differentially expressed after adjusting for multiple testing in the RNA sequencing study (paper IV). Other gene expression studies based on purified cell types¹⁰⁶ or PBMCs¹⁷⁴ from blood identified several differentially expressed genes in CD compared to non-CD. One reason for the discrepant results might be that stabilised blood contains a mixture of blood cells contributing to the RNA pool, as compared to studies of purified cell types. In disease, proportion changes of leukocyte subtypes can disguise disease-associated gene expression in leukocytes¹⁷⁵. PBMC samples, although not consisting of a single cell type, still contain RNA from less diverse cell types compared to whole blood¹⁷⁶. Stabilised blood has been shown to yield a reduced number of expressed genes compared with PBMCs, with larger inter-individual expression variation and a smaller number of differentially expressed genes^{176,177}. In a study by Feezor *et al.*¹⁷⁷, the authors performed *ex vivo* stimulation with *Staphylococcus* enterotoxin B of blood from healthy donors. The blood was stabilised or leukocytes were isolated, and microarray analysis showed reduced responsiveness in the gene expression profiles from stabilised blood compared with isolated leukocytes. However, in a study of gene expression in stabilised whole blood compared to PBMCs as tools to evaluate perioperative inflammatory response in patients with advanced heart failure, the molecular fingerprints were similar between the two sample materials¹⁷⁶. Since the handling of blood samples prior to PBMC isolation can affect gene expressions¹⁷⁸, using stabilised blood might reduce the influence of pre-analysis factors.

Studies of gene expression in stabilised blood on a gene set level, in contrast to gene level, showed promising results (paper IV) and should be investigated further. It would also be of interest to investigate to what extent blood gene expression reflects duodenal gene expression. For that purpose, analyses of the datasets from papers III and IV should first be performed using the same set of bioinformatics tools.

Transcriptional changes identified in relation to CD in different studies were recently summarised¹⁰⁷, and some of those findings were replicated in our RNA sequencing studies in intestinal mucosa and blood. CD-associated processes

implicated in biopsies and blood in our work — and in most of the studies in the summary — included metabolism, cell cycle, signal transduction, immune- and inflammatory response-related processes, and transport. CD-associated contexts identified in multiple studies could be suitable targets for further investigations.

The difference in cell compositions that could arise from changes that occur in CD mucosa (*e.g.*, reduced villous height, crypt elongation, increase of IELs in the epithelial layer, and influx of inflammatory cells into the lamina propria) could affect gene expression analyses based on non-CD and CD tissue. Likewise, differences in ratios of immune cells in blood could influence gene expression analyses based on whole blood. This might facilitate the separation of samples based on disease/not disease, but could also influence the interpretation of biological context, and should be considered. However, the results are still of interest both in relation to the investigations into CD pathogenesis, as well as in the development of biomarkers on a gene set level for diagnostic purposes.

Non-CD subjects included in papers I–IV are individuals with suspected CD that are subjected to an intestinal biopsy sampling, but are found not to have CD. The reasons for biopsy sampling vary, including genetic risk, anti-TG2 positivity, a need to exclude CD as a diagnosis, or symptoms. This is the population from which we need to separate CD cases, so it is well suited as a control population, but it would add to the project to characterise gene expression in cases without suspected CD, preferably with a CD-incompatible HLA type.

No controls for differential diagnostics were included in papers I–IV. The incidence of differential diagnoses in regards to CD in Swedish children should be considered in this context. For example, although IBD and CD can produce similar clinical symptoms such as diarrhoea and abdominal pain, IBD has a lower prevalence than CD¹⁷⁹. Duodenal alterations are sometimes found in IBD and are more common in children¹⁸⁰. However, no cases of IBD were found in our study population. Additionally, biopsy evaluation is not the sole basis for the diagnosis of CD, but is part of a diagnostic procedure where symptoms, heredity, CD-specific antibodies, and HLA type are also taken into consideration, and there is a requirement for a clinical response to a GFD before a final CD diagnosis can be established¹. The expressions of selected biomarkers in other inflammatory conditions of the small intestine should be investigated, and it would be very interesting to determine which pathways are shared and which are disease specific. In a study of CD and ulcerative colitis, expressions of genes in shared genetic susceptibility loci and specific CD risk loci

were investigated and revealed both common and opposite expression patterns in the two diseases¹⁸¹.

There is great potential in the diagnostic use of gene expression analysis performed by real-time PCR; it has a good reproducibility and is easily performed with instrumentation available at most clinical laboratories today. Rapid changes can be reflected, since the expression profile is a snapshot of what is happening in the tissue at the sampling time point¹⁸². The efforts made by us and others to characterise gene expression in CD could be summarised into a biomarker profile, based on single biomarkers and/or biological processes, to facilitate CD diagnostics. However, the loss of visualisation compared to the histopathologic assessment is one issue when it comes to gene expression analysis in intestinal biopsies, another is, as mentioned, that gene expression in other diseases of the gastrointestinal tract could show partly the same profile both in the intestine and blood¹⁸¹. The need for a reference population to compare with individual gene expressions could make it difficult to implement the method clinically. Qualitative rather than quantitative transcriptional diagnostic signatures have been proposed as a stable approach to gene expression profiling for diagnostic purposes¹⁸³, which makes use of expression ratios between gene pairs within the same individual instead of relative quantities in relation to other samples. This approach could be investigated for CD diagnostics. Additionally, since the CD population is genetically heterogenic¹³, the same phenotype (*i.e.* inflammatory reaction to gluten) could in theory be accomplished by different genetic predispositions, which could also affect the expressions of different genes. Pathway information has been shown to be useful for predicting individual risk for disease^{184,185}. Although a few solitary biomarkers could be advantageous in the diagnosis of CD, basing gene expression profiles on pathway information instead of single genes might also reveal disease heterogeneity between patients and add stability to a diagnostic method based on gene expressions. Biomarkers could be selected to reflect pathways, or, as the cost of RNA sequencing decreases, the pathways could even be characterised using this unbiased method in the future.

