

**Gene expression analysis approaches to study barrier  
dysfunction in celiac disease and pathogenesis of  
colitis-associated cancer**

A Dissertation

Submitted in Partial Fulfilment of the Requirements  
for the Degree of

Doctor rerum naturalium (Dr. rer. nat.)

to the Department of Biology, Chemistry, Pharmacy  
of Freie Universität Berlin

by

**Danielle Cardoso da Silva**

Berlin, 2021

Supervisor: Prof. Dr. Michael Hummel

Second examiner: Prof. Dr. Volker Haucke

Date of defense: 06/09/2021

To my parents

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Dr. Hummel for making my coming to Berlin possible by choosing me as his doctorate student and for his support on the writing of this thesis. I would also like to thank Prof. Dr. Volker Haucke for accepting the invitation to review my thesis. I am grateful for Prof. Dr. Britta Siegmund for being my co-supervisor and giving me important feedback on my work over the years. I am also grateful for PD. Dr. Michael Schumann for supervising me throughout my PhD studies and for arranging a place for me at his lab. It was a pleasure to work with his team.

I thank the Klinische Physiologie group, Prof. Dr. Michael Fromm and Prof. Dr. Jörg-Dieter Schulzke for welcoming me to the group and for all the support. Especial thanks to Dr. Rita Rosenthal and Prof. Dr. Dorothee Günzel for the support and the willingness to help with scientific matters and any other help that I have ever needed. Moving to a foreign country is not easy and I would not have made it without you. Thanks to In-fah Lee, Anja Fromm and Britta Jebautzke for the indispensable work and the help.

I would like to thank my colleagues from AG Schumann: Claudia Heldt, Dr. Federica Branchii Deborah Delbue, Jacob Wiese and Daphni Siampali for being the best coworkers and personal friends. Thanks for the conversations, for the support, the fun, the advice, the fellowship, and the encouragement. I also thank the new members of the group Subhakankha Manna for doing the blinded analysis and Violaine Dony for the help despite the short time we worked together.

I thank my collaborators from the Department of Genetics of the University of Groningen Dr. Iris Jonkers, Dr. Sebo Withoff, Renée Moerkers and Joram Mooiweer for the development of the Caco-2 KO clones, the RNA-Seq analysis and the exchange of experience. I also thank my collaborator Dr. January Weiner from the CUBI BIH for the analysis of the RNA-Seq of the

intestinal cells exposed to OPN. I thank Dr. Anja Kühn for performing the IHC and for the productive discussions of the results we got. I thank Maximilian Sehn for sharing the work on the colitis-associated cancer project.

I also thank the people from AG Siegmund and AG Hummel for helping in the development of this work. Especial thanks to Hedwig Lammert with the Nanostring RNA analysis and to Dr. Rainer Glauben for integrating me in the seminars and lab meetings and helping with the FACS analysis.

I thank my colleagues of the first cohort of the GRK2318 TJ-Train for being great colleagues, making a good work environment and sharing good times.

In a personal level I have many people to be grateful for.

I thank God for all the great things I have experienced in my life. The care, the mercy, the hope in my darkest hour. I only made it to here through His grace and favor. I am grateful for my parents and my brother, for the unconditional love and support to pursuing my education and for being as present as possible despite the huge distance in between us. Thank you for believing in me and praying for me.

I am thankful for the friends I made here in Germany. When one is far from home, friends become family, and can I say that I made a new family for myself. In this regard, I dedicate a very special thanks to Deborah Delbue for being my best friend and partner in this PhD journey far away from home. You were the greatest surprise, and the best gift God has given me in my moving to Germany. We have survived these years by holding each other's hand throughout all the challenges and difficulties. Thank you for the friendship, for staying with me in the highs and lows, for we collected so many great memories from our travels and happy hours after long days of work. Thanks to Veronica Melo Costa for the three of us supported one another through the hardships. Thank you for being so special, for sharing your house, your parents, and helping

whenever you could. Thanks to Gabriel, boy prodigy, for sharing great and unforgettable moments in our travels. Thank you for your friendship.

Special thanks to Ale Arias for being the best neighbor, friend, and jogging partner. Thank you for being so thoughtful and true, for the conversations and the night-outs. Thanks to Carlos Ayala-Torres for being the best person ever! Thank you for the friendship, the honesty, the good vibes. You are the best! And thanks to Edgar Dolores for being there for us, for the friendship, for your company and for the best guacamole ever!

I thank my friends in Brazil, who have maintained our friendship despite the distance. Special thanks to Muhammed Manasfi for the immense help, both professionally and personally. I do not have words to thank you. Thanks for being there when I most needed. Thanks to Flavia Vasconcelos for supporting me.

Thanks to Fa Stollenwerk for the professional help and support. You were fundamental for my success in this enterprise.

Finally, I would like to give special thanks to my partner Jose Mayala. For being the companion in my day-by-day struggles, for incentivizing me to fight and not give up. Thanks for supporting, believing in me and taking care of me.

## **ABSTRACT**

### **TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE**

Celiac Disease (CeD) is an autoimmune disease that develops in genetically predisposed individuals after the ingestion of gluten. It induces a malabsorption syndrome, commonly provoking diarrhea, weight loss and vitamin deficiency and the only standard treatment so far is a gluten-free diet. Celiac patients present impaired epithelial barrier function with lower TEER and increased permeability to disaccharides. In addition, tight junction strands are discontinuous and decreased in celiac patients. Changes in barrier function are mostly attributed to the immune process, however, it was shown that treated patients may present impaired barrier function, despite the lack of symptoms. Moreover, risk loci for CeD were found in genes related to cell-cell adhesion, including LPP and C1orf106. LPP is involved in focal adhesions formation and E-cadherin cell-cell adhesion. C1orf106 inhibits the degradation of E-cadherin indirectly and its depletion causes reduction in TEER.

In this context, we sought out to study the effect of LPP and C1orf106 in barrier function in intestinal cell lines and in patients with CeD. Our results show that cells depleted of LPP and C1orf106 present changes in tight junction protein content and present a reduced ability to re-assemble the tight junctions after a calcium switch assay. In patients, we did not see significant changes regarding LPP or C1orf106 protein content, but further analysis of electrical resistance and RNA may provide further insights into the importance of both proteins in the barrier impairment in CeD.

### **THE ROLE OF OSTEOPONTIN IN THE PATHOGENESIS OF CAC**

IBD patients present an increased risk of developing colorectal cancer, namely colitis-associated cancer (CAC). In the case of CAC, the immune response in the IBD plays an important role in tumorigenesis and results in a different progression process than sporadic colorectal carcinoma (CRC). For example, the early mutation of APC seen in CRC does not occur as

frequently in CAC and when it does, it is only at the final stages of progressions. On the other hand, p53 mutations occur very early in CAC progression, whereas in CRC it is a late finding. CAC pathogenesis is not as well understood as CRC and there is still much to clarify regarding CAC progression.

Then, we decided to study CAC progression in samples from patients who underwent colectomy and performed an RNA analysis for immune-related genes. The results of this experiment showed osteopontin (OPN) as the most upregulated gene in both CAC coming from ulcerative colitis and Crohn's disease patients, impelling us to investigate it further. OPN was found in both epithelial cells and stromal cells and one of its receptors, CD44, was also identified in both cell compartments, with a tendency for being increased in CAC epithelial cells. OPN is known to promote tumorigenesis in several cancer types, especially solid tumors, and one of its key functions is the induction of epithelial to mesenchymal transition (EMT). Indeed, findings from the RNA analysis and immunohistochemical analysis of the patients' samples point out to the presence of the EMT process in CAC. When we sought out to study OPN effects in cell lines, OPN activated ERK1/2, but not STAT3, AKT or P-65/NFκB. However, we failed to reproduce EMT by exposing the cells to OPN. Finally, we decided to perform an RNA-Seq analysis of the cells treated with OPN and found changes in mitochondrial respiratory chain, especially downregulation of complexes III and IV. These results suggest a new role for OPN in the tumorigenesis of CAC.

## **ZUSAMMENFASSUNG**

### **PRIMÄRER BARRIEREDEFEKT BEI ZÖLIAKIE**

Die Zöliakie ist eine Autoimmunerkrankung, die nach Glutenaufnahme bei genetisch Prädisponierten entsteht. Sie verursacht ein Malabsorptionssyndrom, das üblicherweise mit Diarrhö, Gewichtsverlust und Vitaminmangelercheinungen einhergeht und bei der die bislang einzig zur Verfügung stehende Behandlung die glutenfreie Diät ist. Zöliakie-Patienten weisen eine defekte epitheliale Barrierefunktion auf, die sich durch verminderte transepitheliale Widerstände und eine erhöhte mukosale Permeabilität für Disaccharide einhergeht. Entsprechend sind die epithelialen Tight Junction- (TJ-)Stränge bei Zöliakie unterbrochen bzw. in ihrer Zahl reduziert. Die Veränderungen der Barrierefunktion wurden bislang immer auf die der Zöliakie zugrunde liegenden Immunreaktion zurückgeführt. In diesem Sinne wurden sie als sekundär zur mit distinkten Zytokinsekretion einhergehenden T-Zellreaktion interpretiert. Allerdings gibt es *ex vivo* Daten zur Barrierefunktion behandelter Zöliakie-Patienten, die trotz Therapie weiterhin eine defizitäre Barrierefunktion aufweisen. Außerdem wurden Zöliakie-Risiko-Genloci identifiziert, die mit der interepithelialen Adhäsion verbunden sind, insbesondere die Gene LPP und C1orf106. LPP wurde beschrieben in Zusammenhang mit der Ausbildung von Focal Adhesions beschrieben und weiterhin mit der Ausbildung E-Cadherin-abhängiger Interzellularbrücken. C1orf106 hemmt den Abbau von E-Cadherin. Es ist zudem bekannt, dass die Verminderung von C1orf106 einen Barrieredefekt verursacht. In diesem Zusammenhang begannen wir eine Studie, die das Ziel hatte, die Effekte von LPP und C1orf106 auf die Barrierefunktion von intestinalen Epithelzellen zu untersuchen. Ergebnisse dieser Studie beinhalten u.a., dass ein Knock-out von LPP oder C1orf106 mit Veränderung der TJ-Proteinkomposition einhergeht und, dass die Assemblierung von TJ dysfunktional ist. Bei Patienten mit Zöliakie fanden wir zwar keine signifikanten Proteinmengenänderungen für LPP oder C1orf106. Es kann aber sein, dass weitergehende funktionelle Barriere- und RNA-Untersuchungen einen genaueren Einblick in die Bedeutung der beiden Proteine für die Barrierefunktion bei Zöliakie ermöglichen.

## **DIE ROLLE VON OSTEOPONTIN IN DER PATHOGENESE DES COLITIS-ASSOZIIERTEN KARZINOMS**

Bei Patienten mit chronischen entzündlichen Darmerkrankungen (CED) besteht ein höheres Risiko, ein Colitis-assoziiertes Karzinom zu entwickeln (CAC). Dabei ist auszugehen, dass die CED-assoziierte mukosale Immunantwort bei CED eine große Rolle spielt. Es ist inzwischen klar, dass der Prozess der CAC-Karzinomgenese sich deutlich von dem des sporadischen Kolorektales Karzinoms (CRC) unterscheidet. Dazu gehört, dass die beim CRC sehr früh auftretende Mutation im APC-Gens nicht oder nur sehr spät in der Karzinogenese des CAC auftritt. Auf der anderen Seite treten p53 Mutation sehr viel frühzeitiger beim CAC als beim CRC auf. Insgesamt ist die CAC-Pathogenese weitestgehend unverstanden. Wir begannen daher eine Studie zur Aufklärung der Karzinogenese-Mechanismus bei CAC unter Verwendung von chirurgischen Resektaten von CAC-Patienten (Kolektomiepräparate), isolierten RNA und führten eine Expressionsanalyse von Genen durch, die mit dem mukosalen Immunsystem assoziiert sind. Resultate dieser Expressionsanalyse ergaben, dass Osteopontin (OPN) das am stärksten hochregulierte Gen sowohl bei CAC auf dem Boden einer Colitis ulcerosa als auch bei CAC auf dem Boden eines Morbus Crohns ist. Immunhistochemisch konnte OPN sowohl in Epithelzellen als auch im Stromazellen identifiziert werden. CD44, einer der OPN-Rezeptoren, wurde ebenfalls in beiden Kompartimenten gefunden. Zudem ergab sich eine Tendenz für eine höhere Expression in Epithelzellen. Für OPN ist bekannt, dass es die Tumorigenese verschiedener Karzinomtypen unterstützt. Zudem ist es einer der zentralen Induktoren der Epithelial-zu-mesenchymalen Transition (EMT). Damit in Einklang ergab die Expressionsanalyse sowie auch die immunhistochemische Analyse der CAC-Patientenproben den Nachweis EMT-spezifischer Genexpression bei CAC. In dieser Situation wechselten wir auf ein Zelllinien-basiertes System und konnten zeigen, dass OPN ERK1/2 aber nicht STAT3, Akt oder p65/NF $\kappa$ B aktiviert. Allerdings konnten wir nicht die Induktion von EMT durch OPN in den Zelllinien nachweisen. Zuletzt führten wir an Zellen, die mit OPN behandelt worden waren, eine RNA-Seq-Analyse durch und konnten

zeigen, dass OPN Veränderungen der Gene der mitochondrialen Atemkette verursacht, insbesondere der Komplexen III und IV. Diese Resultate legen eine neue Rolle für Osteopontin in der CAC-Tumorigenese nahe.

## Table of Contents

Acknowledgements.....	I
Abstract.....	IV
TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE.....	IV
THE ROLE OF OSTEOPONTIN IN THE PATHOGENESIS OF CAC .....	IV
Zusammenfassung .....	VI
Primärer Barrieredefekt bei Zöliakie.....	VI
Die Rolle von Osteopontin in der Pathogenese des Colitis-assoziierten Karzinoms .....	VII
List of figures.....	5
List of tables .....	6
List of abbreviations.....	7
1 Introduction .....	13
Intestinal mucosal barrier .....	13
Microbiota.....	13
Mucus layer.....	14
Epithelial cells .....	15
Intraepithelial lymphocytes .....	18
Epithelial cell junctions: the apical junctional complex.....	19
Adherens junctions .....	19
Tight junctions.....	20
Bicellular TJ .....	20
Tricellular TJ .....	20
Claudins.....	20
TJ-associated MARVEL proteins.....	22
Occludin .....	22
MARVEL D3 .....	23
Tricellulin.....	23
Junctional adhesion proteins (JAMs) .....	23
Scaffold proteins .....	23
Tight junctions of the intestine .....	24

The intestinal immune system .....	24
Organized lymphoid structures.....	24
Effector cells.....	25
Mononuclear phagocytes .....	26
Other innate cells.....	27
Diseases in this study .....	27
Celiac disease .....	27
Inflammatory Bowel Disease .....	32
Colitis-associated cancer.....	34
Osteopontin .....	36
Structure and function .....	36
Epithelial to mesenchymal transition .....	40
2    Aim .....	42
TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE.....	42
THE ROLE OF OPN IN THE PATHOGENESIS OF CAC.....	42
3    Materials and methods.....	44
Reagents.....	44
Cell culture .....	44
Establishment of LPP and C1orf106 knock-out Caco2 cell lines .....	44
Characterization of genomic sequence of LPP and C1orf106 knock-out Caco2 cells.....	46
Transepithelial electrical resistance measurements .....	47
Protein extraction from cell lines.....	47
Protein extraction from tissue samples.....	47
Western Blot .....	48
Calcium Switch assay .....	49
Immunofluorescence .....	50
RNA extraction from cell lines .....	51
RNA quantification and cDNA synthesis .....	52
Recruitment of patients and samples collection .....	53
RNA extraction of formalin-fixed-paraffin-embedded samples .....	54
<i>In silico</i> analysis of gene expression data .....	54
Immunohistochemistry of FFPE slides .....	57

OPN exposure of intestinal cell lines in filters .....	57
OPN treatment of intestinal cell lines in plates .....	58
Real time-quantitative PCR.....	58
RNA-seq.....	59
Bioinformatics of the RNA-Seq data .....	60
Statistics .....	60
Devices and consumables .....	60
4 RESULTS.....	65
TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE.....	65
Establishment of knock-out clones of Caco-2 cells.....	65
Characterization of the barrier function in the Caco-2 knock-out clones .....	66
Transepithelial electrical resistance.....	66
Tight junction proteins content .....	66
Tight junction re-assembling capacities of the knock-out clones.....	68
Barrier function of celiac patients.....	69
Patients' characteristics .....	69
THE ROLE OF OPN IN THE PATHOGENESIS OF CAC.....	71
Patients' characteristics .....	71
Gene expression analysis of colitis-associated patients' samples.....	72
OPN might account for a poorer prognosis .....	77
Osteopontin is expressed by epithelial and stromal cells in patients.....	78
Mechanistic network of OPN .....	78
Immunohistochemistry for OPN signaling .....	79
OPN as an EMT-inducing molecule .....	82
Protein analysis of intestinal cell lines exposed to osteopontin.....	82
Analysis of the phosphorylation of P56 after OPN treatment.....	85
mRNA analysis of cell lines after osteopontin incubation .....	86
RNA-Seq of cell lines .....	87
Gene enrichment sets of the RNA-Seq analysis.....	90
5 Discussion.....	95
Tracking a primary barrier dysfunction in celiac disease.....	95
Role of osteopontin in cac .....	98

6	Conclusions .....	105
	TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE.....	105
	THE ROLE OF OPN IN THE PATHOGENESIS OF CAC.....	105
7	Appendix – list of publications.....	107
8	References .....	108
9	Supplementary material .....	130

## LIST OF FIGURES

Figure 1.1. The intestinal crypt.....	18
Figure 1.2. Osteopontin signaling pathways.....	39
Figure 4.1. Protein analysis of Caco-2 knock-out (KO) clones.....	65
Figure 4.2. Transepithelial electrical resistance (TEER) of Caco-2 clones.....	67
Figure 4.3. Protein content of tight junctional proteins.....	68
Figure 4.4. Tight junction re-assembling after calcium depletion and replacement.....	69
Figure 4.5. LPP and C1orf106 protein content in celiac mucosa.....	71
Figure 4.6. Heatmaps of the 20 most up- and downregulated genes in the Nanostring.....	74
Figure 4.7. Osteopontin expression might account for a poorer prognosis.....	77
Figure 4.8. Immunohistochemical analysis of osteopontin in colonic tissue.....	78
Figure 4.9. Osteopontin is predicted as an upstream regulator.....	79
Figure 4.10. Immunohistochemistry of putative osteopontin targets.....	80
Figure 4.11. Immunohistochemistry of putative osteopontin targets.....	81
Figure 4.12. Epithelial to mesenchymal transition in colitis-associated cancer.....	83
Figure 4.13. Western blot protein analysis of intestinal cells exposed to osteopontin.....	84
Figure 4.14. NFκB P-65 phosphorylation after osteopontin exposure.....	85
Figure 4.15. Quantitative PCR evaluation of HT29/B6 for EMT markers.....	86
Figure 4.16. Quantitative PCR evaluation of T84 for EMT markers.....	87
Figure 4.17. RNA-Seq analysis of intestinal cell lines exposed to osteopontin.....	88
Figure 4.18. ROC curves of the enriched gene sets in HT29/B6 cells.....	91
Figure 4.19. ROC curves of enriched gene sets in T84 cells.....	93
Figure 5.1. RNA-Seq analysis of Caco-2 knock-out clones.....	97
Supplementary Figure 1. Densitometric analysis of Caco-2 knock-out clones.....	130
Supplementary Figure 2. Densitometric analysis of tight junctional proteins in Caco-2 knock-out clones.....	131
Supplementary Figure 3. Densitometric analysis of HT29/B6 cells exposed to osteopontin.....	132
Supplementary Figure 4. Densitometric analysis of T84 cells exposed to osteopontin.....	133