Conclusions

Based on the findings in papers I–IV, the following conclusions can be drawn in regard to CD in children and adolescents:

-  Gene expression fold changes are large between active CD and non-CD small intestinal mucosa.
-  Gene expression fold changes are small between active CD and non-CD stabilised whole blood.
-  Biological processes are affected in active CD. This is detectable in both small intestinal biopsies and blood.
-  Minor histological changes in the CD small intestinal mucosa are accompanied by alterations in gene expression. This could be clinically relevant to help establish a CD diagnosis when histopathology is inconclusive for CD, but needs further evaluation.
-  An expression profile based on a few individual genes in small intestinal biopsies has the potential to be used as a diagnostic tool in CD, but it requires validation by analysis of gene expression in small intestinal biopsies from differential diagnoses.
-  A diagnostic gene expression profile for CD based on whole blood and/or intestinal biopsies could be feasible on a gene set level, but this needs further investigation.



Thesis contribution and future perspectives

When this project was initiated, the small intestinal biopsy was central in the CD diagnostic procedure. The project set out to decrease dependency on the histopathologic assessment for a CD diagnosis by a gene expression profile, either based on small intestinal biopsies or blood. A gene expression profile in blood would have the extra benefit of easy sampling, in addition to relieving the need for a small intestinal biopsy for CD diagnosis and follow-up.

In 2012, the non-biopsy criteria for a CD diagnosis were presented by ESPGHAN, where cases with high anti-TG2 levels (>10 times ULN) could be diagnosed without a small intestinal biopsy under certain circumstances. The need for a gene expression profile then shifted more toward cases with lower anti-TG2, where the biopsy was still mandatory for a CD diagnosis. Our results indicate that active CD study subjects have an altered gene expression profile in the small intestine compared with non-CD subjects, and that some expression alterations can be detected before the presence of CD-associated changes of the small intestine can be confirmed by histopathology. Conversely, gene expression of individual genes in whole blood does not provide large enough differences between active CD and non-CD to be used for diagnostic purposes. However, the analysis of expressions on a gene set level indicated differences between active CD and non-CD and should be investigated further.

A gene expression profile based on small intestinal biopsies could be used as a method of assessing small intestinal status in cases where extra information is needed to establish a CD diagnosis, such as in mild histological lesions and/or low anti-TG2 levels, when the pathologist has technical issues with the histopathologic assessment, or in other cases where additional information regarding intestinal status in CD is needed. This information could help establish a correct CD diagnosis without the need for a repeat biopsy, which would reduce time, effort, and cost for the health care system and save the patient from unnecessary biopsy samplings.

Future work will be focused on further investigations into gene expression in CD and non-CD study subjects with positive anti-TG2 and/or minor histological alterations using RNA sequencing to find out what biological processes and genes are affected at this stage and how that relates to the risk of developing CD or to return to normal histopathology. Additionally, information from RNA sequencing could suggest pharmacological targets for CD treatment. A final group of biomarkers (on a

gene level and/or a gene set level) should be selected and validated in CD study subjects, healthy controls, and subjects with differential diagnoses.

It would also be interesting to investigate the potential of gene expression analysis in FFPE intestinal biopsies in CD diagnostics, as FFPE material is available as an archival material from most CD cases. This provides the added possibility to investigate the relationship between histopathology and gene expression using successive sections from the FFPE biopsies for analysis based on histopathology or gene expression.

Acknowledgements

Ett stort varmt tack till alla som stöttat och hjälpt till i det här projektet genom åren!

Jan Söderman – Din kunskap har varit en trygghet som gjort att jag vågat mig på det här projektet. Tack för att du hjälpt mig att fokusera på rätt saker och att tro på mig själv. **Ulf Jansson** – du har hela vägen varit lika kunnig, lugn och entusiastisk. **Ewa Grodzinsky** – när jag träffar dig så blir dagen lite bättre, tack för att du alltid är så glad och positiv! Tack även för dina konstruktiva råd och ditt stöd. **Mats Fredrikson** – tack för kloka råd genom åren, alltid lika lugn. **Inga-Lena Hultman** – vart skulle det här projektet (och jag) varit utan dig? **Tack!**

Tack till **Ingvar Jarlsfelt** för att du tog dig tid att hjälpa oss med histopatologiska bedömningar och för att du bidragit med din kunskap till projektet.

Personalen på Konsultations- och behandlingscentralen (KBC) och Dagkirurgienheten – Hade ni inte hjälpt till så hade det inte blivit något material till den här studien. Ni har engagerat och intresserat er och varit en konstant och väldigt positiv del av det här projektet. Tack även till alla patienter som ställt upp i studien.

Tack **Hanna Odén Poulsen och Johanna Rundbäck** för ert stöd, och för allt ni har gjort för att underlätta på vägen.

Kollegor och vänner på Laboratoriemedicin - Tack för att ni gör detta till en trevlig arbetsplats att komma till varje dag. Ett speciellt tack till **Molekylärgänget** för stöd, uppmuntran och fika. **Malin och Linda** – tack kära vänner för att ni funnits där som stöd precis när man behövt det som mest. Ett stort tack också för att ni hjälpt till att samla in material till studien. **Lisa** – en vapendragare och vän, tack! **Olaf** – du ställer alltid upp när man behöver hjälp, tack! **Bettan, Sara, Sofie och Renate** – tack för råd och stöd och glada hejarop.

Till **alla mina vänner** – Tack för att ni är ni. Jag är så glad att ha er i mitt liv! Snart har jag lite mer tid för våra äventyr.

Min familj – tack för ert stöd. Speciellt tack till min **storasyster Malin** med familj för all hjälp med att ro detta projekt i land.

Min svärfamilj – tack för att ni får mig att känna mig som en del av er familj, och tack för att ni alltid, alltid ställer upp!

Henrik – min allra bästa vän, min stora kärlek och min stöttepelare i livet. Du gör mig glad varje dag! Tillsammans klarar vi allt. Tack för stöd och uppmuntran under alla år. Även om du inte alla gånger förstått vad det är som är så viktigt så har du alltid ställt upp. **Carl och Casper** – mina ♥♥. Tack för att ni visar mig vad som verkligen är viktigt i livet.

This thesis was supported by grants from Futurum – the Academy for Health and Care, Region Jönköping County, and the Medical Research Council of Southeast Sweden.

References

1. Husby S, Koletzko S, Korponay-Szabo IR, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;**54**(1):136-60.
2. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* 1990;**65**(8):909-11.
3. Corazza GR, Villanacci V, Zambelli C, et al. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2007;**5**(7):838-43.
4. Weile B, Hansen BF, Hagerstrand I, et al. Interobserver variation in diagnosing coeliac disease. A joint study by Danish and Swedish pathologists. *APMIS* 2000;**108**(5):380-4.
5. Taavela J, Koskinen O, Huhtala H, et al. Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. *PloS one* 2013;**8**(10):e76163.
6. Collin P, Kaukinen K, Vogelsang H, et al. Antiendomysial and antihuman recombinant tissue transglutaminase antibodies in the diagnosis of coeliac disease: a biopsy-proven European multicentre study. *European journal of gastroenterology & hepatology* 2005;**17**(1):85-91.
7. Rostami K, Villanacci V. Microscopic enteritis: novel prospect in coeliac disease clinical and immuno-histogenesis. Evolution in diagnostic and treatment strategies. *Dig Liver Dis* 2009;**41**(4):245-52.
8. Namatovu F, Sandstrom O, Olsson C, et al. Celiac disease risk varies between birth cohorts, generating hypotheses about causality: evidence from 36 years of population-based follow-up. *BMC gastroenterology* 2014;**14**:59.
9. Kohlmann A, Haschke-Becher E, Wimmer B, et al. Intraplatform reproducibility and technical precision of gene expression profiling in 4 laboratories investigating 160 leukemia samples: the DACH study. *Clinical chemistry* 2008;**54**(10):1705-15.
10. Kohlmann A, Kipps TJ, Rassenti LZ, et al. An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase. *Br J Haematol* 2008;**142**(5):802-7.

11. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, et al. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* 2016;**17**(5):257-71.
12. Moreno ML, Rodriguez-Herrera A, Sousa C, et al. Biomarkers to Monitor Gluten-Free Diet Compliance in Celiac Patients. *Nutrients* 2017;**9**(1).
13. Lundin KE, Sollid LM. Advances in coeliac disease. *Curr Opin Gastroenterol* 2014;**30**(2):154-62.
14. Ludvigsson JF, Leffler DA, Bai JC, et al. The Oslo definitions for coeliac disease and related terms. *Gut* 2013;**62**(1):43-52.
15. Dube C, Rostom A, Sy R, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* 2005;**128**(4 Suppl 1):S57-S67.
16. Gee S. The celiac affection. *St Bartholomew's Hosp Rep* 1888;**24**:17-20.
17. Dicke WK, Weijers HA, Van De Kamer JH. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr* 1953;**42**(1):34-42.
18. Berger E, Buergin-Wolff A, Freudenberg E. [Diagnostic Value of the Demonstration of Gliadin Antibodies in Celiac Disease]. *Klin Wochenschr* 1964;**42**:788-90.
19. Chorzelski TP, Beutner EH, Sulej J, et al. IgA anti-endomysium antibody. A new immunological marker of dermatitis herpetiformis and coeliac disease. *Br J Dermatol* 1984;**111**(4):395-402.
20. Dieterich W, Laag E, Schopper H, et al. Autoantibodies to tissue transglutaminase as predictors of celiac disease. *Gastroenterology* 1998;**115**(6):1317-21.
21. Kuja-Halkola R, Lebwohl B, Halfvarson J, et al. Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins. *Gut* 2016;**65**(11):1793-98.
22. Mustalahti K, Catassi C, Reunanen A, et al. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;**42**(8):587-95.
23. Myleus A, Ivarsson A, Webb C, et al. Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr* 2009;**49**(2):170-6.
24. Ivarsson A, Persson LA, Nystrom L, et al. Epidemic of coeliac disease in Swedish children. *Acta Paediatr* 2000;**89**(2):165-71.