## LIST OF TABLES

Table 1.1. Distribution of claudins by organ.....	21
Table 3.1. guideRNA sequences for CRISPR/Cas9-mediated knock-out .....	45
Table 3.2. Primers to test genomic disruption .....	46
Table 3.3. Characterization of genomic sequence.....	48
Table 3.4. Antibodies used for Western blot analysis .....	50
Table 3.5. Antibodies used for Immunofluorescence .....	51
Table 3.6. Reverse transcription mixes .....	52
Table 3.7. Reverse transcription cycles.....	53
Table 3.8. Custom genes added to the Nanostring panel.....	56
Table 3.9. TaqMan Probes .....	58
Table 3.10. Devices .....	61
Table 3.11. Chemicals and kits .....	62
Table 3.12. Consumables .....	63
Table 4.1. Clinical characteristics of celiac patients.....	70
Table 4.2. Clinicopathological characteristics of included patients .....	73
Table 4.3. Most up- and downregulated genes in CDAC vs CD .....	75
Table 4.4. Most up- and downregulated genes in UCAC vs UC .....	76
Table 4.5. Genes differentially expressed in HT29/B6.....	89
Table 4.6. Genes differentially expressed in T84.....	90
Table 4.7. Gene enrichment analysis of HT29/B6 cells exposed to osteopontin for 24 hours....	91
Table 4.8. Gene enrichment analysis of T84 cells exposed to osteopontin for 24 hours.....	92
Table 4.9. Mitochondrial respiratory chain gene sets.....	94

## LIST OF ABBREVIATIONS

ADP	Adenosine Diphosphate
AJ	Adherens Junction
AKT	Ak strain transforming
ANOVA	Analysis of Variance
AOM	Azoxymethane
APC	Adenomatous polyposis coli
ARF6	ADP-ribosylation factor 6
BCA	Bicinchoninic acid
BSP-1	Bone Sialoprotein 1
CAC	Colitis-associated Cancer
CBCs	Columnar Base Cells
CCND1	Cyclin D1
CD	Crohn's Disease
CDAC	Crohn's disease-associated cancer
CDH1	E-cadherin gene
CeD	Celiac Disease
CLDN	Claudin gene
CRB3	Crumbs 3 gene
CRC	Sporadic Colorectal Carcinoma
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRP	C-reactive Protein
DC	Dendritic cell
DGP	Deamidated gliadin peptides

DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DSS	Dextran Sulfate Sodium
EATL	Enteropathy-associated T cell Lymphoma
E-cad	E-cadherin
ECL	Extracellular Loop
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EHEC	Enterohemorrhagic <i>Escherichia coli</i>
EMA	Anti-endomysium
EMT	Epithelial to mesenchymal transition
EpCAM	Epithelial Cellular Adhesion Molecule
ER	Endoplasmic reticulum
ERK	Extracellular-signal-regulated kinase
ESR	Erythrocyte Sedimentation
ETA-1	Early T-lymphocyte Activator 1
FCS	Fetal Calf Serum
FFPE	Formalin-fixed paraffin-embedded
FN1	Fibronectin
FOXP3	Forkhead box P3
GALT	Gut-associated lymphoid tissue
GEF	Guanine Exchange Factor
GF	Growth Factor
GFD	Gluten-free Diet

GFP	Green Fluorescent Protein
GRHL2	Grainyhead-like 2
GTP	Guanosine Triphosphate
GWAS	Genome-wide association studies
H&E	Hematoxylin and Eosin
HIF1 $\alpha$	Hypoxia-inducible factor-1 alpha
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HMG	High Motility Group
HNF	Hepatocyte Nuclear Factor
HRP	Horseradish Peroxidase
IBD	Inflammatory Bowel Disease
IECs	Intestinal Epithelial Cells
IELs	Intraepithelial Lymphocytes
IFN $\gamma$	Interferon gamma
Ig	Immunoglobulin
IL	Interleukin
ILC	Innate Lymphoid Cells
ILDR	Immunoglobulin-like domain containing receptor
ILT	Isolated Lymphoid Tissue
INAVA	Innate Immunity Activator Protein
iNKT	invariant Natural Killer T cells
ISCs	Intestinal Stem Cells
JAM	Junctional Adhesion Proteins
JNK	c-Jun N-terminal kinase 1
kDa	kilodalton

KDR	Kinase Insert Domain Receptor
KO	Knock-out
KRAS	Kirsten rat sarcoma
LPP	Lipoma Preferred Partner
LRCs	label retaining cells
Lrg5	Leucine-rich repeat-containing G-protein coupled receptor 5
Lrig1	Leucine-rich Repeats and Immunoglobulin-like Domain 1
LSR	Lipolysis-stimulated lipoprotein receptor
MAGUK	Membrane-associated Guanylate Kinase
MAIT	Mucosal-associated invariant T cells
MAPK	Mitogen-Activated Protein Kinase
MARVEL	MAL and related proteins for vesicle trafficking and membrane link
MEKK1	Mitogen activated protein kinase kinase 1
MEM	Modified Eagle Medium
MHC	Major Histocompatibility Complex
MICA	MHC class I chain-related protein A
MMP	Metalloproteinase
mTOR	mammalian Target Of Rapamycin
NES	Nuclear Export Signal
NFkB	Nuclear Factor kappa B
NIK	Nuclear factor-Inducing Kinase
NK	Natural Killer
NKG2D	Natural Killer Group 2D
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
OGDHL	Oxoglutarate Dehydrogenase L
Olm4	Olfactomedin 4

OPN	Osteopontin
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PI3K	Phosphatidylinositol 3-kinase
PIGR	Polymeric Ig Receptor
PKC	Protein Kinase C
PLC $\gamma$	Phospholipase C- $\gamma$
Pro	Proline
PSC	Primary Sclerosis Cholangitis
PTS	Phosphotransferase Systems
RCD	Refractory Celiac Disease
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute Medium
SDS	Sodium Dodecyl Sulfate
SED	Subepithelial dome
Ser	Serine
SH3	Src homology 3
SIBLING	Small Integrin Binding Ligand N-linked Glycoprotein
SIgA	Secretory IgA
SMAD	Small body size Mothers Against Decapentaplegic
SNP	Single Nucleotide Polymorphism
SPP1	Secreted Phosphoprotein 1
STAT	signal transducer and activator of transcription
TA	Transit-Amplifying

TAMP	TJ-associated MARVEL
Tcd	Clostridium difficile Toxin
TCR	T Cell Receptor
TEER	Transepithelial Electrical Resistance
TG2	Transglutaminase 2
TGF $\beta$	Transforming Growth Factor $\beta$
Thr	Threonine
TJ	Tight Junction
TNF $\alpha$	Tumor Necrosis Factor alpha
TR1	FOXP3- Treg Type 1
Treg	Regulatory T cell
Tris	Trisaminomethane
UC	Ulcerative Colitis
UCAC	Ulcerative colitis-associated cancer
uPA	urokinase-type Plasminogen Activator
Wnt	Wingless-related integration site
WTS	Whole Transcriptome Sequencing
ZEB	Zinc Finger E-Box Binding Homeobox
ZO	Zonula Occludens

# 1 INTRODUCTION

## INTESTINAL MUCOSAL BARRIER

Organs that interact directly with the external environment such as skin, lung, intestine and kidney nephrons are lined with epithelial cells that have the dual task of protecting the organism from harmful external stimuli while allowing the passage of beneficial agents, such as nutrients. The structure and components of the epithelial barrier vary according to the organ function. For instance, the intestinal epithelial barrier, the object of study of this project, is formed from lumen to the lamina propria by the intestinal microbiota, mucus layer, epithelial cells monolayer and immune cells in the lamina propria (1). Impairment of the intestinal epithelial barrier is related to many diseases from malabsorption syndromes, celiac disease, inflammatory bowel disease, to colorectal cancer. This chapter presents an overview of the components of the intestinal barrier, the importance of tight junctions (TJ) to maintaining the barrier selective function, the diseases studied in this thesis and their relation to an impaired intestinal barrier function.

### Microbiota

Recent estimates report the “standard male” human cell numbers to be  $3.0 \times 10^{13}$  and the resident microbiota cell number,  $3.8 \times 10^{13}$ , resulting on a ratio of 1:1.3. Of note, the colonic microbiota alone corresponds to almost 100% of the bacterial population in the human body (2).

The luminal microbiota composition consists mostly of Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria phyla. Notably, *Bacteroides* is not only the most abundant but also the most variable genus (3). Beyond simply residing in the intestine, the microbiota maintains constant interaction with the host and perform a crucial role in some of the host’s biological processes, namely metabolism of some nutrients and the prevention of pathogenic bacteria colonization (4). As an example of the beneficial role of the microbiota, butyrate producing bacteria such as *Faecalibacterium prausnitzii* positively impact barrier function by reducing

severity of DSS colitis in mice and preventing increase in permeability to Cre-EDTA (5) and negatively correlates with inflammatory markers in patients that undergone bariatric surgery (6).

In an opposite way, dysbiosis of the intestinal microbiota is associated with disruption of the barrier function and disease. *Clostridium difficile* is the leading cause of nosocomial diarrhea and disrupts intestinal barrier function through the secretion of toxins TcdA and TcdB (7). TcdA promotes stronger disruption of the epithelium with the redistribution of Zonula occludens (ZO-1), occludin, E-cadherin, F-actin and tubulin (7). Enterohemorrhagic *Escherichia coli* (EHEC) is responsible for foodborne diarrhea that destroys the mucus layer of the intestine and attaches to the epithelial cells. Once attached, it promotes displacement of occludin from the membrane and increase of claudin-2 with concomitant drop in TEER values (8).

## **Mucus layer**

The mucus layer covers the epithelium as mechanical protection coating against sheer stress and microbial infiltration and maintains intestinal homeostasis. It is quite complex in structure and has a crucial role in the barrier which was however neglected for the first decades of intestinal research due to its being washed off in the first steps of immunohistochemical preparations. The intestinal mucus is a viscoelastic secretion from goblet cells located in the columnar epithelial monolayer that provides a selectively permeable layer for the diffusion and absorption of nutrients (9). Its main structural and functional constituents are the mucins proteins, but account only for about 1-5% of it. The mucus barrier is 90-95% water, 1% electrolytes (NaCl, KCl, NaOH<sub>2</sub>, PO<sub>4</sub><sup>3-</sup>, Mg<sup>+2</sup>, Ca<sup>+2</sup>) lipids 1-2% and other components (10). Mucins are large proteins characterized by specific domain "mucin domain" with multiple repetitive aa sequences in Pro, Thr and Ser, the PTS domain. Thr and Ser are heavily O-glycosylated (10). They are divided into transmembrane and secreted gel-forming mucins. Transmembrane mucins 1, 3, 4, 12, 13, 15, 17 and 21 are expressed in the intestine (11).

The main function of the mucus layer is to protect the epithelium against mechanical, chemical and biological attacks and maintaining intestinal homeostasis. Ancillary non-mucus proteins originated in the interstitial fluid play an important role in mucus defensive function. Among those are found defensive proteins (alpha-, beta-defensins, lysozyme, lactoferrin, statherins), IgA, IgM, GFs, structural proteins (secretory leukocyte proteinase inhibitor, pancreatic secretory trypsin inhibitor) and glycoproteins (12). Defensins are a family of 2-5 kDa antimicrobial peptides produced by Paneth cells and that play a role in innate immunity (13).  $\alpha$ -defensins antiviral capabilities include inhibition of HIV replication, in vitro protection against influenza A, enveloped and non-enveloped viruses and present anti-microbial effects through pore-formation (13,14).  $\beta$ -defensins target bacteria, virus and yeast also have a role in inflammation and fertility (15). Lysozyme is an antimicrobial protein which hydrolyzes bacterial cell wall peptidoglycan. Produced by neutrophils and macrophages (16).

## **Epithelial cells**

Initially thought to be the only and utmost barrier of the intestine, the intestinal epithelial cells (IECs) are currently known to be in communication with all the other components of the barrier in a reciprocal relationship and regulation. The study of the intestinal cells can only be in the light of the complex architecture of the villi and crypts, where stem cells generate the different kinds of IECs which differentiate while migrating towards the tip of the villi.

### *Intestinal architecture*

The small intestine presents a peculiar architecture that allows for maximal surface area to improve nutrient absorption. The villi are finger-like protrusions covered by a monolayer of terminally differentiated epithelial cells and are connected to crypts, which, in opposite, are well-like structures in which the multipotent intestinal stem-cells reside together with secretory Paneth cells (17). Using techniques such as chemical mutagenesis it was proved that all IECs come from

an single progenitor, forming labelled “ribbons” as the cells migrate and differentiate from crypt bottom to villus tip (18).

### *Intestinal stem cells*

There are two different types of stem cells in the intestinal crypts, the proliferating Lgr5+ or crypt base columnar (CBC) characterized by the expression of Lgr5, Olm4 CD133 and Lrig1 (19), and the quiescent stem cells, also called label retaining cells (LRCs), which are thought to play an important role in regeneration of damaged epithelium (20). Of notice, recent evidence suggest that either progenitor or differentiated cells of the intestine can also regain stem cell phenotype to regenerate the epithelium (21).

### *Transit-amplifying cells*

The ever-proliferating Lgr5+ stem cells replenish the intestinal crypt with daughter-cells of the transit-amplifying (TA) compartment, which migrate upwards while being exposed to gradients of different morphogens and signaling molecules. By lateral inhibition, they commit either to the secretory or the absorptive lineage, which will bring about the terminally differentiated intestinal cell lines: enterocytes, Paneth cells, goblet cells, enteroendocrine cells, tuft cells and M cells (22).

### *Goblet cells*

Goblet cells secrete the mucins that form the mucus layer of the intestine (9). Differentiation of secretory progenitors into goblet cells is dependent of the inhibition of Notch signaling (23).

### *Enteroendocrine cells*

There are up to 15 different subtypes based on the hormone they produce. They are scattered throughout the mucosa representing approximately 1% of the IECs (24).

### *Paneth cells*

Paneth cells are secretory cells and the main source of antimicrobial peptides in the intestine. They reside the base of the Lieberkühn crypts and contain a large ER and Golgi.

Discharge secretory granules into the lumen of the intestine (25). In the differentiation process, cells committed to be Paneth cells migrate downwards to the crypt bottom where the mature Paneth cells are (26). Being so close to the CBCs, Paneth cells are important for the maintenance of their stem-cell state (27).

### *Microfold cells*

Microfold (M) cells are responsible for the transport of microorganisms and antigens into the mucosal lymphoid follicles helping to strengthen mucosal innate immunity, but since they represent a weaker point in the epithelial barrier, it is possible to some pathogens to have developed mechanisms of infection through the M cells (28).

### *Tuft cells*

The term “Tuft cell” is used to describe a type of cell that can be found not only in the intestine, but also trachea and lungs of human beings and whose function remained unknown for many decades. Intestinal tuft cells come from the secretory progenitor and have an important role in immunity, especially through their interaction with Group 2 innate lymphoid cells (ILC2s) (29).

### *Enterocytes*

Enterocytes or columnar cells are the only absorptive lineage of the intestinal crypt and are highly polarized cells presenting a brush border specialized in the absorption and transport of nutrients. They are the most frequent type of IECs, accounting for approximately 80% of them (30). A model of the intestinal crypt and the IECs together with the dedifferentiation model of adult cells in order to repair tissue damage is shown in Figure 1.1.

## Gene expression analysis to study celiac disease and colitis-associated cancer

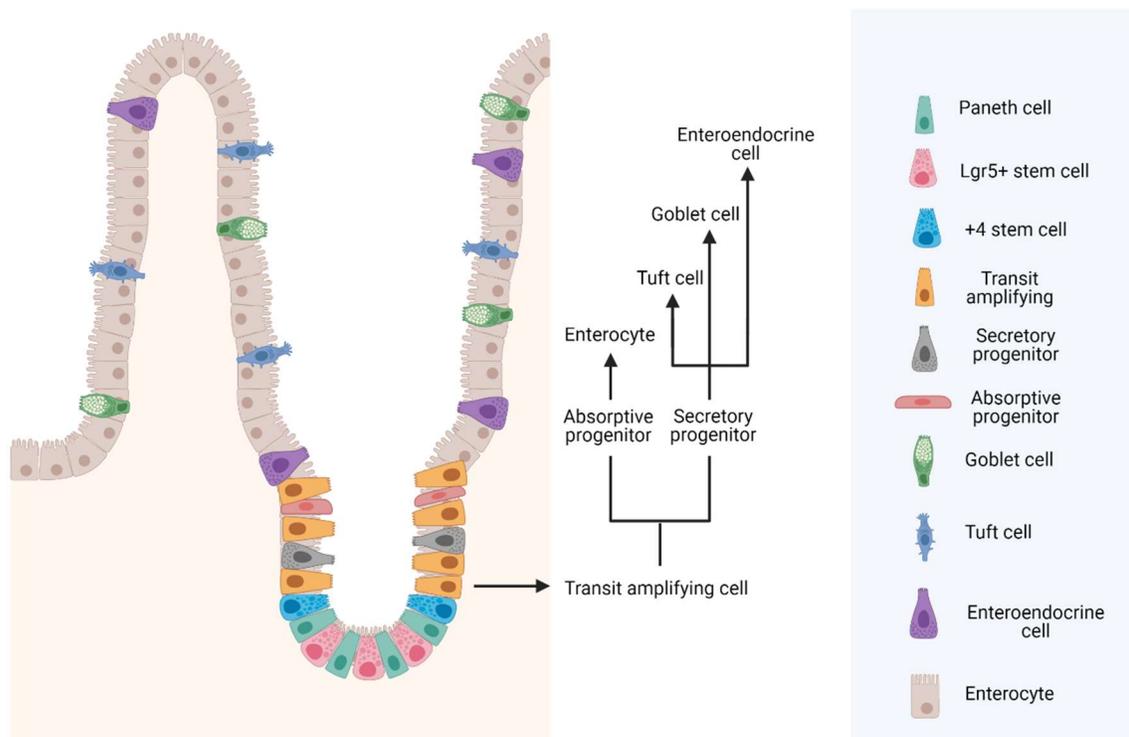


Figure 1.1. The intestinal crypt. Lgr5+ stem cells at the base of the crypt maintain cell renewal. Paneth cells reside at the base of the crypt and help maintain the stem cell permissive environment. +4 stem cells are quiescent “reserve stem cells”. Transit amplifying cells are rapidly proliferating and differentiate into secretory and absorptive progenitors. Absorptive progenitors differentiate into enterocytes, whereas secretory progenitors differentiate into Goblet, Paneth, Tuft and Enteroendocrine cells. Image created with BioRender.com

### Intraepithelial lymphocytes

Intraepithelial lymphocytes (IELs) are a unique type of T lymphocytes within the epithelium of the small intestine. They are located at the basement membrane and occur in a healthy tissue at a frequency of 10-15 IELs/100 epithelial cells (31). They are classified based on ontogeny as naturally occurring (Type B) and adaptively induced (Type A) IELs (32). Naturally occurring IELs are tissue resident TCR $\gamma\delta$ + T cells existing independently of microbial colonization of the gut (33), whereas adaptive IELs express the TCR $\alpha\beta$ +CD8 $\alpha\beta$ + and TCR $\alpha\beta$ +CD4+ T cells generated after local tissue damage (34).

All TCR $\alpha\beta$ +CD8 $\alpha\beta$ + IELs express Natural Killer receptors NKG2D as well as CD94 and NKR-P1A (35). The IELs are important in celiac disease innate response through the NKG2D (36). IELs are thought to participate in the surveillance and maintenance of barrier function due to their proximity to the intestinal lumen and their ability to respond to tissue stress via NK receptor or classical antigen-specific TCR interaction (37).

## **EPITHELIAL CELL JUNCTIONS: THE APICAL JUNCTIONAL COMPLEX**

As mentioned in the beginning of this study, the epithelial barrier has the function of protecting the body from harmful stimuli and providing absorption of nutrients. This dual task is possible due to the structures through which epithelial cells establish contact with each other. In the particular case of the intestine, the columnar epithelial cells are highly polarized in apical and basolateral domains and the limits of those are also determined by the same structure: the apical junctional complex. The apical junctional complex is formed by the TJ and Adherens junctions (AJ).

## **ADHERENS JUNCTIONS**

Adherens junctions are cell junctions that localize below the tight junctions and are formed by homophilic interactions between E-cadherins (E-cad) from neighboring cells via their N-terminal extracellular domain (38). E-cad is a large protein with long extracellular and cytosolic domains, the latter interacting with various intracellular proteins, being  $\beta$ -catenin the most frequent and best studied interaction, and linking it to the actin-myosin network, vesicle transport and cell polarity complexes (39). The AJ also transduces mechanical signals from junctions to the nucleus, even inciting the transcription of oncogenes (40).

## **TIGHT JUNCTIONS**

### **Bicellular TJ**

TJ are cell junctions localized far up on the lateral membrane of epithelial cells. They were first observed by freeze-fracture microscopy as strands forming a net structure on the side of intestinal cells and as “kissing points” on electron micrographs (41). They are mainly constituted by claudins, TJ-associated MARVEL (MAL and related proteins for vesicle trafficking and membrane link) proteins (TAMPs) and cytosolic scaffolding proteins from the membrane-associated guanylate kinase (MAGUK) family (42).

### **Tricellular TJ**

Contacts between cells occur laterally between two cell or on the corners between three cells. The bicellular TJ form in the lateral contact points, they are different from the tricellular TJ in structure and composition. The tricellular TJ forms a central tube with diameter estimated to be around 10 nm, and length up to 1  $\mu$ m, allowing for the passage of macromolecules (43). The major constituents of tricellular TJ are tricellulin and the proteins from the angulin family: angulin-1 (Lipolysis-stimulated lipoprotein receptor (LSR)); angulin-2 (immunoglobulin-like domain containing receptor (ILDR1)) and angulin-3 (ILDR2; LISCH-like or C1orf32) (44).

### **Claudins**

Claudins are the proteins responsible for establishing cell-cell contact in the TJ via homophilic and heterophilic interactions. 27 claudins have been described, being 26 found in humans and one (claudin-13) only found in mice. They are small proteins, with molecular masses ranging from 21 to 34 kDa and present four transmembrane helices: a short intracellular N-terminal domain, a longer intracellular C-terminal region, a small intracellular loop and two extracellular loops (ECL1 and 2) (45). There is a signature domain in the ECL1 and a COOH-terminal PDZ binding motif that mediates interaction to the PDZ domains of the

scaffolding/adaptor proteins (46). Claudins can be divided by phylogeny in eight subgroups or four major clusters: cluster 1 (subgroups A and B) claudin-3 -4, -5, -6, -9 and 8; cluster 2 (subgroups D and E) claudin-1, -7, -19, -2, -14, -20; cluster 3 (subgroup F) claudin-10, -11, -15, -18, and cluster 4 (subgroups C, G and H) claudin-21, -22, -24, -12, -16, -25, -23, -26, and -27 (47).

### *Barrier- and pore-forming claudins*

Claudins are the ones responsible for the regulation of transport through the paracellular pathway and for that reason can present properties for preventing transport (barrier) or promoting transport (pore) of ions and water. The pore-forming claudins can be divided into cation permissive: claudin-2, -10b and -15,-16 and -21 and anion permissive: claudin-10a and claudin-17 (48–50). The barrier-forming claudins are claudin-1, -3, -4, -5, -6, -8, -9, -11, -14, and -18 (50).

TJ are ubiquitous in epithelial tissues and, consequently, claudins. A different set of claudins is found in the various epithelial tissues of the human body and they described in table 1.1.

Table 1.1. Distribution of claudins by organ

<b>Organ</b>	<b>Claudins</b>	<b>Reference</b>
Choroid plexus	1, 2, 5, 11	(51–53)
Cochlea	1, 2, 3, 8, 9, 10, 11, 12, 14, 18	(54)
Distal respiratory tract	3-5, 7, 8, 15, 18-1	(52,55,56)
Epidermis	1, 3, 4, 5, 7, 8, 11, 12, 17	(57)(58)
Epididymis	2, 4, 5, 7, 10	(59)
Exocrine pancreas	1-5, 7	(60)(61)
Eye	1, 4, 7, 10	(62)
Gall bladder	1-4, 10, 7, 8	(63)
Intestine	1-5, 7, 8, 10, 12, 15, 18, 20, 21, 23	(64,65)

## Gene expression analysis to study celiac disease and colitis-associated cancer

Kidney nephron	Glomerulus	5, 6, 1	(66–68)
Kidney nephron	Proximal tubule	2, 10a, 17, 6, 9	(48,49,66,69,70)
Kidney nephron	Upper thin descending limb	2	(66,70)
Kidney nephron	Lower thin descending limb	7, 8	(71)
Kidney nephron	Thin ascending limb	3, 4, 16, 19	(66,72)
Kidney nephron	Thick ascending limb	3, 10a, 10b, 16, 18, 19	(49,66,72–75)
Kidney nephron	Macula densa	10	(48)
Kidney nephron	Distal tube, connecting tubule and collecting duct	7, 8, 10	(48,71,76)
Kidney nephron	Collecting duct	3, 4, 7, 8, 10a, 10b, 14, 18	(48,49,66,71,73,77)
Liver		1-3- 5-9, 14	(61,78–80)
Mammary gland		1-5, 7, 8, 15, 16	(81–84)
Proximal respiratory tract		1, 3-5, 7, 10, 18	(55,56)
Retinal pigment epithelium		3,10, 19	(85)
Ovary		1, 5	(86)
Prostate		1, 3, 4, 5, 7, 8, 10	(87)
Salivary gland		1, 2, 3, 4, 7, 8, 10, 12	(88–90)
Seminiferous tubule		3, 5, 11	(52)(91)
Stomach		3, 4, 5, 12, 18, 23	(55,92,93)
Taste bud		4, 6, 7, 8	(94)
Urinary bladder		4, 8, 12	(95)

**TJ-associated MARVEL proteins****Occludin**

Occludin was the first described TJ integral protein (96), however, its function in the TJ is not yet known. It has four transmembrane domains, two extracellular loops, short N-terminal and long C-terminal region. The C-terminal region can bind to the MAGUK family proteins ZO-1, ZO-2

and ZO-3 as well as F-actin (97). Even though being expressed in all epithelial TJ, occludin knock-out mice do not present barrier impairment (98).

### **MARVEL D3**

MARVEL D3 (MD3) occurs in two isoforms: MD3-1 and MD3-2 and was the last MARVEL protein described. It also has four transmembrane domains; however, the C-terminal region does not bind to ZO-1 in contrast to occludin and tricellulin. MD3 co-localizes with occludin and ZO-1, however, it is not essential for TJ formation and barrier function (99).

### **Tricellulin**

As mentioned above, tricellulin (MARVEL D2) participates in the tricellular TJ, forming the central tube. Overexpression of tricellulin increases transepithelial electrical resistance (TEER) and its down-regulation results in impaired barrier function (100,101).

### **Junctional adhesion proteins (JAMs)**

There are three JAM proteins: JAM-1, -2 and -3 (also called JAM-A, -B and -C) all around 40 kDa. They possess a single transmembrane domain and an extracellular region whose folding resembles an immunoglobulin. They are, in fact, members of the immunoglobulin superfamily; play an important role in TJ assembly and localize in the TJ laterally to the claudins (102).

### **Scaffold proteins**

The TJ scaffold proteins belong to the MAGUK family and includes ZO-1, -2 and -3. They are very large proteins and present three PDZ domains, which is important for binding membrane proteins; one Src homology 3 (SH3) and a guanylate kinase domain, all of them important for the assembly and homeostasis of the TJ (103). They mediate the interaction between the TJ structure and its transmembrane proteins and intracellular structures and signaling molecules for the regulation of TJs and are essential for efficient TJ assembly and tricellulin localization at the tricellular TJ (104).

### **Tight junctions of the intestine**

The expression of claudins has been thoroughly assessed in mouse and rat intestines. The most expressed claudins are 2, 3, 7 and 15 and only 6, 16, 19, 22 and 24 were not detected (64).

Many claudins present regional distribution throughout the intestine. Claudin-8 is more expressed in the colon, conversely, claudin-15 is more expressed in the duodenum. Claudin-2 only appears in deep crypts. Claudins-2, 8, 10, 12, 15 and -18 seemed to be restricted to the TJ, whereas 1, 3, 4, 5 and 7 were also found at the basolateral membrane (64)

### **THE INTESTINAL IMMUNE SYSTEM**

The intestine harbors the largest compartment of the human immune system, presenting both adaptive and innate immunity. It is constantly exposed to antigens from the microbiome and the diet and it is also the entry point for many pathogens into the organism. The adaptive immune response is located in the gut-associated lymphoid tissue (GALT) and the draining lymph nodes, whereas the innate immune cells are widespread throughout the lamina propria and epithelium (105).

#### **Organized lymphoid structures**

The GALT are the main lymphatic organs in the mucosa and submucosa of the intestine and are composed of lymphoid aggregates surrounded by a “follicle-associated epithelium” with M cells, which transports antigens from lumen to the dendritic cells at the subepithelial dome (SED). The GALT also includes smaller lymphoid aggregates that are collectively termed isolated lymphoid tissues (ILTs) (106).

Peyer’s patches are the best characterized GALT structures and are located on the small intestine. They consist of numerous B cell follicles surrounded by smaller T cell areas. They are not encapsulated and always contain germinal centers (107). ILTs are microscopic structures

containing germinal centers and primarily consisting of B cells without a defined T cell zone. The human intestine contains approximately 30000 ILTS (108)

### **Effector cells**

The effector cells in the lamina propria are very diverse, consisting of B cells, T cells and numerous innate immune populations, including dendritic cells, macrophages, eosinophils, and mast cells.

#### *Lamina propria T cells*

There are proximately twice as more CD4+ than CD8+ T cells in the intestinal lamina propria and since most of them present effector memory characteristics, they are believed to be originated in the thymus and primed in different secondary lymphoid organs before homing to the intestine (109). The repertoire of T cells in the lamina propria is very diverse, containing IL2+, IL2+IFN $\gamma$ +, IFN $\gamma$ +, IL-17+, forkhead box P3 (FOXP3)+ IL-10-producing regulatory T cell (T<sub>Reg</sub>) and FOXP3-Treg type 1 (T<sub>R1</sub>) cells. IL-10 and IL-17 producing cells seem to be more present in the colon (110).

#### *B cells*

Intestinal lamina propria contains a large number of plasma cells increasing to the distal end of colon in number and in proportion of IgA secreting cells from 75% in the duodenum to 90% in the colon. The rest of them secrete IgM. The production of secretory IgA (SIgA) is dependent of the presence of microbiota and polymeric Ig receptor (PIGR) expression, which transports the SIgA to the lumen (111)

#### *Innate lymphoid cells*

There are three types of ILCs: ILC1, ILC2 and ILC3. ILC1 is expressed along the small and large intestines in equal proportions, whereas ILC2 is not described in the colon and ILC3 proportion is higher in the colon (112). ILC2 function is strongly influenced by circadian rhythm-

dependent feeding cycles, which are controlled by the hormone vasoactive intestinal peptide (113). Most of the ILC3 cells produce IL-22 and express the NKp46 receptor and are found in the ILTs and colonic patches, but they also populate the lamina propria (114).

### *Invariant T cells*

There are minor subsets of T cells that express invariant forms of the TCR including CD3+CD161<sup>hi</sup>CD8 $\alpha\alpha$ + (or CD4-CD8-) mucosal-associated invariant T cells (MAIT cells), and invariant natural killer T cells (iNKT cells). MAIT cells were only described in human jejunum, accounting for 2-3% of lamina propria T cells. They produce cytokines and exert cytolytic activity upon recognizing bacteria infected cells (115). iNKT react to self-antigens and bacterial lipids, by recognizing glycolipids presented by the Major Histocompatibility Complex MHC class I-molecule CD1d (116).

### **Mononuclear phagocytes**

Mononuclear phagocytes are macrophages and dendritic cells (DCs). The distinction of these two subtypes is sometimes questioned because they share many surface markers such as CD11c, MHC class II, CD11b and CX<sub>3</sub>C-chemokine receptor (CX<sub>3</sub>CR1) (117).

### *Macrophages*

Macrophages are abundant in the intestinal lamina propria and fulfill an array of functions to maintain homeostasis such as phagocytosis of microorganisms and dead cells, production of mediators to promote tissue renewal and maintenance of T cells. They produce great amounts of IL-10, suppressing the immune response triggered by pattern recognition receptors (118) and promote survival of FOXP3+ T<sub>Reg</sub> cells (119).

### *DCs*

There are two subsets of DCs in the lamina propria: the CD103+CD11b- and CD103+CD11b+. CD103+CD11b- play a role in the initiation of adaptive immune response by cross-presenting antigens to CD8+ T cells (120).

### **Other innate cells**

#### *Eosinophils*

Eosinophils can account for 30% of all myeloid cells and are believed to assist maintain the populations of CD103+ DCs, IgA+ plasma cells, and FOXP3+ Treg cells, through secretion of transforming growth factor  $\beta$  (TGF $\beta$ )-activating metalloproteinases (121).

#### *Mast cells*

Mast cells are found in healthy mucosa and submucosa of the intestine and secrete molecules that mediate barrier function, permeability, peristalsis, and vascular tone as well as interact with the enteric nervous system (122).

## **DISEASES IN THIS STUDY**

Within the present thesis research was made on different autoimmune diseases of the intestine, celiac disease, and colitis-associated cancer (which involves the study of inflammatory bowel disease as well). On the first view, these diseases fall into two groups with only limited commonalities. However, recent research has uncovered the mucosal barrier as a central structure in the initiation of all three diseases.

### **Celiac disease**

Celiac disease (CeD) is a T-cell-dependent immune-modulated disease that develops in genetically susceptible individuals upon gluten ingestion (123). Classically known to mainly affect the duodenum, the condition nowadays is regarded as a systemic malabsorptive disease presenting chronic diarrhea, abdominal pain, weight loss or failure to thrive, nutrient/vitamin

deficiency besides presenting extraintestinal manifestations ranging from neurological, ocular, dermatologic, oral, musculoskeletal, cardiovascular, pulmonary, renal, cardiovascular, hepatic, endocrine and reproductive manifestations (124). Prevalence for celiac disease worldwide is around 1,4% with Asia present the highest (1.8%) and South America, the lowest (1.3%) (125). An interesting fact is that CeD incidence and prevalence are increasing, as seen in studies which measured incidence overtime (126). CeD can develop at any age, but recent cohort studies reported that most patients develop CeD before 10 years of age (127). Prevalence rates are higher in females than in males (126), but it can be biased due to male patients being less likely to undergo biopsy examination than females, even though presenting CeD-related symptoms (128).

Genetic predisposition to CeD is dependent on the expression of specific human leukocyte antigen (HLA)-DQ2 and/or HLA-DQ8 haplotypes in the membrane of antigen-presenting cells (129). However, they are not a sufficient cause for celiac disease development since they are expressed in virtually 40% of the general population (130).

Gluten is a nomenclature that comprises the storage proteins of wheat, barley, rye, and other related grains. Of those, gliadins are the class of proteins that are most important etiologic factors of gluten-related disorders (131). Gluten proteins present high solubility in alcohols, due to their high glutamine and proline content, which also renders them resistant to complete digestion in the human intestine. Different gliadin peptides are formed as a result of this partial digestion being the immunogenic 33-mer (pp.57-89) known to induce a strong adaptive response, and the 25 AA peptide (pp. 31-35) reported to induce IL-15 expression in enterocytes and dendritic cells (132).

Upon ingestion of gluten-containing food, the gluten peptides permeate the epithelial barrier and reach the lamina propria, where they are deamidated by the enzyme tissue-transglutaminase

2 (TG2), enhancing their affinity to the specific HLA haplotypes (133). Those are necessary for triggering the immune response by activating gluten-specific CD4+ T-cells with the immunogenic gliadin peptides. Competent gluten-specific CD4+ T-cells are a small population of 0.5%-2% of all intestinal CD4+ T cells found only in celiac patients (134) and upon activation, they promote a pro-inflammatory phenotype and go to the lamina propria where they initiate the immune response by secreting great amounts of IFN $\gamma$  and IL-21 (135). Concomitantly, gluten- and TG2-specific B-cells differentiate into plasma cells and produce antibodies against deamidated gliadin peptides (DGPs) and TG2, which are used in diagnostics as specific markers of CeD (135).

Hallmarks of CeD include the positive serology for anti-TG2, anti- endomysium (EMA) and DGP antibodies; positive HLA-DQ2 or -DQ8 testing and the histological finding of villous atrophy and crypt hyperplasia that resolve with the adherence to a strict gluten-free diet (GFD), which is the only standardized treatment until now (131). Of notice, a percentage of patients does not respond to a GFD and present persistent malabsorption and villous atrophy even after 1 year of strict GFD; being, for that reason, regarded as refractory celiac disease (RCD) patients (136).

The RCD patients are divided in two different categories depending on their population of IELs. On the one hand, RCDI are those who present an increased, but normal repertoire of IELs, with no T-cell receptor (TCR) clonality. On the other hand, RCDII patients present abnormal IEL population of more than 25% of CD103+ or CD45+ lacking surface CD3 or 50% of IELs expressing intracellular CD3 $\epsilon$  but no CD8 and/or clonal rearrangement of the TCR $\gamma$  chain (137). The RCDII patients are of special concern due to the increased risk of developing enteropathy-associated T-cell lymphoma (EATL) a disease presenting 5-year overall survival rate of 44-58% (137).

### *Barrier impairment in Celiac disease*

Epithelial barrier impairment in celiac disease is studied since the 1970s, when it was observed that patients with villous atrophy present increased permeability to disaccharides and

decreased permeability to monosaccharides (138–140). Moreover, macrostructural changes are also observed in the TJ network where celiac patients present fewer and discontinued TJ strands that can be recovered with a gluten-free diet (141). Around 50% reduction in TEER was observed in active celiac patients (142,143) as well as a partial recovery in gluten-free adherent patients (142,144). In the background of those structural and functional alterations there are molecular changes in celiac disease intestinal mucosa. Increased pore-forming claudins-2 and -15 together with diminution of protein content of ZO-1, occludin, barrier-forming claudins -3, -5 and -7 as well as membrane displacement of those proteins were reported in celiac patients samples (143,145). In addition, there is evidence for a correlation between an impaired polarization process and the molecular changes described above (144).

The immune response in CeD also affect the barrier properties of the intestinal mucosa. To exemplify that, TNF $\alpha$  and IFN $\gamma$  exposure also displace ZO-1, occludin, and E-cadherin from the plasma membrane of Caco-2 cells (145). Comparably, TGF $\beta$  prevents epithelial cells from polarizing correctly by inhibiting TJ assembly (143). Furthermore, the innate immunity in CeD disrupts the barrier by provoking apoptosis of epithelial cells through gliadin-induced IL-15 (146). IL-15 promotes the expression of MHC class I chain-related protein A (MICA) in epithelial cells and the survival of IELs presenting NKG2D receptor. IELs induce apoptosis in epithelial cells through interaction of MICA and NKG2D. Moreover, the innate immunity in CeD disrupts the barrier by provoking apoptosis of epithelial cells indirectly by IL-15. IL-15, whose expression can be triggered by gliadin fragments (147), on the one hand, promotes the expression of MICA in epithelial cells and, on the other hand, promotes survival of NKG2D-expressing IELs (137), which cause epithelial lysis through the interaction between the NKG2D receptor to the MICA molecule (147).