25. Tapsas D, Hollen E, Stenhammar L, et al. Unusually High Incidence of Paediatric Coeliac Disease in Sweden during the Period 1973 - 2013. *PloS one* 2015;**10**(12):e0144346.
26. Grodzinsky E, Franzen L, Hed J, et al. High prevalence of celiac disease in healthy adults revealed by antigliadin antibodies. *Annals of allergy* 1992;**69**(1):66-70.
27. Lionetti E, Castellaneta S, Francavilla R, et al. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med* 2014;**371**(14):1295-303.
28. Vriezinga SL, Auricchio R, Bravi E, et al. Randomized feeding intervention in infants at high risk for celiac disease. *N Engl J Med* 2014;**371**(14):1304-15.
29. Szajewska H, Shamir R, Mearin L, et al. Gluten Introduction and the Risk of Coeliac Disease: A Position Paper by the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr Gastroenterol Nutr* 2016;**62**(3):507-13.
30. Green PH, Lebwohl B, Greywoode R. Celiac disease. *J Allergy Clin Immunol* 2015;**135**(5):1099-106; quiz 107.
31. Sandstrom O, Rosen A, Lagerqvist C, et al. Transglutaminase IgA antibodies in a celiac disease mass screening and the role of HLA-DQ genotyping and endomysial antibodies in sequential testing. *J Pediatr Gastroenterol Nutr* 2013;**57**(4):472-6.
32. Fasano A, Berti I, Gerarduzzi T, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* 2003;**163**(3):286-92.
33. Ciccocioppo R, Kruzliak P, Cangemi GC, et al. The Spectrum of Differences between Childhood and Adulthood Celiac Disease. *Nutrients* 2015;**7**(10):8733-51.
34. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annual review of immunology* 2013;**31**:443-73.
35. Sollid LM, Markussen G, Ek J, et al. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* 1989;**169**(1):345-50.
36. Stamnaes J, Sollid LM. Celiac disease: Autoimmunity in response to food antigen. *Seminars in immunology* 2015;**27**(5):343-52.

37. Abadie V, Sollid LM, Barreiro LB, et al. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annual review of immunology* 2011;**29**:493-525.
38. Margaritte-Jeannin P, Babron MC, Bourgey M, et al. HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. *Tissue Antigens* 2004;**63**(6):562-7.
39. Megiorni F, Mora B, Bonamico M, et al. HLA-DQ and risk gradient for celiac disease. *Hum Immunol* 2009;**70**(1):55-9.
40. Kang JY, Kang AH, Green A, et al. Systematic review: worldwide variation in the frequency of coeliac disease and changes over time. *Aliment Pharmacol Ther* 2013;**38**(3):226-45.
41. Bevan S, Popat S, Braegger CP, et al. Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999;**36**(9):687-90.
42. van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007;**39**(7):827-9.
43. Hunt KA, Zhernakova A, Turner G, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008;**40**(4):395-402.
44. Garner CP, Murray JA, Ding YC, et al. Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet* 2009;**18**(21):4219-25.
45. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;**42**(4):295-302.
46. Ostensson M, Montén C, Bacelis J, et al. A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease. *PLoS one* 2013;**8**(8):e70174.
47. Trynka G, Zhernakova A, Romanos J, et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut* 2009;**58**(8):1078-83.
48. Trynka G, Hunt KA, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011;**43**(12):1193-201.
49. Withoff S, Li Y, Jonkers I, et al. Understanding Celiac Disease by Genomics. *Trends in genetics : TIG* 2016;**32**(5):295-308.

50. Jabri B, Sollid LM. T Cells in Celiac Disease. *J Immunol* 2017;**198**(8):3005-14.
51. Kumar V, Gutierrez-Achury J, Kanduri K, et al. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. *Hum Mol Genet* 2015;**24**(2):397-409.
52. Smyth DJ, Plagnol V, Walker NM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 2008;**359**(26):2767-77.
53. Akobeng AK, Thomas AG. Systematic review: tolerable amount of gluten for people with coeliac disease. *Aliment Pharmacol Ther* 2008;**27**(11):1044-52.
54. Wieser H. The precipitating factor in coeliac disease. *Baillieres Clin Gastroenterol* 1995;**9**(2):191-207.
55. Marasco G, Di Biase AR, Schiumerini R, et al. Gut Microbiota and Celiac Disease. *Dig Dis Sci* 2016;**61**(6):1461-72.
56. Rostami Nejad M, Ishaq S, Al Dulaimi D, et al. The role of infectious mediators and gut microbiome in the pathogenesis of celiac disease. *Arch Iran Med* 2015;**18**(4):244-9.
57. Tjon JM, van Bergen J, Koning F. Celiac disease: how complicated can it get? *Immunogenetics* 2010;**62**(10):641-51.
58. Floch MH. Topography of the Small Intestine. In: Floch MH, Floch NR, Pitchumoni CS, et al., eds. *Netter's Gastroenterology*. 2nd ed. Philadelphia, PA: Saunders Elsevier, 2010:235-36.
59. Dickey W, Hughes DF. Histology of the terminal ileum in coeliac disease. *Scand J Gastroenterol* 2004;**39**(7):665-7.
60. Floch MH. Duodenal Mucosa and Duodenal Structures. In: Floch MH, Floch NR, Pitchumoni CS, et al., eds. *Netter's Gastroenterology*. 2nd ed. Philadelphia, PA: Saunders Elsevier, 2010:93-94.
61. Floch MH. Gross Structure of the Small Intestine. In: Floch MH, Floch NR, Pitchumoni CS, et al., eds. *Netter's Gastroenterology*. 2nd ed. Philadelphia, PA: Saunders Elsevier, 2010:237-38.
62. Adler DG, Farraye FA, Crawford JM. Gastrointestinal Tract Endoscopic and Tissue Processing Techniques and Normal Histology. In: Odze RD, Goldblum JR, eds. *Odze and Goldblum Surgical Pathology of the GI Tract, Liver, Biliary Tract and Pancreas*. 3rd ed. Philadelphia, PA: Elsevier Saunders, 2015:4-33.e3.
63. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol* 2009;**71**:241-60.