These alterations point to a disturbance in the function of the TJ and even though they mostly disappear after GFD, the small remaining difference points to a genetic cause.

### *Evidence for genetic cause for barrier impairment*

The first study that pointed out for a primary cause for barrier impairment in celiac disease reported that the lactulose/mannitol ratio of healthy relatives of celiac disease patients is significantly higher than control patients' (148). With the advent of genome wide association studies (GWAS) more non-HLA loci contributing to CeD risk were described (149,150). So far 39 loci comprising 57 independent genetic SNP variants have been identified; however, around 80% of the variants are located in noncoding regions of the genome, suggesting the genetic variation impacts transcription regulation rather than the gene sequence (151). Four of those loci were predicted to have an impact in cell-cell adhesion, including the genes LPP and C1orf106 (151).

### *Lipoma preferred partner*

Lipoma preferred partner (LPP) was described in a subset of lipomas, one of the most common mesenchymal tumor types in humans and characterized for having translocations involving chromosome segment 12q13-q1, as a preferred fusion partner for *HMG1C*, a member of the high mobility group (HMG) protein gene family. It contains a proline-rich domain in the N-terminal region, a leucine-zipper and three C-terminal LIM domains, indicating it should be categorized as a LIM family protein (152). After its characterization, LPP has been reported to mostly localize in focal adhesions and possess a nuclear exportation signal (NES) sequence and the ability to act as a transcription factor (153). LPP was also identified among proteins proximal to E-cadherin. Despite having a PDZ domain, it is not dependent of ZO-1 to localize to cell contacts. LPP knock-out cells have weaker E-cadherin adhesion contacts and had an impaired barrier re-establishment as evaluated by a calcium-switch assay (154). LPP was found to be a risk factor for celiac disease in a genome wide association study (GWAS) (155) and also has a role in cancer metastasis progression (156).

### *C1orf106*

C1orf106 (INAVA) was found in GWAS as a risk factor for celiac disease, ulcerative colitis (157) and Crohn's disease (158). It is highly expressed in IECs, where it co-localizes to ZO-1 at the TJ. C1orf106 function is thought to be the regulation, by inducing ubiquitination and degradation, of the cytohesins 1 and 2 (159,160). Cytohesins are GEFs (guanine exchange factor) for the GTPase ADP-ribosylation factor 6 (ARF6), which in turn promotes internalization of E-cad e AJ disassembling (161). Caco-2 cells which were knocked-down of C1orf106 present lower TEER and increased permeability to lucifer yellow (160). In addition, C1orf106 KO in mice rendered mice more susceptible to barrier impairment after TNF $\alpha$  exposure (159).

### **Inflammatory Bowel Disease**

Inflammatory bowel disease (IBD) comprises a spectrum of diseases characterized by a chronic inflammatory process in the gastrointestinal tract, being Crohn's disease (CD) and Ulcerative colitis (UC) the two broadest subtypes of IBD. Epidemiology of IBD varies greatly worldwide. Both diseases clinically present abdominal pain and diarrhea. UC often causes rectal bleeding more than CD. CD patients often have weight loss and perianal disease (162).

Diagnosis of IBD often requires colonoscopy examination, with biopsy acquisition for histologic examination. Laboratory exams that assist in the diagnosis are erythrocyte sedimentation, (ESR), serum C-reactive protein (CRP) and fecal calprotectin (162). High calprotectin correlates with histologic grade of mucosal inflammation with a sensitivity of 94% and a specificity of 64% (163).

### *Ulcerative colitis*

UC presents a greater incidence in Europe, ranging from 0.97 to 57.9 per 100000 and greater prevalence in Europe, from 2.42 to 505 (164). It is characterized by a mucosal chronic inflammation that starts in the rectum and progresses up into the colon in a continuous way (165).

The pathogenesis of UC is not completely understood, nonetheless, it is believed to develop through a combination of genetic and environmental factors (166). Around 260 risk loci for UC were found in whole genome sequencing experiments (167–169). 67% of which are shared with CD (170) and the strongest genetic signals coming from HLA regions (171). Regarding environmental factors, UC incidence is rising in industrialized countries and urbanization, exposure to pollution and changes in diet are considered contributory factors (172). Interestingly, smoking is protective against UC and the disease only develops after the person has quit smoking (173).

Gut microbiota in UC is altered with depletion of protective bacteria (174). The epithelial barrier is thought to also play a role in the disease pathogenesis, especially regarding impaired production of antimicrobial molecules by Paneth and goblet cells (175). Inflammation also plays an important role in barrier dysfunction, as the secretion of TNF $\alpha$ , IFN $\gamma$  and IL-13 affect intestinal barrier function (176,177).

The immune response in UC is classically described as a TH2 response (178), however, some findings state otherwise, for example, IL-23 being largely expressed in UC mucosa (179), the increased numbers of Th17 and Th9 lymphocytes and the fact that anti-IL-23 drugs such as mirikizumab and ustekinumab are effective in the treatment of UC (180,181).

### *Crohn's disease*

CD incidence is greater in Oceania, ranging from 12.9 to 29.3 and prevalence in Europe ranging from (164). Different from UC, CD is characterized for causing transmural inflammation and by skip lesions that can be found anywhere in the GI tract (182). Similar to UC, CD is thought to develop through a combination of genetic and environmental factors.

There are around 200 risk loci described for CD (170,183), being the most important NOD2, IL-23, HLA ATG16L1, IRGM, and LRRK2/MUC19 SLC22/OCTN on 5q31 and TNF (184). Among

the environmental factors associated to CD risk are smoking, oral contraceptive use, antibiotics and anti-inflammatory drugs, urban development (185,186).

CD is treated with corticosteroids, mostly for symptom management; immunomodulators and nowadays mostly with biologicals against TNF, integrins and IL-12 (187–189).

### **Colitis-associated cancer**

IBD presents as its most deadly complication the development of intestinal cancer, specifically called colitis-associated cancer (CAC). This type of cancer is relatively rare, incidence of 1,7 in all IBD patients (190), representing 2% of all cancer cases, and it is associated with significant morbidity and mortality reaching up to 15% of cases. The risk factors associated with CAC are the age at onset of IBD, duration of disease, anatomic extent, histological changes, primary sclerosing cholangitis (PSC) and family history of cancer, which is the only independent factor on this list. Regarding the age of onset, patients that develop IBD before reaching 15 years of age have 40% of risk, whereas this risk drops to 25% in those who developed it between 15 and 39 years. The duration of IBD is a very important risk factor because it is also determinant for the surveillance interval and it is calculated to be around 8% for both UC and CD after 20 years of disease. The anatomic extent is especially important for UC, since CD can present patchy lesions in the whole extension of the gastrointestinal tract, whereas UC is restricted to the colon. In UC, the disease is denominated extensive when it extends beyond the splenic flexure and those patients have a higher risk of developing CAC (incidence = 7); left-side UC is restricted to the descending colon and presents an intermediate risk for CAC (SIR= 1.7), finally, proctosigmoiditis is restricted to the end of the colon and has low to inexistent risk of developing CAC (190). For CD, any colon involvement accounts for an SIR of 2. Both diseases in studies made in referral centers, extensive colon involvement accounts for an SIR of 18.2 (190). Histological changes are an important risk factor as well as diagnostic finding for CAC, the more extensive or multiple, the greater the risk of developing or even determine an initial stage of CAC.

PSC is a hepatic disease intrinsically associated with UC and its presence accounts for 4-fold increased risk for cancer development. Finally, family history features an independent factor for all cancer types as well as for CAC.

Conversely from sporadic colorectal carcinoma (CRC), CAC does not progress from an adenoma to carcinoma but rather from a dysplastic lesion, which are usually flat lesions and often difficult to identify during colonoscopy. Dysplasia is present in 75 to 90% of CAC patients, however, CAC may occur without a prior history of dysplasia. Besides, some lesions are only identified in histological examination and for that are called microscopic dysplasia or invisible dysplasia. For that reason, IBD patients undergo a strict surveillance protocol to diagnose CAC as early as possible. This surveillance is done with periodic colonoscopies and collection of random biopsies every 4 cm of colon up to 30-40 samples that are histologically examined afterwards. When a visible lesion is identified, it is immediately resected. When there is suspicion of a lesion, or when an invisible lesion is detected in histology, it should be followed by a chromoendoscopy, in which a fluorescent dye is locally injected to improve the detection of lesions. High-grade invisible dysplasia, multifocal low-grade dysplasia or unresectable lesions are indications for colectomy, which is the standard treatment for CAC (190).

### *Inflammation-driven carcinogenesis*

Tumor progression differs in between CRC and CAC not only in the type of initial lesion, but also in the chronology of genetic alterations and the fact that the chronic inflammatory process contributes to tumorigenesis. On the one hand, CRC tumor progression is very well described in a model of sequential somatic alterations beginning with APC mutation, followed by microsatellite instability and KRAS mutation and finally presenting p53 mutation/allelic deletion in its final stage (191). On the other hand, even though microsatellite and chromosomal instability also play a role in CAC as well as KRAS mutation, CAC often does not present APC mutation, or it is only observed in the final stages of progression. Moreover, p53 mutation is an early event occurring in

85% of cases and p53 allelic deletion occurs in 50% of cases and it is known to sustain activation of NF $\kappa$ B aggravating the extent of colitis and accelerating the progression to high grade dysplasia and carcinoma (192). Finally, the inflammatory milieu of IBD is complex and inflammatory cytokines such as interleukin (IL)-6, IL-11 and TNF $\alpha$  were shown to contribute to colorectal cancer tumorigenesis (193–195). Among the cytokines that are secreted in the microenvironment and influence the fate of immune and epithelial cells is osteopontin (OPN), which is a secreted glycoprotein that acts as a cytokine and promotes differentiation of immune cells such as macrophages and dendritic cells (191).

## **OSTEOPONTIN**

Human OPN is a glycoprotein important for many biological processes, such as bone homeostasis, angiogenesis, cell migration, inflammatory process, and tumor progression (196). It is expressed by osteoclasts, osteoblast, immune cells such as DCs, NKs, T and B cells as well as an array of epithelial cells from the intestine, kidney, bladder, breast, lung, gallbladder and different cell populations from different organs such as Kupfer cells of the liver, islet cells of the pancreas, Leydig cells of the testis, Hoffbauer cells from the placenta, follicular cells of the thyroid et cetera (197).

### **Structure and function**

OPN, which is also known as Secreted Phospho-Protein 1 (SPP1), Bone Sialoprotein 1 (BSP-1) and Early T-lymphocyte Activation 1 (ETA-1), was first identified as a component of bone extracellular matrix (198). It is a member of the SIBLING (Small Integrin-Binding Ligand, N-linked Glycoprotein) family and it is codified by the SPP1 gene, which is a single-copy gene with 7 exons, mapped in a primary culture of human bone cells to chromosome 4q13 (199). The product of this gene is a ~34 kDa protein that contains several domains which are highly conserved among species such as the main integrin binding motif GRGDS (or only RGD) (199), transglutaminase-reactive glutamines (200), the thrombin cleavage motif SVVYGRL, calcium binding sites and 2

putative heparin binding domains. OPN binds to integrin  $\alpha v\beta 3$  (201),  $\alpha v\beta 1$ ,  $\alpha v\beta 5$  (202), CD44 (203),  $\alpha 8\beta 1$  (204),  $\alpha 9\beta 1$  through the SVVYGRL domain (205), as well as  $\alpha 4\beta 1$  and  $\alpha 4\beta 7$  (206).  $\alpha 4\beta 1$  also binds a different domain of the N-terminal OPN (207).  $\alpha v\beta 6$  binding depends on amino acids upstream of the RGD and  $\alpha 5\beta 1$  binding depends on both RGD and amino acids downstream of it (208).

### *Isoforms*

OPN exists in a variety of splicing isoforms and posttranslational modifications, e.g., phosphorylation, glycosylation, and cleavage.

### Soluble OPN

OPN-a is the full-length protein, OPN-b lacks exon 5 which contains phosphorylated Ser and Thr; and OPN-c lacks exon 4, which contains the transglutaminase binding sequence and due to that cannot form polymeric complexes (209). Those isoforms present different expression patterns in different tissues and even prognostic value for some cancers. For example: OPN-c is detected in breast cancer cells, but not in the healthy surrounding tissue; conversely, OPN-a and -b are found in both (210). In addition, OPN-c increased expression correlates with tumor grade and poor prognosis (211). In pancreatic cancer, OPN-b is associated with poorer prognosis and OPN-c with metastasis (212). OPN-c is used as a biomarker to distinguish between prostate cancer and prostate benign hyperplasia (213). Finally, OPN-a and -c induce invasiveness in glioma cells, but not OPN-b (214).

### Intracellular OPN

iOPN It is a truncated form of OPN lacking the signal sequence (215). It has been shown to play important roles in immune cells. NK cells lacking iOPN present impaired expansion and increased apoptosis (216). In follicular T helper cells iOPN supports differentiation by inhibiting Bcl-6 degradation (217).

## PTM

Among the posttranslational modifications suffered by OPN, cleavage by thrombin is the most largely studied and has been shown to separate the integrins and CD44 binding domains (218). Thrombin-cleaved OPN N-terminal fragment binds to integrins  $\alpha v\beta 3$ ,  $\alpha v\beta 5$ ,  $\alpha 9\beta 1$  and  $\alpha 4\beta 1$  to promote cell adhesion (208). On the other hand, the C-terminal fragment binds to CD44v6 and v3 to promote invasion and tumorigenesis (219). OPN is also cleaved by MMP3 and MMP7, which enhances binding to integrins  $\beta 1$  and inducing migration (220).

## *Osteopontin signaling*

OPN is reported to activate several signaling pathways in different cells (Figure 1.2). By interaction with CD44, it activates PI3K/AKT pathway through phospholipase C and regulates gene expression in immune cells. Proliferation signals from OPN usually are associated with interaction with EGFR, for example: in prostate cells, OPN induces proliferation by EGFR/PI3K/AKT (221). Another example is OPN induction of cell motility in breast cancer cells by PI3K/AKT/NF $\kappa$ B production of urokinase type plasminogen activator (222).

## *Role of OPN in inflammation*

OPN regulates several monocyte/macrophage functions such as adhesion, migration, differentiation, and phagocytosis. OPN inhibits macrophage apoptosis by interaction with integrin  $\alpha 4$  and CD44 (223,224). In addition, it regulates DC migration by CD44 and  $\alpha v$ , inhibits apoptosis and induces Th1 polarization (225,226). In T cells, it induces Th1 and Th17 polarization by inducing the expression of IFN $\gamma$  and IL-17A (227). OPN induces neutrophil migration through ERK and p38 (228). Moreover, in NK cells, it induces expansion and differentiation by mTOR activity (229).

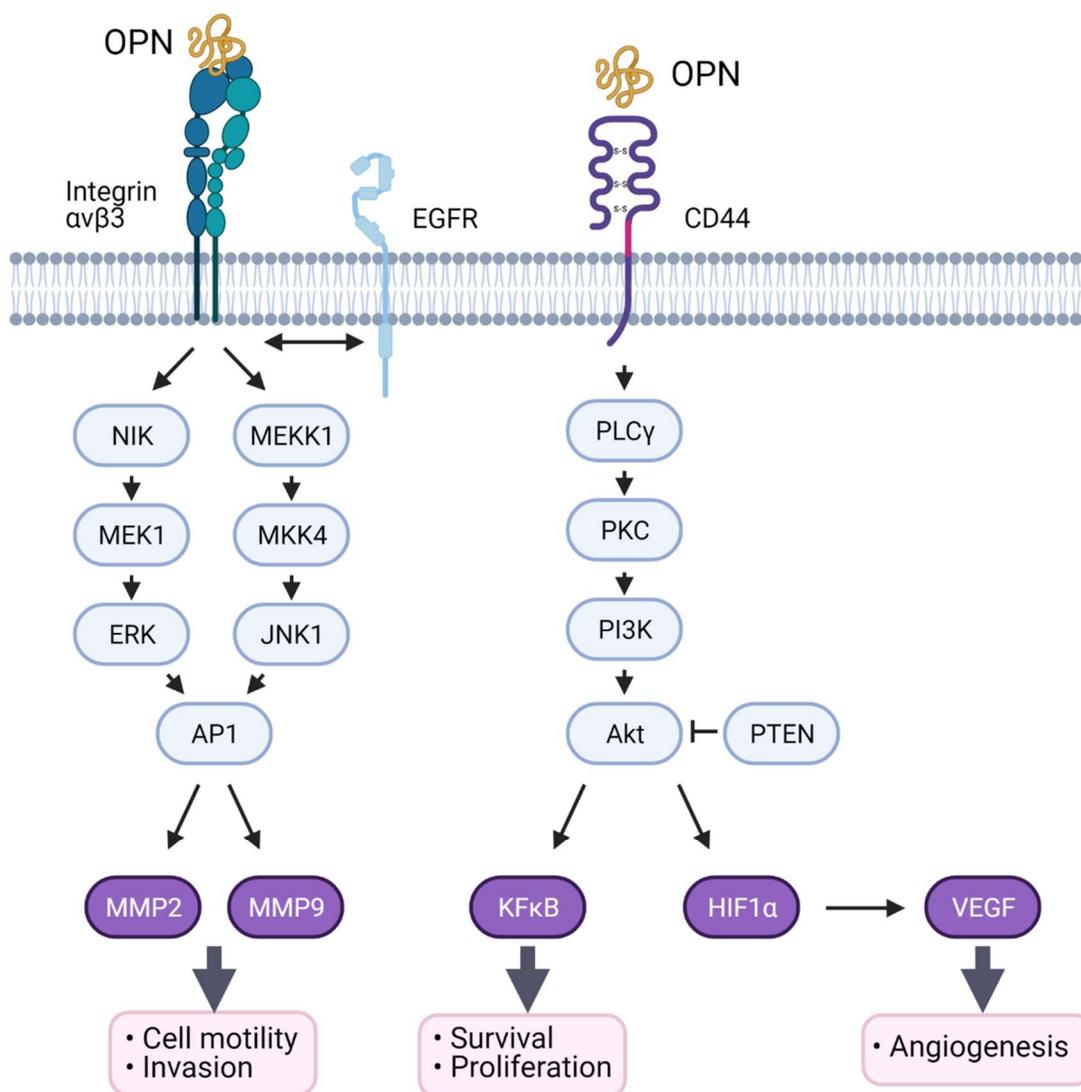


Figure 1.2. Osteopontin signaling pathways. Osteopontin binds to Integrin  $\alpha v \beta 3$  activating AP1 through nuclear factor-inducing kinase (NIK)/extracellular signal-related kinase (ERK) and mitogen activated protein kinase kinase1 (MEKK1)/c-Jun N-terminal kinase 1 (JNK) signaling pathways. Transactivation with epidermal growth factor receptor (EGFR) promotes phosphorylation of ERK and activation of AP1. Upon binding to CD44, OPN activates anti-apoptotic signals in tumor cells through phospholipase C- $\gamma$  (PLC $\gamma$ )/protein kinase C (PKC)/phosphatidylinositol 3-kinase (PI3K)/Akt pathway. Akt is also involved in activation of HIF1 $\alpha$ , leading to angiogenesis via VEGF. Image created with BioRender.com.

### *Role of OPN in tumorigenesis*

Its role in cancer progression has been thoroughly studied, mostly in solid tumors. OPN is increased in many types of cancer, such as breast, prostate, squamous cell carcinoma, melanoma, osteosarcoma and glioblastoma (230). OPN promotes tumorigenesis in many different ways: it was reported to promote invasion in melanoma (231); growth in breast cancer through the up-regulation of hypoxia-inducible factor-1 alpha (HIF1 $\alpha$ ) (232). OPN levels correlate with poorer prognosis in colorectal cancer (233). Of notice, there is a special interest in the promotion of epithelial to mesenchymal transition, which confers the migratory and invasive capabilities to epithelial cells and is crucial for solid tumors to metastasize.

### **EPITHELIAL TO MESENCHYMAL TRANSITION**

EMT is a natural process in embryogenesis, crucial for the neural crest migration (234), and in wound healing (235), however, in cancer it means a greater concern for treatment and prognosis. In the process, epithelial cells lose their epithelial characteristics, such as apical-basolateral polarization and cell-cell and cell-extracellular matrix junctions and acquire a flat morphology and migratory capabilities as well as produce enzymes that degrade the ECM, contributing to the invasion of tissues. These changes are the consequence of several transcriptional changes regulated by a set of transcription factors: SNAI1, SNAI2, TWIST1, TWIST2, ZEB1 and ZEB2 that are categorized as EMT-core genes. These transcription factors suppress the transcription of epithelial genes such as E-cadherin, EpCAM, and tight junction genes and induce the expression of mesenchymal genes such as N-cadherin, vimentin, fibronectin and metalloproteinases (236). SNAIL directly suppresses expression of E-cadherin by binding to its promoter (237). ZEB1 is also known for repressing E-cadherin expression and inducing vimentin expression (238). Furthermore, SNAIL activates expression of MMPs, facilitating the degradation of the basement membrane and invasion (239).

*Role of OPN in EMT*

As mentioned above, OPN is an inducer of EMT in cancer. In breast cancer, for example, OPN was shown to increase the transcription factors TWIST, SNAIL and SLUG. OPN promotes TWIST phosphorylation, which in turn binds to Bmi-1 and causes EMT (240). Also in hepatocellular carcinoma OPN activates TWIST, promoting invasion and decreasing cell-adhesion (241). Activation of TWIST by OPN occurs through expression of HIF1- $\alpha$ , which in turn binds to the TWIST promoter and activates its transcription, a phenomenon observed in ovarian and breast cancers (232,242). Apart from affecting cancer cells directly, OPN is also reported to influence the tumor microenvironment, enhancing metastasis (243).

## **2 AIM**

### **TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE**

It is known that CeD impairs barrier function, and that this impairment is not completely explained by the inflammatory process only, since treated patients may still present some barrier impairment and healthy relatives of celiac patients also present some impairment in barrier function. Other findings are the risk loci presenting genetic polymorphisms associated with CeD development and cell-adhesion in LPP and C1orf106 genes. In an attempt to study mechanistically the impact of those two genes in barrier impairment, we used knock-out cell models for both genes and measured barrier function parameters such as TEER, TJ protein western blot, calcium switch. In addition, to expand our findings to patients' samples, we performed protein analysis through Western Blot of LPP and C1orf106 in celiac patients.

### **THE ROLE OF OPN IN THE PATHOGENESIS OF CAC**

CAC pathogenesis is different from sporadic CRC and not as well understood. CAC develops in IBD patients and the immune response is known to contribute to tumorigenesis. Thus, we examined the RNA of CAC patients' samples using Nanostring for the human immunology panel plus a list of custom genes. The results pointed out to OPN being the most upregulated gene in CAC and we decided to investigate it further. We analyzed OPN by immunohistochemistry to learn in which cell compartment it was found in the tissue as well as proteins putatively involved in OPN signaling. Since one of the most studied functions of OPN in tumorigenesis is the induction of EMT, we investigated whether there were signs for it in CAC patients' samples in the Nanostring analysis as well as using immunohistochemistry. At last, we used epithelial intestinal cell lines HT29/B6 and T84 as models to study the impact of OPN in cells. After incubation of cells with human recombinant OPN, we performed western blot for phosphorylated proteins to determine which signaling pathway was activated by OPN. In addition, immunofluorescence staining was

### Gene expression analysis to study celiac disease and colitis-associated cancer

performed to analyze NF $\kappa$ B translocation. RT-qPCR was performed to analyze the expression of EMT-related genes and, finally, we performed an RNA-Seq analysis on the cell lines exposed to OPN to have a general idea of the changes triggered by OPN.

### **3 MATERIALS AND METHODS**

#### **REAGENTS**

Human recombinant OPN was purchased from R&D Systems Inc. (Minneapolis, USA), TNF $\alpha$ , TGF $\beta$ 1 and IFN $\gamma$  from Peprotech (Cranbury, USA), while IL-22 and IL-15 from BioLegend (San Diego, USA).

#### **CELL CULTURE**

Caco-2 cells were grown in MEM AQmedia (Invitrogen, Thermo Fisher Scientific Inc, Waltham, USA) supplemented with 15% fetal calf serum (FCS) (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) and 1% penicillin/streptomycin (Gibco, Thermo Fisher Scientific Inc., Waltham, USA); T84 cells were kept in DMEM/Ham's F12 (Corning Inc., Manassas, USA) supplemented with 10% FCS (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) and 1% penicillin/streptomycin (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) antibiotics and HT29/B6 cells were kept in RPMI (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) supplemented with 10% FCS and 1% antibiotics. All cells were kept in an incubator at 37° C and 5% CO<sub>2</sub> (Heraeus, Hanau, Germany).

#### **ESTABLISHMENT OF LPP AND C1ORF106 KNOCK-OUT CACO2 CELL LINES**

The Knock-out of LPP and C1orf106 genes in Caco-2 cells and the genetic characterization of the clones were done in the Department of Genetics of the University of Groningen by our collaborators Dr. Iris Jonkers, Dr. Sebo Withoff, Renée Moerkens and Joram Mooiweer.

Knocking-out of the LPP and C1orf106 genes in Caco-2 cells was achieved by using CRISPR/Cas9-mediated genome engineering. For that, the 20-nt sgRNA sequence containing Cas9-gRNA complex and a GFP selection marker was cloned into the pSpCas9(BB)-2A-GFP (PX458) plasmid (a courtesy of Feng Zhang, Addgene plasmid, Watertown, USA). Successful introduction of sgRNA into the plasmid was confirmed by Sanger sequencing. Following

## Gene expression analysis to study celiac disease and colitis-associated cancer

validation, the PX458-sgRNA plasmid and PX458 empty plasmid (control) were nucleofected into Caco2 cells according to the manufacturer's protocol (Lonza, Basel, Switzerland). Briefly, cells ( $1 \times 10^6$ ) resuspended in nucleofector solution were electroporated with 2  $\mu$ g plasmid and seeded in culture plates. The guideRNA sequences for LPP and C1orf106 knockout are shown in Table 3.1.

Table 3.1. guideRNA sequences for CRISPR/Cas9-mediated knockout

Target gene	Exon	sgRNA sequence 5' to 3'
C1orf106	3	Sense: TGCAGTGCACAAGCAGCAGA Antisense: TCTGCTGCTTGTGCACTGCA
LPP	3	Sense: CCACCCAAAAAGTTTGCCCC Antisense: GGGGCAAACCTTTTTGGGTGG

For Caco2 Empty Control D4, D6, E5, F10: 48 hours post nucleofection, GFP-positive cells were single-cell sorted using a SH800S cell sorter (Sony Biotechnology, San Jose, USA) and grown in separate wells in a 96-well culture plate.

For Caco2 LPP and C1orf106 knock-out cell lines and control line Empty Control B4: 48 hours post nucleofection, GFP-positive cells were sorted using a SH800S cell sorter (Sony Biotechnology, San Jose, USA) and cryopreserved in bulk. GFP-positive cells were thawed, grown, and seeded as single cells using SH800S cell sorter (Sony Biotechnology, San Jose, USA) or diluted in maintenance media, plated, and sequestered as single cells using PYREX® cloning cylinders (Corning Inc., Manassas, USA). Cloning cylinders were only used in generating Caco2 knock-out line C1orf106 Cyl2.

From residual unsorted cell suspension, DNA was isolated, the CRISPR/Cas9-targeted region of the genome was amplified by PCR and the efficiency of CRISPR/Cas9-mediated gene

disruption was analyzed by T7 Endonuclease I mismatch detection assay. The sgRNA and primer sequences are provided in Supplementary Table 3.2.

Table 3.2. Primers to test genomic disruption

<b>Target</b>	<b>Primer sequence 5' to 3'</b>
gRNA target site in C1orf106	Forward: ACAAGAAAGAAGAGGCTTAT Reverse: GACCTCTTTCTGATCACTTC
gRNA target site in LPP	Forward: CTTTATCAGGATGCATTTAG Reverse: GAGTTTGAATAAGCTGCTAA

### **CHARACTERIZATION OF GENOMIC SEQUENCE OF LPP AND C1ORF106 KNOCK-OUT CACO2 CELLS**

Single cell-derived clones were grown to confluency in 96-wells plates and DNA was isolated. The CRISPR/Cas9-targeted region of the genome was amplified by PCR and disruption of target genes was validated by Sanger sequencing. As the Caco2 cell line is tetraploid, additional sequencing was performed to validate target gene disruption on the individual chromosomes. Briefly, the targeted genomic region was amplified by PCR for each Caco2 cell clone, PCR products were purified using the Qiaquick PCR purification kit (Qiagen, Hilden, Germany). Purified PCR products were cloned into competent E. coli cells using the CloneJET PCR Cloning kit (ThermoFisher Scientific Inc., Waltham, USA), after which each E. coli cell harbors one purified PCR product that is derived from a single chromosome. E. coli cells were plated on agar plates containing Ampicillin for plasmid selection and grown overnight. For each Caco2 knock-out clone, 12 E. coli colonies were picked and grown in suspension culture in LB medium. The next day, plasmids were isolated and purified using the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, Waltham, USA). The PCR product insert in the plasmid was amplified by PCR and sequenced by Sanger sequencing, to validate gene disruption in

individual chromosomes of each knock-out line. Results of sequencing individual chromosomes are shown in Table 3.3.

### **TRANSEPITHELIAL ELECTRICAL RESISTANCE MEASUREMENTS**

Caco-2 cells were plated in 12 mm millicell inserts (Millipore) with 4  $\mu$ m pore. TEER was measured using chopstick electrodes in different days for up to 19 days. Total resistance was corrected for the resistance of the empty inserts and the average of 8 biological replicates per time-point was obtained for each individual experiment. Statistical analysis was performed in Graphpad Prism v5 using the 2-way ANOVA test with Bonferroni post-test comparing all clones to the control clones EC B4 and EC D4.

### **PROTEIN EXTRACTION FROM CELL LINES**

Cells were washed with cold PBS + and then lysed using total lysis buffer (10 mM Tris-Cl pH7.5; 150 mM NaCl; 0.5% Triton X-100; 0.1% SDS) supplemented with phosphatase and protease inhibitors. Cells were harvested using a cell scraper, transferred to a microcentrifuge tube, and then left on ice for 1h, being vortexed every 10 min. Cells were then centrifuged at 12000 g at 4°C for 10 minutes. The supernatant was collected to a new microcentrifuge tube and kept at -20° C.

### **PROTEIN EXTRACTION FROM TISSUE SAMPLES**

Biopsies were placed in glass Teflon dounce homogenizers on ice and lysed with either total lysis buffer (10 mM Tris-Cl pH7.5; 150 mM NaCl; 0.5% Triton X-100; 0.1% SDS) supplemented with protease and phosphatase inhibitors. The samples were homogenized with the dounces until no fragments could be seen, then were transferred to syringes and passed through 0.8 mm needles for 10 times and then for insulin 0.3 mm needles for 10 times to further homogenize the tissue. Samples were centrifuged at 12000 g and 4°C for 10 minutes.

Table 3.3. Characterization of genomic sequence

Clone	Total # PCR products sequenced (out of 12)*	Total # WT alleles sequenced	# different alleles sequenced (out of 4)	Allele 1	Allele 2	Allele 3	Allele 4
LPP B5	7	0	3	Del. G	Del. T; Transition G to A	Del. TG	
LPP B11	8	0	2	Del. G	Del. TG		
C1orf106 Cyl 2	5	0	3	Del. 100 bp	Del. C	Ins. T	
C1orf106 C2	7	0	4	Del. 100 bp	Del. C	Ins. T	Ins GTGCA

\* Result of sequencing an allele on one of the four chromosomes (PCR product) 12 times via subcloning in *E. coli*. The lower numbers are due to sequencing an empty vector that self-closed without a PCR product insert.

## WESTERN BLOT

Protein lysates were quantified using the Pierce BCA (Invitrogen) kit and read at the Spectrophotometer at 640nm. Protein samples were prepared in Laemmli sample buffer containing 5%  $\beta$ -mercaptoethanol in order to achieve 15 ng of protein in 10 $\mu$ l of sample. The electrophoretic separation was performed in pre-cast acrilamyde gels (Bio-Rad Laboratories Inc., Hercules, USA), at 100V. Proteins were transferred to Polyvinylidene fluoride membranes (Perkin Elmer, Weiterstadt, Germany) in a semi-dry Fast Blot system (Bio-Rad Laboratories Inc., Hercules, USA). Unspecific protein epitopes were blocked by a PVP-40 solution at 1% and 0.05 SDS, for 1h. Membranes were incubated in overnight at 4°C with primary antibodies (Table 3.4), washed with TBS-Tween 0.05% and incubated with secondary antibodies (Table

3.4) for 2h before being washed again with TBS-T. Membranes were exposed to the chemiluminescent reagent Lumi-LightPLUS western blotting kit (Roche, Basel, Switzerland) for 1-3 minutes and then developed at the Fusion FX7 (VilberLourmat, Eberhardzell, Germany). Densitometric analysis was performed using the Image Studio™ Lite (LI-COR Biosciences, Lincoln, Nebraska USA).

### **CALCIUM SWITCH ASSAY**

Caco-2 cells were cultivated in Millicell 12 mm diameter, 4 µm pore, cell inserts (MilliporeSigma, Burlington, USA) for 2 weeks before the start of the experiment. Cells were washed 4 times with PBS- and then put in DMEM calcium-free (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) with 2.5% FCS low grade Calcium, Glutamax 1% (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) and 1% penicillin/streptomycin (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) for 6 hours to disrupt cell adhesions. After 6 hours, 1.8 mM CaCl<sub>2</sub> was added to the media and cells were left to re-establish cell contacts. TEER was measured in different time-points with chopstick electrodes for 8 biological replicates for each clone.

Table 3.4: Antibodies used for Western Blot Analysis

Name	Company	dilution	species
Anti-ERK1/2	Cell signaling	1:1000	Rabbit
Anti-phospho ERK1/2 (Thr 202/Tyr 204)	Cell signaling	1:1000	Rabbit
Anti-AKT	Cell signaling	1:1000	Rabbit
Anti-phospho AKT (Tyr 308)	Cell signaling	1:1000	Rabbit
Anti-STAT3	Cell signaling	1:1000	Rabbit
Anti-phospho STAT3 (Tyr 705)	Cell signaling	1:1000	Rabbit
Anti- $\beta$ -actin	Invitrogen	1:5000	Mouse
Anti-LPP	Cell signaling	1:1000	Rabbit
Anti-C1orf106	Abcam	1:1000	Rabbit
Anti-C1orf106	Atlas	1:500	Rabbit
Anti-Claudin-1	Invitrogen	1:1000	Rabbit
Anti-Claudin-2	Invitrogen	1:1000	Rabbit
Anti-Claudin-3	Life Technology	1:1000	Rabbit
Anti-Claudin-4	Invitrogen	1:1000	Mouse
Anti-Claudin-5	Invitrogen	1:1000	Mouse
Anti-Claudin-7	Invitrogen	1:1000	Rabbit
Anti-Claudin-8	Invitrogen	1:1000	Rabbit
Anti-occludin	Invitrogen	1:1000	Rabbit
Anti-Par3	Millipore	1:1000	Rabbit
Anti-E-cadherin/clone E36	BD	1:1000	Mouse
Anti-Rabbit conjugated to peroxidase	Jackson Immunology	1:10000	Goat
Anti-Mouse conjugated to peroxidase	Jackson Immunology	1:10000	Goat

## IMMUNOFLUORESCENCE

Cells were cultured in Millicellinserts (MilliporeSigma, Burlington, USA) were washed with PBS+ and fixed with 2% paraformaldehyde (PFA) for 15 minutes, then stored in PBS+ at 4°C. Cells were permeabilized with 5% TritonX-100 for 5 minutes, blocked with blocking solution (PBS+ 6% Goat serum) for 30 minutes. The incubation with primary antibodies (Table 3.5) was

performed for 1h at 37°C, followed by wash with blocking solution and incubation with secondary antibody or phalloidin (Table 3.5) again for 1h at 37°C. Then, cells were incubated with DAPI diluted 1: 2000 in PBS for 10 min, room temperature protected from the light, washed in PBS then washed in distilled water then mounted to glass slides with Mount Fluor (Thermo Fisher Scientific Inc., Waltham, USA). Slides were analyzed in a LSM780 confocal microscope (Carl Zeiss AG, Oberkochen, Germany).

Table 3.5: Antibodies used for Immunofluorescence

Name	Company	dilution	species
Anti-phospho-P65	Cell signaling	1:100	Rabbit
Phalloidin 488	Dyomics	1:100	
Phalloidin 647	Dyomics	1:100	
Anti-Mouse secondary antibody conjugated to AlexaFluor 488	Life Technology	1:250	Goat
Anti-Rabbit secondary antibody conjugated to AlexaFluor 488	Life Technology	1:250	Goat

## RNA EXTRACTION FROM CELL LINES

Total RNA was extracted using the *mirVana*<sup>™</sup> mRNA Isolation Kit (Thermo Fisher Scientific Inc., Waltham, USA) according to manufacturer's recommendations. Cells were scraped with Lysis Binding buffer, and then the RNA Homogenate solution was added at a 1/10 of the Lysis buffer volume. Samples were vortexed and placed on ice for 10 minutes. A volume equal to the Lysis buffer of Phenol Chloroform was added, then samples were thoroughly vortexed and centrifuged at 10000 x g for 5 minutes at room temperature. The aqueous phase was collected in a new tube and 1.25 volume of 100% ethanol ACS was added. Samples were then

transferred to filter cartridges and washed once with washing buffer 1 and twice with washing buffer 2/3. The RNA was then eluted in nuclease-free water at 95 °C and stored at -80 °C.

### RNA QUANTIFICATION AND CDNA SYNTHESIS

RNA quantification was performed using the NanoDrop 1000 (Thermo Fisher Scientific Inc., Waltham, USA). 1µg of total RNA was used to make cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, USA) according to the manufacturer's instructions. Reactions were prepared according to Table 3.6 and the reverse transcription reaction was performed according to the Steps described in Table 3.7.

Table 3.6: Reverse transcription mixes

	Component	Volume per reaction
RT Master Mix	10x RT Buffer Mix	2 µl
	10x Random Primers	0.8 µl
	25x dNTP Mix (100mM)	2 µl
	MultiScribe™ Reverse Transcriptase (50 U/µl)	1 µl
	Nuclease-free H <sub>2</sub> O	4,2 µl
RNA sample	RNA sample	up to 10 µl
	Nuclease-free H <sub>2</sub> O	Q.S.* to 10 µl
	Total per reaction	20 µl

\*Q.S. = Quantity sufficient

Table 3.7: Reverse transcription cycles

Steps	Temperature	Time
1	25 °C	10 minutes
2	37 °C	120 minutes
3	85 °C	5 minutes
4	4 °C	Hold

## RECRUITMENT OF PATIENTS AND SAMPLES COLLECTION

Patient recruitment and targeted tissue dissection of paraffin-embedded samples were performed by Maximilian Sehn.

Between January 2018 to June 2020 at the Charité Campus Benjamin Franklin in Berlin celiac patients were recruited regardless of disease status and control patients with non-celiac gastrointestinal complains unrelated to the duodenum. Exclusion criteria were patients younger than 18 years, inflammatory bowel disease, oncology, or any non- celiac condition affecting the duodenum. All patients signed a written consent with the ethical approval *EA4/116/18*.