64. Crawley SW, Mooseker MS, Tyska MJ. Shaping the intestinal brush border. *J Cell Biol* 2014;**207**(4):441-51.
65. Bao F, Green PH, Bhagat G. An update on celiac disease histopathology and the road ahead. *Archives of pathology & laboratory medicine* 2012;**136**(7):735-45.
66. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992;**102**(1):330-54.
67. Rostami K, Kerckhaert J, von Blomberg BM, et al. SAT and serology in adult coeliacs, seronegative coeliac disease seems a reality. *Neth J Med* 1998;**53**(1):15-9.
68. Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *European journal of gastroenterology & hepatology* 1999;**11**(10):1185-94.
69. Adelman DC, Murray J, Wu TT, et al. Measuring Change In Small Intestinal Histology In Patients With Celiac Disease. *Am J Gastroenterol* 2018;**113**(3):339-47.
70. Shan L, Molberg O, Parrot I, et al. Structural basis for gluten intolerance in celiac sprue. *Science* 2002;**297**(5590):2275-9.
71. Di Sabatino A, Vanoli A, Giuffrida P, et al. The function of tissue transglutaminase in celiac disease. *Autoimmunity reviews* 2012;**11**(10):746-53.
72. Meresse B, Malamut G, Cerf-Bensussan N. Celiac disease: an immunological jigsaw. *Immunity* 2012;**36**(6):907-19.
73. Pelaseyed T, Bergstrom JH, Gustafsson JK, et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunological reviews* 2014;**260**(1):8-20.
74. Barone MV, Zimmer KP. Endocytosis and transcytosis of gliadin peptides. *Mol Cell Pediatr* 2016;**3**(1):8.
75. Fasano A. Intestinal permeability and its regulation by zonulin: diagnostic and therapeutic implications. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2012;**10**(10):1096-100.

76. Lundin KE, Scott H, Hansen T, et al. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J Exp Med* 1993;**178**(1):187-96.
77. Ferguson A, Murray D. Quantitation of intraepithelial lymphocytes in human jejunum. *Gut* 1971;**12**(12):988-94.
78. Abadie V, Jabri B. IL-15: a central regulator of celiac disease immunopathology. *Immunological reviews* 2014;**260**(1):221-34.
79. Boulpaep EL. Blood. In: Boron WF, Boulpaep EL, eds. *Medical Physiology*. 3rd ed. Philadelphia, PA: Elsevier Inc., 2017:429-46.e1.
80. Brodin P, Davis MM. Human immune system variation. *Nat Rev Immunol* 2017;**17**(1):21-29.
81. Giersiepen K, Lelgemann M, Stuhldreher N, et al. Accuracy of diagnostic antibody tests for coeliac disease in children: summary of an evidence report. *J Pediatr Gastroenterol Nutr* 2012;**54**(2):229-41.
82. Dieterich W, Ehnis T, Bauer M, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 1997;**3**(7):797-801.
83. Anderson RP, van Heel DA, Tye-Din JA, et al. T cells in peripheral blood after gluten challenge in coeliac disease. *Gut* 2005;**54**(9):1217-23.
84. Baydoun A, Maakaron JE, Halawi H, et al. Hematological manifestations of celiac disease. *Scand J Gastroenterol* 2012;**47**(12):1401-11.
85. Jericho H, Guandalini S. Extra-Intestinal Manifestation of Celiac Disease in Children. *Nutrients* 2018;**10**(6).
86. Ensari A. Gluten-sensitive enteropathy (celiac disease): controversies in diagnosis and classification. *Archives of pathology & laboratory medicine* 2010;**134**(6):826-36.
87. Marsh MN, Johnson MW, Rostami K. Mucosal histopathology in celiac disease: a rebuttal of Oberhuber's sub-division of Marsh III. *Gastroenterology and hepatology from bed to bench* 2015;**8**(2):99-109.
88. Falth-Magnusson K, Jansson G, Stenhammar L, et al. Serum food antibodies analyzed by enzyme-linked immunosorbent assay (ELISA) and diffusion-in-gel (DIG)-ELISA methods in children with and without celiac disease. *J Pediatr Gastroenterol Nutr* 1994;**18**(1):56-62.
89. Werkstetter KJ, Korponay-Szabo IR, Popp A, et al. Accuracy in Diagnosis of Celiac Disease Without Biopsies in Clinical Practice. *Gastroenterology* 2017;**153**(4):924-35.