Duodenal samples were obtained from included patients during endoscopic procedure and were kept in MEM (Invitrogen, Thermo Fisher Scientific Inc, Waltham, USA) supplemented with 10% FCS (Gibco, Thermo Scientific Inc., Waltham, USA) and 1% antibiotics (Gibco, Thermo Scientific Inc., Waltham, USA) and placed on ice, until further processing.

For the colitis-associated cancer project, a total of 60 patients who underwent surgery for colectomy at Charité campus Benjamin Franklin between 2005 and 2015 for one of the following conditions: Crohn's colitis, ulcerative colitis, Crohn's associated cancer, ulcerative colitis associated cancer, sporadic colorectal carcinoma and controls (diverticulitis patients). Each group included ten patients. FFPE (formalin-fixed-paraffin-embedded) samples were

obtained and significant areas of colonic mucosa containing inflammation and/ or carcinoma were identified together with a GI- pathologist. The areas of interest were marked microscopically for targeted tissue dissection. The marked areas were then dissected and sliced for hematoxylin and eosin (H&E) staining, immunohistochemical staining and total RNA extraction.

### **RNA EXTRACTION OF FORMALIN-FIXED-PARAFFIN-EMBEDDED SAMPLES**

RNA extraction from FFPE material and the Nanostring analysis were performed by Maximilian Sehn and Hedwig Lammert.

RNA extraction was subsequently performed using the RNA FFPEasy™-Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Extracted RNA quality and quantity were then assessed with NanoDrop™ (Thermo Scientific Inc., Waltham, USA) and Qubit™ (Thermo Scientific Inc., Waltham, MA).

For the amplification-free NCounter™ RNA expression analysis (Nanostring, Seattle, USA) the commercially distributed Human Immunology v2 Panel containing 594 gene targets was used and a list of 30 custom genes (Table 3.8) was added to the analysis. RNA expression analysis was carried out according to the manufacturer's protocol.

### ***IN SILICO* ANALYSIS OF GENE EXPRESSION DATA**

Nanostring raw data was submitted to the nSolver software (Nanostring, Seattle, USA) where the raw data was normalized according to endogenous control genes and differential expression was assessed for different comparisons between groups. In total, 9 comparisons were made: IBD vs CTRL, CAC vs IBD, CAC vs CRC, CAC vs CTRL, CRC vs CTRL, UC vs CTRL, CD vs CTRL, CD-CAC vs CD, UC-CAC vs UC.

## Gene expression analysis to study celiac disease and colitis-associated cancer

Ratios were then uploaded to Ingenuity Pathway Analysis (Qiagen, Hilden, Germany) and the core analysis was performed. Among the tools used in the core analysis there were the Disease and Functions, Upstream regulator Analysis and Mechanistic Networks tools.

Heatmaps, volcano plots, principal component analysis were made in RStudio software using R version 3.6.0 R. <https://www.R-project.org/>.

Table 3.8. Custom genes added to the Nanostring panel

<b>HGNC gene name</b>	<b>Probe NSID</b>	<b>Total Isoforms</b>	<b>Isoforms Not Hit By Probe</b>
ACTB	NM 001101.2:1010	2	
AKT1	NM 001014432.1:1275	6	
FAPC	NM 000038.3:6850	3	
BUB1B	NM 001211.4:835	1	
CCND1	NM 053056.2:690	2	
CDH1	NM 004360.2:535	6	
CDK8	NM 001260.1:370	6	XM 011534865.1
CLDN2	NM 020384.3:2540	3	
CLDN7	NM 001307.3:175	3	
CLDN8	NM 199328.2:805	1	
CRB3	NM 139161.3:300	3	
DLG1	NM 001098424.1:1460	27	
F2RL2	NM 004101.2:475	2	
FLT1	NM 002019.4:530	5	
GRHL2	NM 024915.3:1818	4	
GSK3B	NM 002093.2:925	4	
KDR	NM 002253.2:1420	1	
KRAS	NM 004985.3:327	4	
MYC	NM 002467.3:1610	1	
OCLN	NM 002538.3:5130	3	
PAWR	NM 002583.2:824	5	XR 944560.1;XR 944561.1
PDCD4	NM 014456.3:1115	3	
PFKFB3	NM 001145443.1:495	7	
PIK3CA	NM 006218.2:2445	3	
PMM1	NM 002676.2:497	8	
PRKCH	NM 006255.3:850	5	
PRKCZ	NM 002744.4:771	14	
PSMB6	NM 002798.1:695	2	
PTEN	NM 000314.4:1351	9	
VEGFA	NM 001025366.1:1325	20	

## **IMMUNOHISTOCHEMISTRY OF FFPE SLIDES**

Immunohistochemical staining was performed in the iPATH. Berlin by Dr. Anja Kühl and Simone Spieckermann.

Paraffin sections (1-2  $\mu\text{m}$ ) were dewaxed prior to heat-induced epitope retrieval using citrate buffer (pH 6). Sections were rinsed with running tap water and incubated with antibodies directed against CD44 (clone IM7, Cell Signaling Technology, Danvers, USA), beta-catenin (clone 6B3, Cell Signaling Technology, Danvers, USA), EpCAM (clone E6V8Y, Cell Signaling Technology, Danvers, USA), pSMAD3 (polyclonal rabbit anti-human, Abcam, Cambridge, UK), at room temperature for 30 minutes. Antibodies were detected using EnVision+ Single Reagent (HRP. Mouse or HRP.rabbit; Agilent Technologies, Santa Clara, USA). HRP was visualized with diaminobenzidine (DAB) as chromogen (Agilent Technologies, Santa Clara, USA). Nuclei were counterstained with hematoxylin (Merck, MilliporeSigma, Burlington, USA) and slides coverslipped with glycerol gelatin (Merck, MilliporeSigma, Burlington, USA). Primary antibodies were omitted in negative control sections.

## **OPN EXPOSURE OF INTESTINAL CELL LINES INFILTERS**

T84 and HT29 cells were seeded in Millicell 12 mm diameter, 3  $\mu\text{m}$  pore, cell inserts (MilliporeSigma, Burlington, USA) for 1 week before incubation with their respective media with 1% FCS (Gibco, Thermo Fisher Scientific Inc., Waltham, USA) for 24h. Cells were then incubated with media containing OPN on both apical and basolateral sides in concentrations and for time-points optimized for each experiment.

Protein analysis was performed by incubating cells with 200 or 500 ng/ml of OPN (R&D Systems Inc., Minneapolis, USA) for 10, 30 and 60 minutes and then cells were lysed for protein extraction. TNF $\alpha$  2000 U/ml and IL-22 10 ng/ml (Peprotech, Cranbury, USA) were used as positive controls for the phosphorylation of ERK1/2 and STAT3, respectively.

For the analysis of phosphorylated P65, Caco-2 cells were cultivated in Millicell inserts (MilliporeSigma, Burlington, USA) for 2 weeks and then incubated with either 200 ng/ml of OPN or 5000 U/ml of TNF $\alpha$  for 20 minutes and later fixed for immunofluorescent staining.

### **OPN TREATMENT OF INTESTINAL CELL LINES IN PLATES**

Intestinal cell lines T84 and HT29/B6 were seeded  $2 \times 10^5$  and  $1 \times 10^5$  cells respectively in 12-well plates and left to attach for 2 days. Normal media was then exchanged for media containing 1% FCS overnight and then incubated with media containing 200 ng/ml of OPN (R&D Systems Inc., Minneapolis, USA) for 3h and 24h. RNA was then extracted from the cells and TGF $\beta$ 1 (Peprotech, Cranbury, USA) and IL-22 (Biolegend, San Diego, USA) was used as positive control for the activation of expression of SNAI1, SNAI2 and TWIST1 genes.

### **REAL TIME-QUANTITATIVE PCR**

Real time-qPCR reactions were performed using 1  $\mu$ L of cDNA template, 1  $\mu$ L of the desired probe (Table 3.9), 10  $\mu$ L of RT-qPCR MasterMix (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, USA) and nuclease-free water to a final volume of 20  $\mu$ L. Comparative CT reactions were performed in triplicates using the 7500 Fast Real-Time PCR System instrument (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, USA). Calculations for gene expression changes were performed using the  $2^{-\Delta\Delta CT}$  method.

Table 3.9: TaqMan Probes

<b>Target</b>	<b>Probe Name</b>	<b>Species</b>	<b>Fluorochrome</b>	<b>Company</b>
SNAI1	Hs00195591_m1	Human	FAM	Applied Biosystems
SNAI2	Hs00161904_m1	Human	FAM	Applied Biosystems
TWIST	Hs00161904_m1	Human	FAM	Applied Biosystems
ACTB	Hs01060665_g1	Human	FAM	Applied Biosystems
MMP7	Hs01042796_m1	Human	FAM	Applied Biosystems
GAPDH	Hs99999905_m1	Human	FAM	Applied Biosystems

## **RNA-SEQ**

RNA-Seq analysis was made by Novogene Co. Ltd. as described below.

RNA-Seq experiments were performed in T84 and HT29 cells exposed to osteopontin (as described above) for 3h and 24h in 3 biological replicates. After incubation, total RNA was extracted using the *mirVana*<sup>™</sup> mRNA Isolation Kit (Thermo Fisher Scientific Inc., Waltham, USA) (as described before), quantified and then sent to Novogene (Novogene Co. Ltd., Beijing, China) where the RNA-Seq analysis was performed.

In short, using 1 µg of total RNA, libraries were made using NEBNext<sup>®</sup>Ultra<sup>™</sup>RNA Library Prep Kit for Illumina<sup>®</sup> (NEB, USA) according to manufacturer's recommendations and index codes were added to each sample. mRNA was purified using poly-T oligo-attached magnetic beads. Fragmentation was performed in NEBNext First Strand Synthesis Reaction Buffer (5x) using divalent cations in high temperature. First strand cDNA was generated using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA was subsequently performed using DNA Polymerase I and RNase H. Overhangs were converted into blunt ends via exonuclease/polymerase activities. Following adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop were ligated to prepare for hybridization. The fragments were purified using AMPure XP system (Beckman Coulter, Beverly, USA) to enrich DNA fragments of 150~200 bp in length. Samples were incubated with 3 µl of USER Enzyme (NEB, USA) at 37 °C for 15 min followed by 5 min at 95 °C before PCR. PCR was carried out using Phusion High-Fidelity DNA polymerase, Universal PCR primers Index (X) Primer. Lastly, PCR products were purified (AMPure XP systems) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Clustering of index-coded samples was done on a cBot Cluster Generation System using PE Cluster Kit cBot-HS (Illumina Inc, San Diego, USA) following manufacturer's instructions. The library preparations were then sequenced on an Illumina platform and paired-end reads were generated.

## **BIOINFORMATICS OF THE RNA-SEQ DATA**

Bioinformatic analysis was performed by our collaborator Dr. January Weiner from CUBI BIH.

RNA-Seq reads were aligned to the Gr38 human genome using STAR aligner (244). Count data were analyzed using the R package DESeq2 (245) and a two-factor (treatment/time point) model with interaction. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure (246). Gene set enrichment analysis was performed using the R package tmod (247).

## **STATISTICS**

Statistical analyses were performed using GraphPadPrism® 5. Mean  $\pm$  standard deviation was plotted, unless stated otherwise. The standard deviation was calculated for 3 or more independent experiments. For TEER analysis of Caco-2 clones, statistical analysis was performed using 2-way ANOVA with Bonferroni posttest. The unpaired student's t-test was used to determine significant differences between two groups in cell line experiments. Mann Whitney test was employed in the analysis with patients' samples in Nanostring counts and immunohistochemical staining intensity analysis. Were considered statistically significant (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

## **DEVICES AND CONSUMABLES**

All devices, the version and the supplier are listed in Table 3.10

Table 3.10. Devices.

<b>Device</b>	<b>Version</b>	<b>Supplier</b>
Centrifuge	PerfectSpin 24R Refrigerated Microcentrifuge	PEQLAB Biotechnologie GmbH, Germany
Centrifuge	Hermle z233MK	Wehingen
Chemiluminescence signal detector	Fusion FX7	Vilber Lourmat, Germany
Fluorometer	Qubit® Fluorometer	Thermo Fisher Scientific Inc., USA
Fragment analyzer	5200 Fragment Analyzer System	Agilent Technologies, USA
Heating block	AccuBlock™	Labnet international, Inc., Corning Inc., USA
Heating block	Digital Dry Bath	Labnet international, Inc., Corning Inc., USA
Incubator for cell culture		Heraeus, Germany
Laminar Flow Workbench	SAFE 2020	Thermo Electron Corporation, Thermo Fisher Scientific Inc., USA
Laser scanning microscope	LSM 780	Carl Zeiss Jena GmbH, Germany
Magnetic stirring		Merck, Berlin
Nanodrop	NandoDrop 1000	Thermo Fisher Scientific Inc., USA
Nanostring machine	nCOUNTER MAX	Nanostring Technologies, USA
pH meter	HI 9017 microprocessor	Hanna Instruments, Germany
Power supply	Blotting device 200/2.0	Bio-Rad Laboratories GmbH, Germany
Resistance measuring device		Institut für Klinische Physiologie, CBF, Germany
RT-qPCR device	7500 Fast Real-Time PCR System instrument	Applied Biosystems, Thermo Fisher Scientific Inc., USA
Scale		Sartorius, Germany
Shaker	Rocking platform	Biometra, Germany
Shakers	Rocking platform	VWR, Germany

## Gene expression analysis to study celiac disease and colitis-associated cancer

Thermocycler	PeqSTAR	Peqlab Biotech. GmbH, Germany
Ussing-chambers		Institut für Klinische Physiologie, CBF,
Vortex device	LSE™	Corning Inc. USA
Vortex device	Vortex mixer	Corning Inc. USA
Water bath	1002	GFL, Germany

Table 3.11. Chemicals and kits

Chemicals/kits	Supplier
4', 6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich Chemie GmbH, Germany
Agarose	Invitrogen, USA
Ammonium persulfate (APS)	MilliporeSigma, USA
BCA-Protein Assay (Reagents A and B)	Pierce, USA
Blotting grade blocker non-fat dry milk	Carl Roth, Germany
Mercaptoethanol	Clontech, Germany
Bovine serum albumin (BSA)	Biomol GmbH, Germany
DMSO (cell culture quality)	Biochrom AG, Germany
Dulbecco's PBS with Mg <sup>2+</sup> /Ca <sup>2+</sup>	Gibco Inc., U.S.A
Dulbecco's PBS without Mg <sup>2+</sup> /Ca <sup>2+</sup>	Gibco Inc., U.S.A
Ethanol	J.T. Backer, Netherlands
FBS	Biochrom, Germany
Glucose	
Glycin	Carl Roth GmbH, Germany

## Gene expression analysis to study celiac disease and colitis-associated cancer

Immersion oil for microscopy	VWR International GmbH, Germany
Immomount	Thermo Fisher Scientific Inc., USA
Lumilight Western Blotting Kit	Roche, Switzerland
Methanol	Merck, Germany
<i>mirVana</i> <sup>™</sup> mRNA Isolation Kit	Thermo Fisher Scientific Inc, USA
Penicillin/streptomycin (P/S)	Carl Roth GmbH , Germany
Pierce Protease Inhibitor mini tablets	Roche, Switzerland
Polyacrylamide mix (30%)	Serva, Germany
Protein-Marker PageRuler	Thermo Fisher Scientific Inc., USA
Sodium azide	Carl Roth GmbH, Germany
Sodium chloride	Serva, Germany
Sodium-Dodecyl sulfate (SDS)	MilliporeSigma, USA
Tetramethylethylenediamine (TEMED)	Thermo Fisher Scientific Inc., USA
Tris	Merck, Germany
Tris Base	Calbiochem, Germany
Tris-HCl 0.5M, pH 6.8	Biorad, Germany
Tris-HCl 1.5M, pH 8.8	Biorad, Germany
Triton X-100	Roche, Switzerland
Trypsin/EDTA	Biochrom, Germany
Tween 20	MillipoerSigma, USA

Table 3.12. Consumables

Consumables	Supplier
12-well-tissue culture plate	TPP Techno Plastic Products AG, Germany

## Gene expression analysis to study celiac disease and colitis-associated cancer

15 ml PPN tube	Greiner, Germany
24-well-tissue culture plate	TPP Techno Plastic Products AG, Trasadingen, Germany
25 cm <sup>2</sup> -tissue culture flask	Corning Incorporated, NY, USA
50 ml PPN tube	Nunc, Germany
6-well-tissue culture plate	TPP Techno Plastic Products AG, Germany
75 cm <sup>2</sup> -tissue culture flask	Corning Inc., USA
Cell and tissue culture dishes	Nunc, Germany
Cell scraper	Coster, Corning Inc., USA
Cryotubes	Corning Inc., USA
Microscope slides	Menzel-Gläser, Thermo Scientific Inc, USA
Microtiter plate 96 wells	Sarstedt, Germany
Mini-PROTEAN TGX, Stain-free gels	Bio-Rad Laboratories Inc.
Nitrocellulose membrane (Amersham Hybond)	GE Healthcare, UK
Pipettes	Eppendorf, Germany
PVDF transfer membrane	Perkin Elmer, Germany
Reaction tubes	Eppendorf, Germany
Standard tips 10, 200, 1000 µl	Eppendorf, Germany
Trans-Blot Turbo Midi 0.2 µm PVDF Transfer Packs	Bio-Rad Laboratories Inc.
Transwell filters (Millicell-HA, 0,6 cm <sup>2</sup> )	MilliporeSigma, USA

---

## 4 RESULTS

### TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE

#### Establishment of knock-out clones of Caco-2 cells

As a model to study the importance of the LPP and C1orf106 genes in barrier function of intestinal epithelial cells, knock-out (KO) clones were established in Caco-2 cells using the CRISPR-Cas 9 at the department of genetics of the University of Groningen. Of all the acquired clones, control clones B4 and D4, LPP KO clones B5 and B11 and C1orf106 KO clones C2 and Cyl2 were selected for our study.

Protein evaluation of the 6 selected clones showed that the LPP KO clones B5 and B11 are depleted from LPP protein band (Figure 4.1), similarly, C1orf106 KO clones C2 and Cyl2 did not present a band for C1orf106 (Figure 4.1). Densitometric analysis of the bands is shown in Supplementary figure 1.

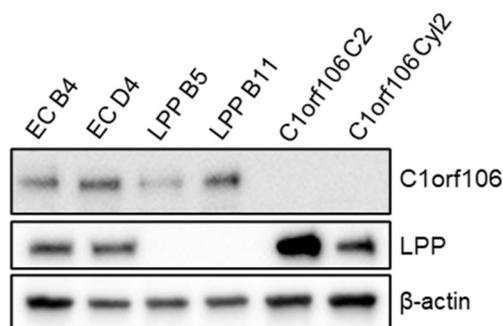


Figure 4.1: Protein analysis of Caco-2 knocked-out (KO) clones. The Caco-2 CRISPRed clones EC B4, EC D4 (controls); LPP KO clones B5 and B11 and C1orf106 KO clones C2 and Cyl2 were lysed and total protein extracted and processed for western blot analysis for LPP and C1orf106 proteins.  $\beta$ -actin was used as loading control. Figure representative of three independent experiments.

## **Characterization of the barrier function in the Caco-2 knock-out clones**

### **Transepithelial electrical resistance**

In order to determine whether the KO promoted functional changes of the tight-junction barrier, the clones were evaluated for their barrier functionality. TEER values were measured for several days starting one day after seeding. In the first week, TEER values were oscillating for all clones since the monolayer is most likely not yet formed. From the second week on culture, all clones tended to stabilize to a certain resistance value range and on the third week no significant changes in TEER in all clones except EC D4 were observed (Figure 4.2A). Control clones EC B4 and EC D4 as well as LPP B5 KO clones presented constantly high resistance values. Looking more closely into three timepoints, one for each week, LPP B11 clone, but not LPP B5, presented statistically lower resistance values (compared to EC D4) in the second week. However, C1orf106 KO clones C2 and Cyl2 were the ones which consistently presented lower TEER over the second week in culture, suggesting there is a barrier impairment effect related to the knock-out of this gene (Figure 4.2B).