90. Lagerqvist C, Dahlbom I, Hansson T, et al. Antigliadin immunoglobulin A best in finding celiac disease in children younger than 18 months of age. *J Pediatr Gastroenterol Nutr* 2008;**47**(4):428-35.
91. Grodzinsky E, Jansson G, Skogh T, et al. Anti-endomysium and anti-gliadin antibodies as serological markers for coeliac disease in childhood: a clinical study to develop a practical routine. *Acta Paediatr* 1995;**84**(3):294-8.
92. Collin P, Maki M, Keyrilainen O, et al. Selective IgA deficiency and coeliac disease. *Scand J Gastroenterol* 1992;**27**(5):367-71.
93. Ferrara F, Quaglia S, Caputo I, et al. Anti-transglutaminase antibodies in non-coeliac children suffering from infectious diseases. *Clin Exp Immunol* 2010;**159**(2):217-23.
94. Parkkola A, Harkonen T, Ryhanen SJ, et al. Transglutaminase antibodies and celiac disease in children with type 1 diabetes and in their family members. *Pediatr Diabetes* 2018;**19**(2):305-13.
95. Ravelli A, Villanacci V, Monfredini C, et al. How patchy is patchy villous atrophy?: distribution pattern of histological lesions in the duodenum of children with celiac disease. *Am J Gastroenterol* 2010;**105**(9):2103-10.
96. Freeman HJ. Emerging drugs for celiac disease. *Expert Opin Emerg Drugs* 2015;**20**(1):129-35.
97. Silvester JA, Kurada S, Szwajcer A, et al. Tests for Serum Transglutaminase and Endomysial Antibodies Do Not Detect Most Patients With Celiac Disease and Persistent Villous Atrophy on Gluten-free Diets: a Meta-analysis. *Gastroenterology* 2017;**153**(3):689-701 e1.
98. Leonard MM, Sapone A, Catassi C, et al. Celiac Disease and Nonceliac Gluten Sensitivity: A Review. *JAMA : the journal of the American Medical Association* 2017;**318**(7):647-56.
99. Hyland EM, Patton JR. Regulation of Gene Expression: Basic Mechanisms. In: Baynes JW, Dominiczak MH, eds. *Medical Biochemistry*. 5th ed: Elsevier Limited, 2019:303-17.
100. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annu Rev Med* 2005;**56**:303-20.
101. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;**45**(6):580-5.

102. Castellanos-Rubio A, Santin I, Martin-Pagola A, et al. Long-term and acute effects of gliadin on small intestine of patients on potentially pathogenic networks in celiac disease. *Autoimmunity* 2010;**43**(2):131-9.
103. Diosdado B, Wapenaar MC, Franke L, et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 2004;**53**(7):944-51.
104. Juuti-Uusitalo K, Maki M, Kaukinen K, et al. cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. *Journal of autoimmunity* 2004;**22**(3):249-65.
105. Bracken S, Byrne G, Kelly J, et al. Altered gene expression in highly purified enterocytes from patients with active coeliac disease. *BMC Genomics* 2008;**9**:377.
106. Quinn EM, Coleman C, Molloy B, et al. Transcriptome Analysis of CD4+ T Cells in Coeliac Disease Reveals Imprint of BACH2 and IFNgamma Regulation. *PLoS one* 2015;**10**(10):e0140049.
107. Castellanos-Rubio A, Bilbao JR. Profiling Celiac Disease-Related Transcriptional Changes. *Int Rev Cell Mol Biol* 2018;**336**:149-74.
108. Tuttüren AEV, Dorum S, Clancy T, et al. Characterization of the Small Intestinal Lesion in Celiac Disease by Label-Free Quantitative Mass Spectrometry. *Am J Pathol* 2018;**188**(7):1563-79.
109. Hogberg L, Webb C, Falth-Magnusson K, et al. Children with screening-detected coeliac disease show increased levels of nitric oxide products in urine. *Acta Paediatr* 2011;**100**(7):1023-7.
110. Comino I, Real A, Vivas S, et al. Monitoring of gluten-free diet compliance in celiac patients by assessment of gliadin 33-mer equivalent epitopes in feces. *Am J Clin Nutr* 2012;**95**(3):670-7.
111. Simone R, Brizzolara R, Chiappori A, et al. A functional soluble form of CTLA-4 is present in the serum of celiac patients and correlates with mucosal injury. *Int Immunol* 2009;**21**(9):1037-45.
112. Frisullo G, Nociti V, Iorio R, et al. T-bet and pSTAT-1 expression in PBMC from coeliac disease patients: new markers of disease activity. *Clin Exp Immunol* 2009;**158**(1):106-14.
113. Derikx JP, Vreugdenhil AC, Van den Neucker AM, et al. A pilot study on the noninvasive evaluation of intestinal damage in celiac disease using I-FABP and L-FABP. *Journal of clinical gastroenterology* 2009;**43**(8):727-33.

114. Manavalan JS, Hernandez L, Shah JG, et al. Serum cytokine elevations in celiac disease: Association with disease presentation. *Hum Immunol* 2009.
115. Cakal B, Beyazit Y, Koklu S, et al. Elevated adenosine deaminase levels in celiac disease. *J Clin Lab Anal* 2010;**24**(5):323-6.
116. Bragde H, Jansson U, Jarlsfelt I, et al. Gene expression profiling of duodenal biopsies discriminates celiac disease mucosa from normal mucosa. *Pediatric research* 2011;**69**(6):530-7.
117. Planas R, Pujol-Autonell I, Ruiz E, et al. Regenerating gene Ialpha is a biomarker for diagnosis and monitoring of celiac disease: a preliminary study. *Translational research : the journal of laboratory and clinical medicine* 2011;**158**(3):140-5.
118. Adriaanse MP, Tack GJ, Passos VL, et al. Serum I-FABP as marker for enterocyte damage in coeliac disease and its relation to villous atrophy and circulating autoantibodies. *Aliment Pharmacol Ther* 2013;**37**(4):482-90.
119. Moron B, Verma AK, Das P, et al. CYP3A4-catalyzed simvastatin metabolism as a non-invasive marker of small intestinal health in celiac disease. *Am J Gastroenterol* 2013;**108**(8):1344-51.
120. Voros P, Sziksz E, Himer L, et al. Expression of PARK7 is increased in celiac disease. *Virchows Arch* 2013;**463**(3):401-8.
121. Russo F, Chimienti G, Clemente C, et al. Adipokine profile in celiac patients: differences in comparison with patients suffering from diarrhea-predominant IBS and healthy subjects. *Scand J Gastroenterol* 2013;**48**(12):1377-85.
122. Galatola M, Izzo V, Cielo D, et al. Gene Expression Profile of Peripheral Blood Monocytes: A Step towards the Molecular Diagnosis of Celiac Disease? *PLoS one* 2013;**8**(9):e74747.
123. Bondar C, Araya RE, Guzman L, et al. Role of CXCR3/CXCL10 axis in immune cell recruitment into the small intestine in celiac disease. *PLoS one* 2014;**9**(2):e89068.
124. Basso MS, Capriati T, Goffredo BM, et al. Citrulline as marker of atrophy in celiac disease. *Intern Emerg Med* 2014;**9**(6):705-7.
125. Viitasalo L, Niemi L, Ashorn M, et al. Early microbial markers of celiac disease. *Journal of clinical gastroenterology* 2014;**48**(7):620-4.
126. Marafini I, Di Sabatino A, Zorzi F, et al. Serum regenerating islet-derived 3-alpha is a biomarker of mucosal enteropathies. *Aliment Pharmacol Ther* 2014;**40**(8):974-81.

127. Bragde H, Jansson U, Fredrikson M, et al. Potential blood-based markers of celiac disease. *BMC gastroenterology* 2014;**14**:176.
128. Monten C, Torinsson Naluai A, Agardh D. Role of proneurotensin as marker of paediatric coeliac disease. *Clin Exp Immunol* 2016;**186**(3):387-92.
129. Yablecovitch D, Oren A, Ben-Horin S, et al. Soluble Syndecan-1: A Novel Biomarker of Small Bowel Mucosal Damage in Children with Celiac Disease. *Dig Dis Sci* 2017;**62**(3):755-60.
130. Manti S, Cuppari C, Tardino L, et al. HMGB1 as a new biomarker of celiac disease in children: A multicenter study. *Nutrition* 2017;**37**:18-21.
131. Choung RS, Murray JA, Marietta EV, et al. Serum alkylresorcinols as biomarkers of dietary gluten exposure in coeliac disease. *Aliment Pharmacol Ther* 2017;**45**(5):643-52.
132. Comincini S, Manai F, Meazza C, et al. Identification of Autophagy-Related Genes and Their Regulatory miRNAs Associated with Celiac Disease in Children. *Int J Mol Sci* 2017;**18**(2).
133. Adriaanse MPM, Mubarak A, Riedl RG, et al. Progress towards non-invasive diagnosis and follow-up of celiac disease in children; a prospective multicentre study to the usefulness of plasma I-FABP. *Scientific reports* 2017;**7**(1):8671.
134. Galatola M, Cielo D, Panico C, et al. Presymptomatic Diagnosis of Celiac Disease in Predisposed Children: The Role of Gene Expression Profile. *J Pediatr Gastroenterol Nutr* 2017;**65**(3):314-20.
135. Piatek-Guziewicz A, Ptak-Belowska A, Przybylska-Felus M, et al. Intestinal parameters of oxidative imbalance in celiac adults with extraintestinal manifestations. *World journal of gastroenterology : WJG* 2017;**23**(44):7849-62.
136. Thatcher SA. Nucleic Acid Isolation. In: Rifai N, Horvath AR, Wittwer CT, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 6th ed. St. Louis, Missouri: Elsevier, 2018:958-58.e12.
137. Buckingham L, Flaws ML. Quality Assurance and Quality Control in the Molecular Laboratory. *Molecular Diagnostics: fundamentals, methods, and clinical applications*. Philadelphia, PA: F. A. Davis Company, 2007:403-25.
138. Duale N, Brunborg G, Ronningen KS, et al. Human blood RNA stabilization in samples collected and transported for a large biobank. *BMC research notes* 2012;**5**:510.

139. Wittwer CT, Makrigiorgos GM. Nucleic Acid Techniques. In: Rifai N, Horvath AR, Wittwer CT, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 6th ed. St. Louis, Missouri: Elsevier, 2018:959-94.
140. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000;**25**(2):169-93.
141. Bjorkman J, Svec D, Lott E, et al. Differential amplicons (DeltaAmp)-a new molecular method to assess RNA integrity. *Biomol Detect Quantif* 2016;**6**:4-12.
142. Chapman JR, Waldenstrom J. With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies. *PLoS one* 2015;**10**(11):e0141853.
143. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**(4):402-8.
144. Dervan A, Shendure J. The State of Whole-Genome Sequencing. In: Ginsburg GS, Willard HF, eds. *Genomic and Precision Medicine*. 3rd ed: Elsevier Inc., 2017.
145. Kamps R, Brandao RD, Bosch BJ, et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int J Mol Sci* 2017;**18**(2).
146. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016;**17**:13.
147. Mastrokoulis A, den Dunnen JT, van Ommen GB, et al. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics* 2012;**13**:28.
148. Quinn EM, McManus R. Quality Control and Analysis of NGS RNA Sequencing Data. *Methods in molecular biology* 2015;**1326**:217-32.
149. Bi R, Liu P. Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments. *BMC bioinformatics* 2016;**17**:146.
150. Williams CR, Baccarella A, Parrish JZ, et al. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC bioinformatics* 2016;**17**:103.
151. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**(1):15-21.

152. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;**14**(4):R36.
153. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* 2013;**41**(10):e108.
154. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;**11**(3):R25.
155. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in bioinformatics* 2015;**16**(1):59-70.
156. Kricka LJ, Park JY. Immunochemical Techniques. In: Rifai N, Horvath AR, Wittwer CT, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 6th ed. St. Louis, Missouri: Elsevier, 2018:348-67.
157. Aoyagi K, Ashihara Y, Kasahara Y. Immunoassays and Immunochemistry. In: McPherson RA, Pincus MR, eds. *Henry's Clinical Diagnosis and Management by Laboratory Methods*. 23rd ed. St. Louis, Missouri: Elsevier Inc., 2017:862-89.
158. Lavant EH, Carlson J. HLA DR-DQ genotyping by capillary electrophoresis for risk assessment for celiac disease. *Methods in molecular biology* 2013;**919**:297-307.
159. Shen G-Q, Abdullah KG, Wang QK. The TaqMan Method for SNP Genotyping. In: Komar AA, ed. *Single Nucleotide Polymorphisms: Methods and Protocols*. Totowa, NJ: Humana Press, 2009:293-306.
160. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 1995;**57**(1):289-300.
161. Lasky-Su J. Statistical Techniques for Genetic Analysis. In: Robertson D, Williams GH, eds. *Clinical and Translational Science*. 2nd ed. London, United Kingdom: Elsevier Inc., 2017:347-62.
162. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003;**13**(11):2498-504.
163. Merico D, Isserlin R, Stueker O, et al. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one* 2010;**5**(11):e13984.

164. Oesper L, Merico D, Isserlin R, et al. WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source code for biology and medicine* 2011;**6**:7.
165. Kucera M, Isserlin R, Arkhangorodsky A, et al. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Res* 2016;**5**:1717.
166. Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006;**7**:3.
167. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**(7):RESEARCH0034.
168. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* 2004;**64**(15):5245-50.
169. Lavant EH, Carlson JA. A new automated human leukocyte antigen genotyping strategy to identify DR-DQ risk alleles for celiac disease and type 1 diabetes mellitus. *Clin Chem Lab Med* 2009;**47**(12):1489-95.
170. Lavant EH, Agardh DJ, Nilsson A, et al. A new PCR-SSP method for HLA DR-DQ risk assessment for celiac disease. *Clinica chimica acta; international journal of clinical chemistry* 2011;**412**(9-10):782-4.
171. Bragde H, Jansson U, Fredrikson M, et al. Celiac disease biomarkers identified by transcriptome analysis of small intestinal biopsies. *Cell Mol Life Sci* 2018;**75**(23):4385-401.
172. Law CW, Chen Y, Shi W, et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;**15**(2):R29.
173. Sangineto M, Graziano G, D'Amore S, et al. Identification of peculiar gene expression profile in peripheral blood mononuclear cells (PBMC) of celiac patients on gluten free diet. *PloS one* 2018;**13**(5):e0197915.
174. Ciccocioppo R, Panelli S, Conti Bellocchi MC, et al. The Transcriptomic Analysis of Circulating Immune Cells in a Celiac Family Unveils Further Insights Into Disease Pathogenesis. *Front Med (Lausanne)* 2018;**5**:182.
175. Hong G, Li H, Li M, et al. A simple way to detect disease-associated cellular molecular alterations from mixed-cell blood samples. *Briefings in bioinformatics* 2018;**19**(4):613-21.

176. Bondar G, Cadeiras M, Wisniewski N, et al. Comparison of whole blood and peripheral blood mononuclear cell gene expression for evaluation of the perioperative inflammatory response in patients with advanced heart failure. *PloS one* 2014;**9**(12):e115097.
177. Feezor RJ, Baker HV, Mindrinos M, et al. Whole blood and leukocyte RNA isolation for gene expression analyses. *Physiol Genomics* 2004;**19**(3):247-54.
178. Goods BA, Vahey JM, Steinschneider AF, et al. Blood handling and leukocyte isolation methods impact the global transcriptome of immune cells. *BMC Immunol* 2018;**19**(1):30.
179. Pascual V, Dieli-Crimi R, Lopez-Palacios N, et al. Inflammatory bowel disease and celiac disease: overlaps and differences. *World journal of gastroenterology : WJG* 2014;**20**(17):4846-56.
180. Hardee S, Alper A, Pashankar DS, et al. Histopathology of duodenal mucosal lesions in pediatric patients with inflammatory bowel disease: statistical analysis to identify distinctive features. *Pediatr Dev Pathol* 2014;**17**(6):450-4.
181. Medrano LM, Pascual V, Bodas A, et al. Expression patterns common and unique to ulcerative colitis and celiac disease. *Annals of human genetics* 2018.
182. Gurwitz D. Expression profiling: a cost-effective biomarker discovery tool for the personal genome era. *Genome Med* 2013;**5**(5):41.
183. Guan Q, Yan H, Chen Y, et al. Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer. *BMC Genomics* 2018;**19**(1):99.
184. Yu X, Zhang J, Sun S, et al. Individual-specific edge-network analysis for disease prediction. *Nucleic acids research* 2017;**45**(20):e170.
185. Haider S, Yao CQ, Sabine VS, et al. Pathway-based subnetworks enable cross-disease biomarker discovery. *Nat Commun* 2018;**9**(1):4746.

Supplementary material

Supplementary materials for papers I–III have been made available by the respective journal. Supplementary material for paper IV will be made available upon request to the corresponding author.

Paper I

<http://links.lww.com/PDR/A69>

Paper II

<http://www.biomedcentral.com/1471-230X/14/176>

Paper III

<https://doi.org/10.1007/s00018-018-2898-5>

Paper IV

Supplementary material will be made available upon request to the corresponding author.

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-156088>

FACULTY OF MEDICINE AND HEALTH SCIENCES

Linköping University Medical Dissertation No. 1672, 2019
Department of Clinical and Experimental Medicine

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se