### **Tight junction proteins content**

Following the TEER measurement, protein content evaluation of occludin, barrier-forming claudin-1, -3, -4, -8; claudin-7; as well as pore-forming claudin-2 were performed on the KO clones (Figure 4.3A). Even though some clones presented individual tendencies, for example, LPP B11 presented lower Claudin-4 than the others and increased occludin content, there were differences consistently seen in both clones for each gene. Of those, only the increase in Claudin-3 in the LPP KO clones was statistically significant (Figure 4.3B). C1orf106 KO clones presented a tendency for increased occludin. LPP KO clones had a tendency for higher claudin-7 (Supplementary figure 2). Changes consistently seen in both clones suggest specific relation to the lack of C1orf106 and LPP genes, respectively (Figure 4.3).

## Gene expression analysis to study celiac disease and colitis-associated cancer

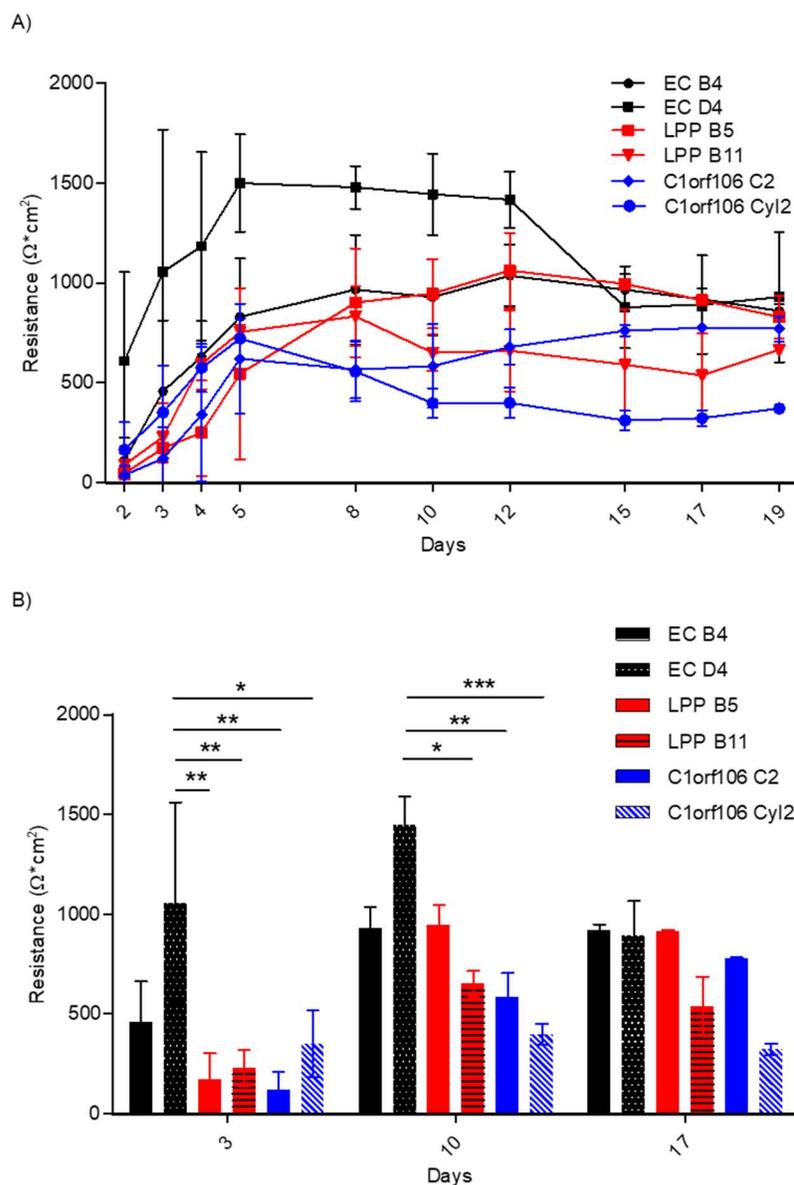


Figure 4.2: Transepithelial electrical resistance (TEER) of Caco-2 clones. Cells from the control clones EC B4 and D4, LPP knock-out clones LPP B5 and B11 and C1orf106 knock-out clones C2 and Cyl2 were seeded in filters and resistance was measured with chopstick electrodes for 19 days. A) shows a graphical representation of the TEER values for each clone overtime and B) shows 3 time-points (days 3, 10 and 17) within the 19 days of measurement where statistical calculations were performed. Statistical analysis was performed using 2-way ANOVA test with Bonferroni post-test and \* =  $p < 0.05$ ; \*\* =  $p < 0.01$  and \*\*\* =  $p < 0.001$ .

## Gene expression analysis to study celiac disease and colitis-associated cancer

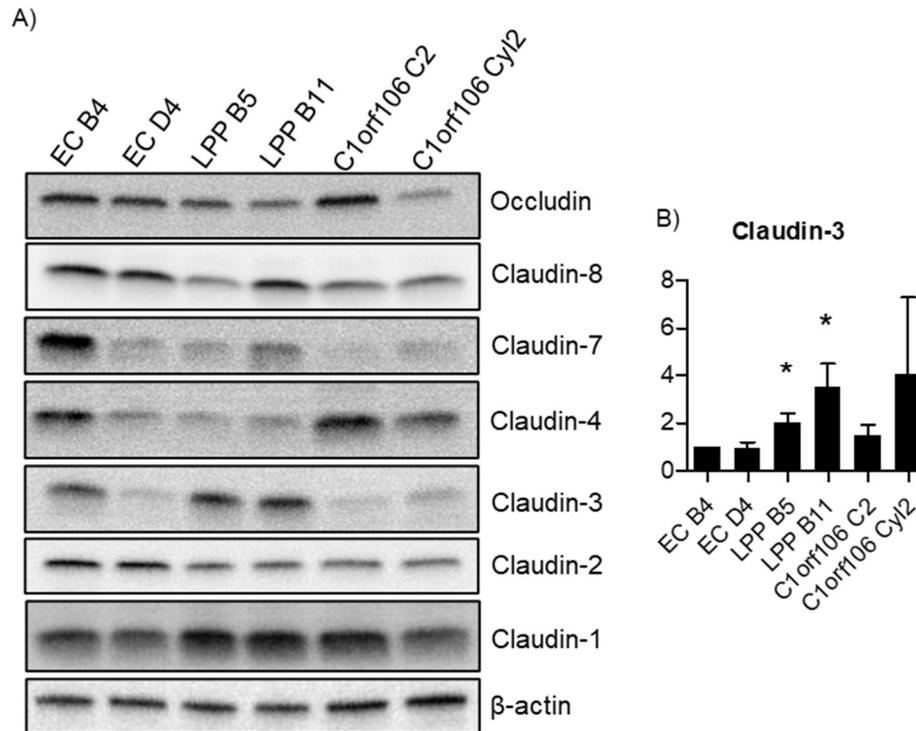


Figure 4.3: Protein content of tight junctional proteins. Caco-2 control clones EC B4 and D4, LPP knock-out clones B5 and B11, and C1orf106 knock-out clones C2 and Cyl2 were cultivated for 2 weeks in filters. Total protein was extracted and western blot analyses with antibodies against occludin, claudin-1, -2, -3, -4, -7, -8 and  $\beta$ -actin were performed A). Densitometric analysis of Claudin-3 shows significant increase in LPP knock-out clones. Statistical analysis performed using unpaired t-test. Image representative of 4 independent experiments.

### Tight junction re-assembling capacities of the knock-out clones

One of the processes that could be impaired by the loss of either LPP or C1orf106 is the re-assembling of TJ. Aiming at evaluating the ability of the KO clones to re-assemble TJ, they were submitted to a calcium-switch assay. In this assay, cells are deprived of calcium and consequently lose cell-cell adhesion, then calcium is then re-introduced, and the time needed for the re-assembling of the cell junctions is monitored by TEER measurements. In all clones, there was a drastic decrease in TEER after calcium deprivation (Figure 4.4). Following calcium replacement, resistance values gradually increased, but at different rates for the KO clones. 16 hours after

calcium re-introduction TEER levels reached a plateau where it is patent that the KO clones revealed lower resistance values than the control clones (Figure 4.4).

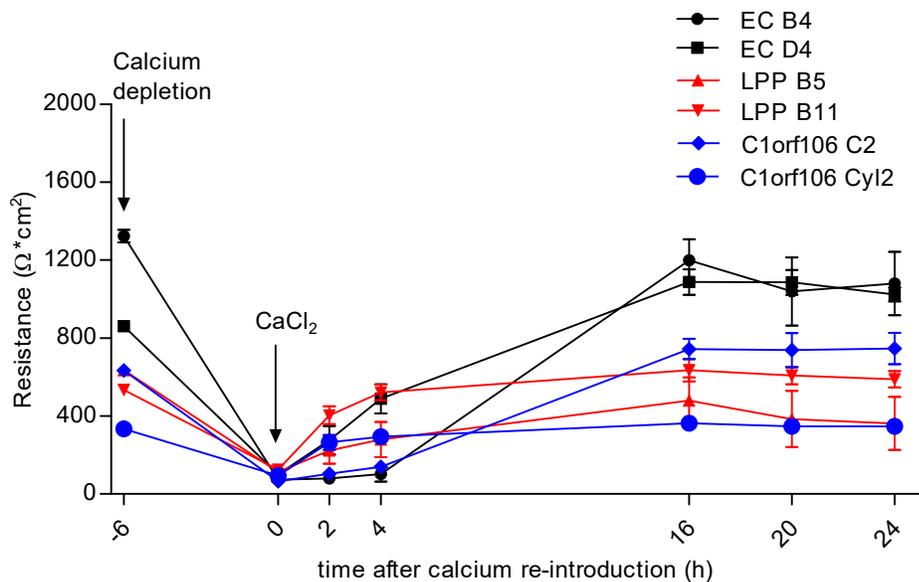


Figure 4.4: Tight junction re-assembling after calcium depletion and replacement. Caco-2 control clones B4 and D4, LPP KO clones B5 and B11 and C1orf106 KO clones C2 and Cyl2 were seeded in filters for 2 weeks, depleted from calcium for 6 hours, then calcium was replaced. Transepithelial electrical resistance was measured in different points in time for 48 hours.

## Barrier function of celiac patients

### Patients' characteristics

Decreased expression of LPP and C1orf106 were reported for a proportion of celiac patients (248). In order to investigate whether the levels of LPP and/or C1orf106 proteins in celiac patients would be associated with disease status or refractory disease, protein analysis was performed in duodenal samples from 25 patients whose characteristics are described below in table 4.1.

## Gene expression analysis to study celiac disease and colitis-associated cancer

Of those, eight were control patients who underwent endoscopic examination for causes not related to celiac disease or affecting the duodenum; seven were celiac patients responding to a gluten-free diet; eight were refractory celiac disease patients and two were active celiac

Table 4.1. Clinical characteristics of celiac patients

<b>Group</b>	<b>number of individuals</b>	<b>Gender f, m</b>	<b>Age median (range)</b>	<b>Anti-TTG IgA (U/ml)</b>	<b>Anti-TTG IgG (U/ml)</b>
Control	8	5, 3	51.5 (19 - 66)	NA	NA
GFD	7	5, 2	52.0 (27 - 68)	2.2	2.1
RCD	8	6, 2	45.5 (20 - 74)	8.1	3.8
ACTIVE	2	2, 0	39.5 (33 - 46)	57.5	12.6

TTG = Tissue transglutaminase

IgA = Immunoglobulin A

IgG = Immunoglobulin G

GFD = Gluten-free diet

RCD = Refractory celiac disease

NA = Not applicable

patients.

Protein content for LPP and C1orf106 was assessed using Western Blot (Figure 4.5A) followed by densitometric analysis of the bands which showed significant scatter for all disease groups, however, no statistical significance difference was found between groups (Figure 4.5B and C).

## Gene expression analysis to study celiac disease and colitis-associated cancer

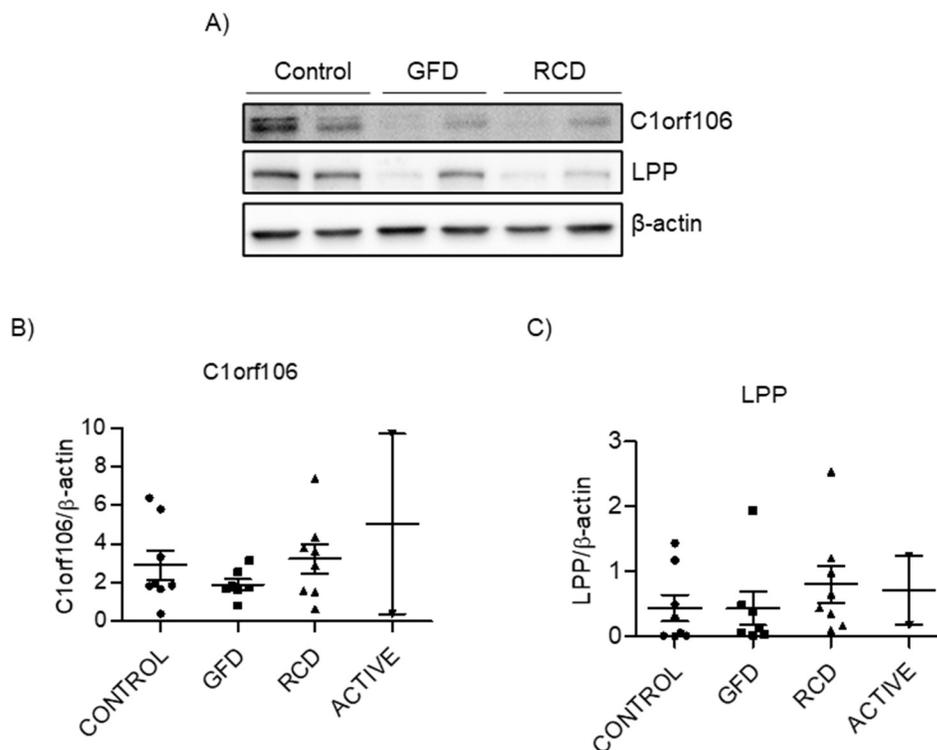


Figure 4.5: LPP and C1orf106 protein content in celiac mucosa. Protein was extracted from duodenal biopsies of celiac patients and controls and protein content of LPP, C1orf106 and  $\beta$ -actin was analyzed using western blot as shown in a representative blot (A). Densitometry of C1orf106 (B) and LPP (C) bands was performed.

## THE ROLE OF OPN IN THE PATHOGENESIS OF CAC

### Patients' characteristics

In order to better understand CAC pathogenesis, patients who underwent colostomy and were diagnosed with Crohn's disease-colitis, ulcerative colitis, colitis-associated cancer or sporadic colorectal cancer were retrospectively recruited from the Charité database after approval by the Charité ethical commission EA/1/204/14. CAC patients were categorized according to the previous IBD condition of the patients in Crohn's disease CAC (CDAC) and ulcerative colitis CAC (UCAC). For the control were selected inflammation-free borders of colon resections of diverticulitis patients. The clinical characteristics of the patients are listed in table 4.2.

### **Gene expression analysis of colitis-associated patients' samples**

In an attempt to have a comprehensive view of gene expression differences between patients' groups, a Nanostring analysis was performed of 624 genes from the human immunology panel plus 30 custom genes. For the purposes of this study, the comparisons between the CAC conditions and their underlying IBD conditions were prioritized. Heatmaps of the 20 most upregulated and 20 most downregulated genes for CDAC vs CD and UCAC vs UC show that CD and CDAC patients clustered separate (Figure 6A) and that the clustering was almost perfect for UC and UCAC patients except for 1 UC patient (Figure 4.6B). The same genes are described in Tables 4.3 and 4.4 where we observe that SPP1 was the most upregulated gene in both comparisons, despite them being totally independent from each other. Among the upregulated genes there are genes from signaling pathways such as MAPK/ERK, AKT, TGF $\beta$  and SRC. Amidst of the downregulated genes, we observed genes related to B-cells and T-cells as well as TNF receptor superfamily members.

Table 4.2. Clinicopathological characteristics of included patients

		Control	CD	UC	CRC	CDAC	UCAC
Gender f/m		5/5	6/4	3/7	4/6	7/3	5/5
Age median (range)		62 (52-80)	42 (20-61)	34.5 (19-51)	77.5 (49-86)	47 (31-90)	45.5 (36-78)
Inflammatory activity	High	6	5	6	-	3	3
	Low	-	5	4	-	1	7 <sup>4</sup>
	None	4	-	-	10	6	
Perforations		4	1	-	1	1	1
Location (inflammation/ tumor)	Pancolitis/ multilocal tumor	-	8 <sup>1</sup>	7	-	-	2
	Right hemicolon	-	-	-	4	5	1
	Transverse Colon	-	-	-	1	0	1
	Left hemicolon	-	2	3	5	4	4
	Rectum (Ileum)	-	- 3	-	-	1 -	2 -
Duration of disease (range)		-	14.5 (4-27)	8 (0-24)	- 1 patient n.a.	20 (3-37) 3 patients n.a.	18 (1-39)
Tumor type	Mucinous	-	-	-	0	6	9
	Non Mucinous	-	-	-	10	4	1
Tumor stage (T)	pT1:	-	-	-	0	1	2
	pT2:	-	-	-	3	2	1
	pT3:	-	-	-	5	2	4
	pT4:	-	-	-	2	5	3
	N0:	-	-	-	6	6	4
	N1a:	-	-	-	3	2	3
	N2:	-	-	-	1	2	3
	M0: M1:	- -	- -	- -	6 4 <sup>2</sup>	8 2 <sup>3</sup>	8 2 <sup>7</sup>
Recurrent tumor(s)		-	-	-	0	0	0
Adenoma(s) present		1	0	1	5	2	6
Steroid treatment		-	7	8	-	2 (8 n.a.)	4 (4 n.a.)
Biological treatments		-	9	8	-	1 (7 n.a.)	0 (6 n.a.)

Age and duration of disease values correspond to the median in years.

<sup>1</sup> One of the above had segmental colitis sparing the transverse colon.

<sup>2</sup> Metastases: Two liver. Once liver and peritoneum. Once lung.

<sup>3</sup> Metastases: Once liver. Once peritoneal metastases.

<sup>4</sup> Histologically reported inflammation without activity level was counted as low/moderate for this synopsis.

<sup>6</sup> One patient: Unknown tumor spread.

<sup>7</sup> Metastases: Twice peritoneal.

## Gene expression analysis to study celiac disease and colitis-associated cancer

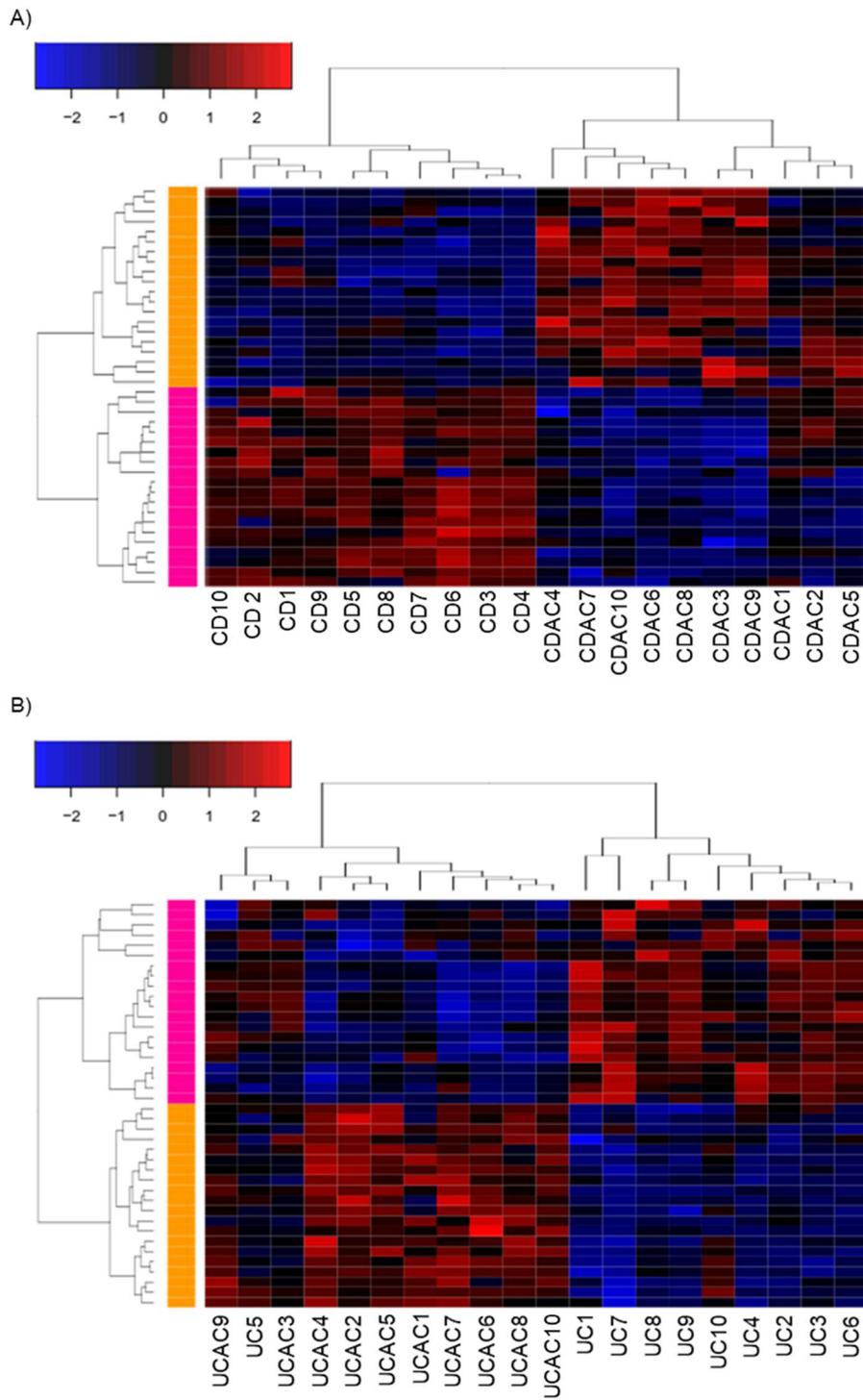


Figure 4.6: Heatmaps of the 20 most up- and downregulated genes in Nanostring. Normalized counts were acquired using the nSolver software and unsupervised heatmaps using k-means clustering algorithm and

## Gene expression analysis to study celiac disease and colitis-associated cancer

featuring the 20 most upregulated and 20-most downregulated genes were made for the comparisons CDAC vs CD (A) and UCAC vs UC (B).

Table 4.3. Most up- and downregulated genes in CDAC vs. CD.

	<b>Gene</b>	<b>Fold change</b>	<b>p-value</b>
<b><i>Upregulated genes</i></b>	SPP1	18.29	<0.0001
	FN1	6.68	<0.0001
	DUSP4	4.60	0.0001
	CLDN2	4.40	0.0344
	CCL26	3.42	0.0002
	CD276	3.01	<0.0001
	THY1	2.83	0.0086
	ICAM5	2.78	0.0102
	F2RL2	2.77	0.0097
	CLEC5A	2.74	0.0123
	MSR1	2.69	0.0185
	TGFBI	2.66	0.0006
	LIF	2.65	0.0054
	HAMP	2.62	0.0035
	PLAU	2.61	0.0272
	TNFSF4	2.3	0.0023
	C6	2.26	0.0228
	CCND1	2.26	0.0009
	ZEB1	2.21	0.0018
	EGR2	2.19	0.0064
<b><i>Downregulated genes</i></b>	PIGR	-23.69	0.0007
	CXCL13	-11.11	0.0003
	MS4A1	-9.77	<0.0001
	CD79A	-7.51	0.0003
	NOS2	-7.34	0.0024
	CR2	-6.58	0.0004
	CCL19	-6.44	0.0001
	CD45 (RA)	-6.33	>0.0001
	TNFRSF17	-6.12	0.0003
	TNFRSF13B	-5.38	0.0002
	CD19	-4.88	0.0001
	IDO1	-4.22	0.0198
	TNFRSF13C	-4.21	>0.0001
	IRF4	-4.14	0.0006
	CD27	-4.08	0.0005
	PLA2G2A	-3.88	0.0007
	SLAMF7	-3.8	0.0009
	CCR7	-3.63	0.0001
	BTLA	-3.57	<0.0001
	BLNK	-3.45	<0.0001

Table 4.4. Most up- and downregulated genes in UCAC vs. UC.

	<b>Gene</b>	<b>Fold change</b>	<b>p-value</b>
<b><i>Upregulated genes</i></b>	SPP1	8.41	0.0007
	CCL26	2.91	0.0220
	GRHL2	2.71	0.0008
	DUSP4	2.67	0.0296
	CCND1	2.65	0.0143
	FN1	2.22	0.0385
	CEACAM6	2.05	0.0319
	CD9	1.78	0.0128
	SRC	1.70	0.0114
	TRAF2	1.68	0.0082
	CD276	1.63	0.0015
	IL13RA1	1.47	0.0089
	AKT1	1.46	0.0049
	PTK2	1.45	0.0186
	CDK8	1.38	0.0154
	GPI	1.36	0.0411
	PPIA	1.34	0.0404
	TRAF4	1.32	0.0323
	CD46	1.24	0.0444
	<b><i>Downregulated genes</i></b>	DEFB4A	-18.33
S100A8		-14.97	<0.0001
PIGR		-11.10	0.011
NOS2		-10.15	0.0002
S100A9		-8.56	0.0002
MS4A1		-8.20	0.0004
IRF4		-7.90	<0.0001
CXCL13		-7.31	0.0014
TNFRSF13B		-7.26	0.0013
IL1B		-7.00	0.0004
CD45 (RA)		-6.87	0.0003
CD19		-6.73	0.0001
CD79A		-5.89	0.0006
CD27		-5.72	0.0004
ARG1		-5.47	0.0023
TNFRSF17		-5.45	0.0003
CD79B		-5.42	0.0002
CCL8		-5.36	<0.0001
CSF3R		-4.89	0.0001
SELL		-4.89	0.0001

### OPN might account for a poorer prognosis

Since OPN was independently found as the most upregulated genes in both comparisons, we decided to further investigate it. Analysis of the osteopontin counts from the Nanostring evaluation shows significant scattering especially for the CAC conditions, with two separate groups of patients expressing either high or low osteopontin (Figure 4.7A). The analysis of the survival data from the CAC patients regarding high or low OPN expression showed a tendency for lower survival rate of the high OPN-expressing patients (Figure 4.7B) even though it was not statistically significant.

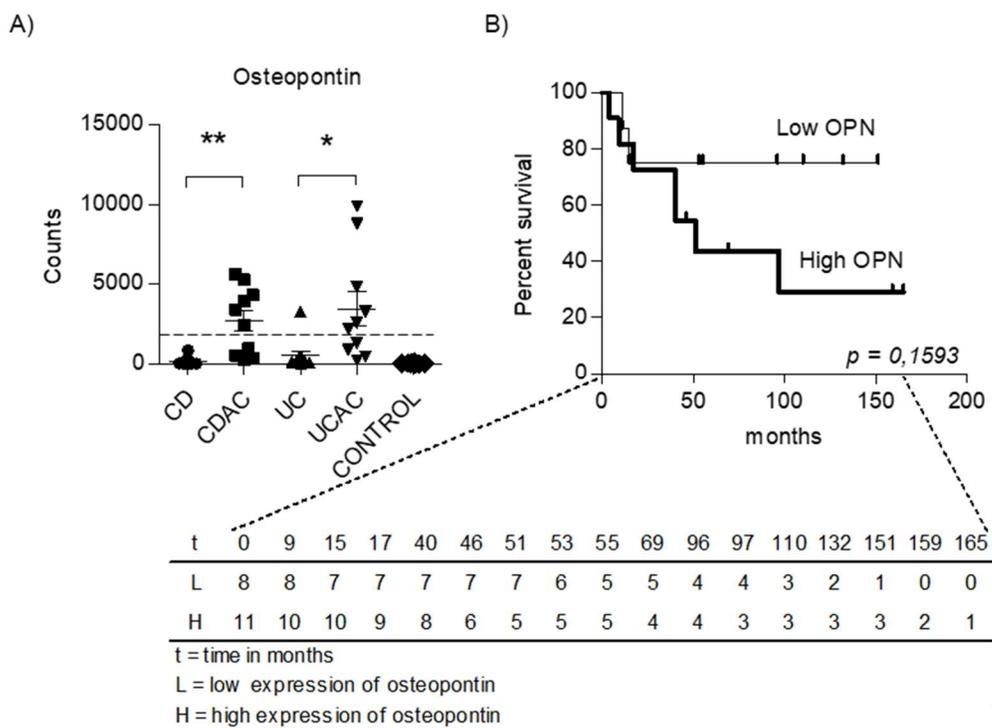


Figure 4.7: Osteopontin expression implicates in lower survival rates. Osteopontin counts resulting from the Nanostring analysis of patients (A). Survival curves of high vs low OPN expressing CAC patients (CDAC and UCAC together) shows a tendency for lower survival of high-OPN-expressing patients (B).

### Osteopontin is expressed by epithelial and stromal cells in patients

As in the Nanostring it is not possible to determine in which cellular compartment OPN protein is found in the colon of CAC patients, immunohistochemical analysis of osteopontin was performed to determine which cellular compartment was responsible for its production. OPN is expressed by both epithelial and stromal cells with a higher epithelial expression in CAC patients than in IBD or controls patients (Figure 4.8). Inserts show OPN nuclear localization in CAC patients, but not in the other groups.

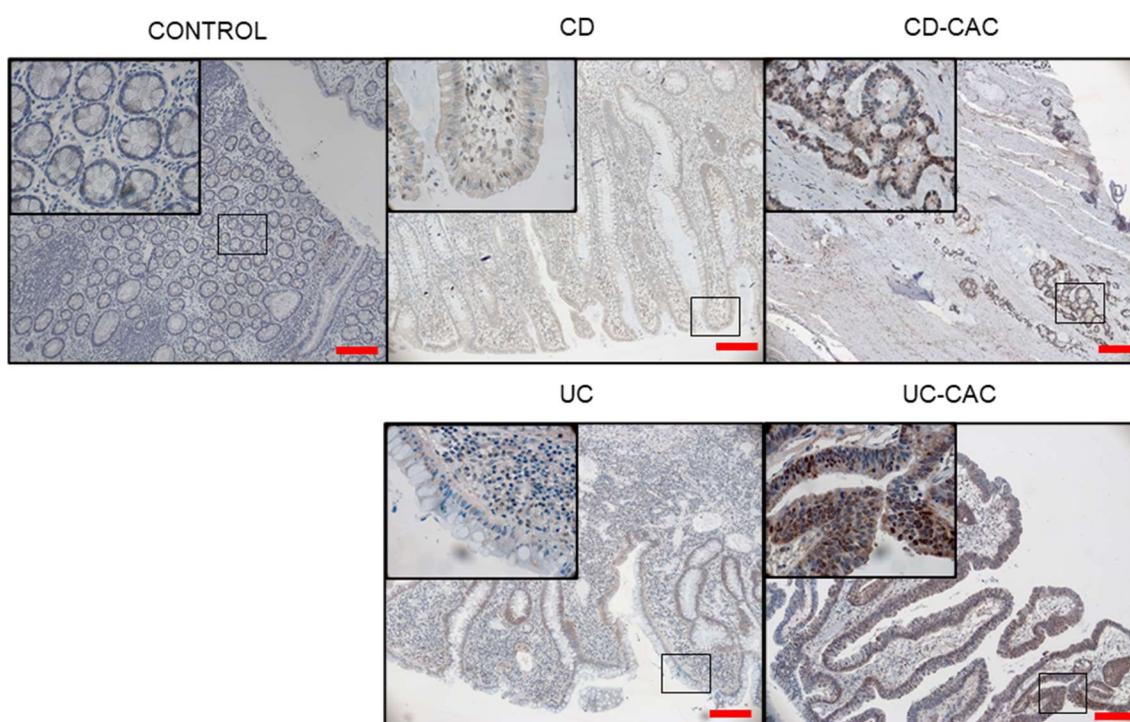


Figure 4.8: Immunohistochemical analysis of osteopontin in colonic tissue. Paraffin embedded samples were stained for osteopontin and analyzed in a confocal microscope LSM 780. Scale bars correspond to 200  $\mu$ m.

### Mechanistic network of OPN

Once we had determined the expression and location of OPN, we sought out to find possible signaling pathways involved in OPN activation. Nanostring data were submitted to Ingenuity

Pathway Analysis to determine which gene expression changes were possibly triggered by OPN in our samples. A mechanistic network of OPN as an activated upstream regulator predicted several putative down-stream targets for osteopontin. Among those, molecules such as  $\beta$ -catenin, SMAD3, STAT3, NF $\kappa$ B and FOS were cited (Figure 4.9).

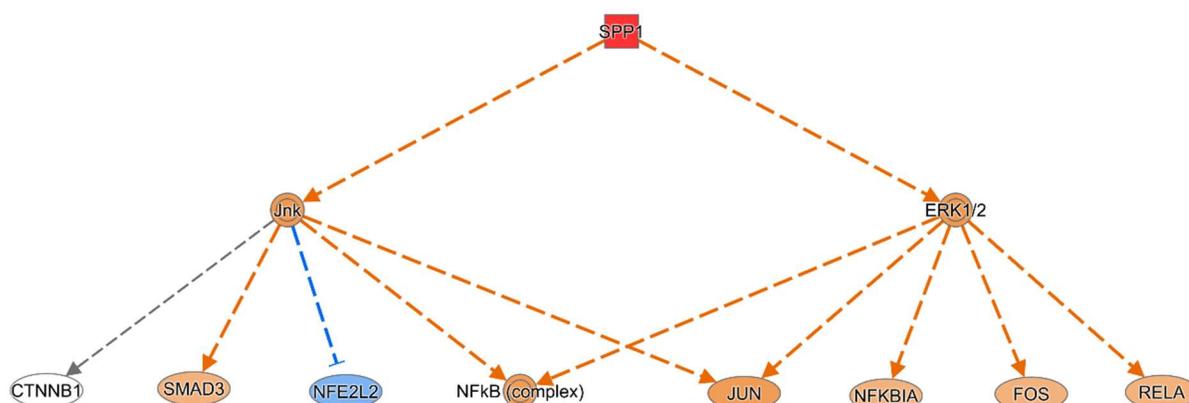


Figure 4.9. Osteopontin is predicted as an upstream regulator. The analysis of the ratios for the gene expression differences between the CAC vs IBD conditions yielded osteopontin as one of the activated putative upstream regulators in the dataset. Image acquired from Qiagen 2000-2020.

### Immunohistochemistry for OPN signaling

We then focused on searching for candidates to examine, including OPN putative targets and genes which were among the most upregulated genes, culminating in a list that contained: OPN, CD44, phospho-SMAD3, phospho-STAT3, cyclin D1, P-65,  $\beta$ -catenin, and GRHL2. Phospho-STAT3, and cyclin D1 were excluded for not presenting visible difference between the groups (data not shown) and GRHL2 and P-65 are not discussed here due to other research interests. Slides of patients' tissue were stained for the selected targets and a blinded analysis of staining intensity was performed. CD44 is one of OPN receptors and showed a tendency for higher expression in the epithelium of CAC patients (Figure 4.10).  $\beta$ -catenin was described as one possible downstream effector of osteopontin (Figure 4.9) and showed no significant different in membrane staining (data not shown), however, nuclear staining was found only in CAC patients

## Gene expression analysis to study celiac disease and colitis-associated cancer

(Figure 4.11A). SMAD3 was also described as downstream of osteopontin, but no differences between disease groups were observed in the analysis of phospho-SMAD3 (Figure 4.11B).

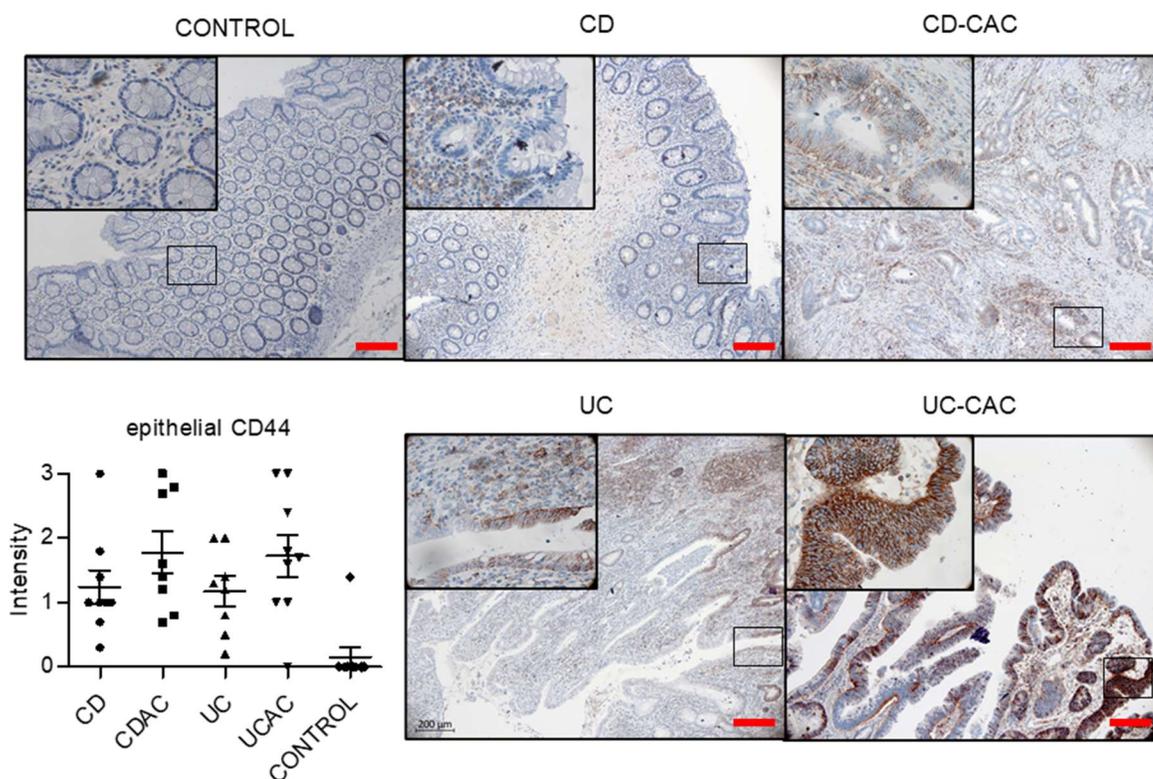


Figure 4.10. Immunohistochemistry of putative osteopontin targets. The osteopontin receptor CD44 was stained in the patients' slides by immunohistochemistry and a blinded analysis of staining intensity was performed. 5 different pictures in 10x magnification were taken using a confocal microscope LSM 780 and were then scored according to staining intensity from 0 to 3 by a blinded evaluator. Results of the blinded analysis are depicted in the dot plot chart and statistical analysis was performed using non-parametric Mann-Whitney test. Scale bars represent 200 μm.

## Gene expression analysis to study celiac disease and colitis-associated cancer

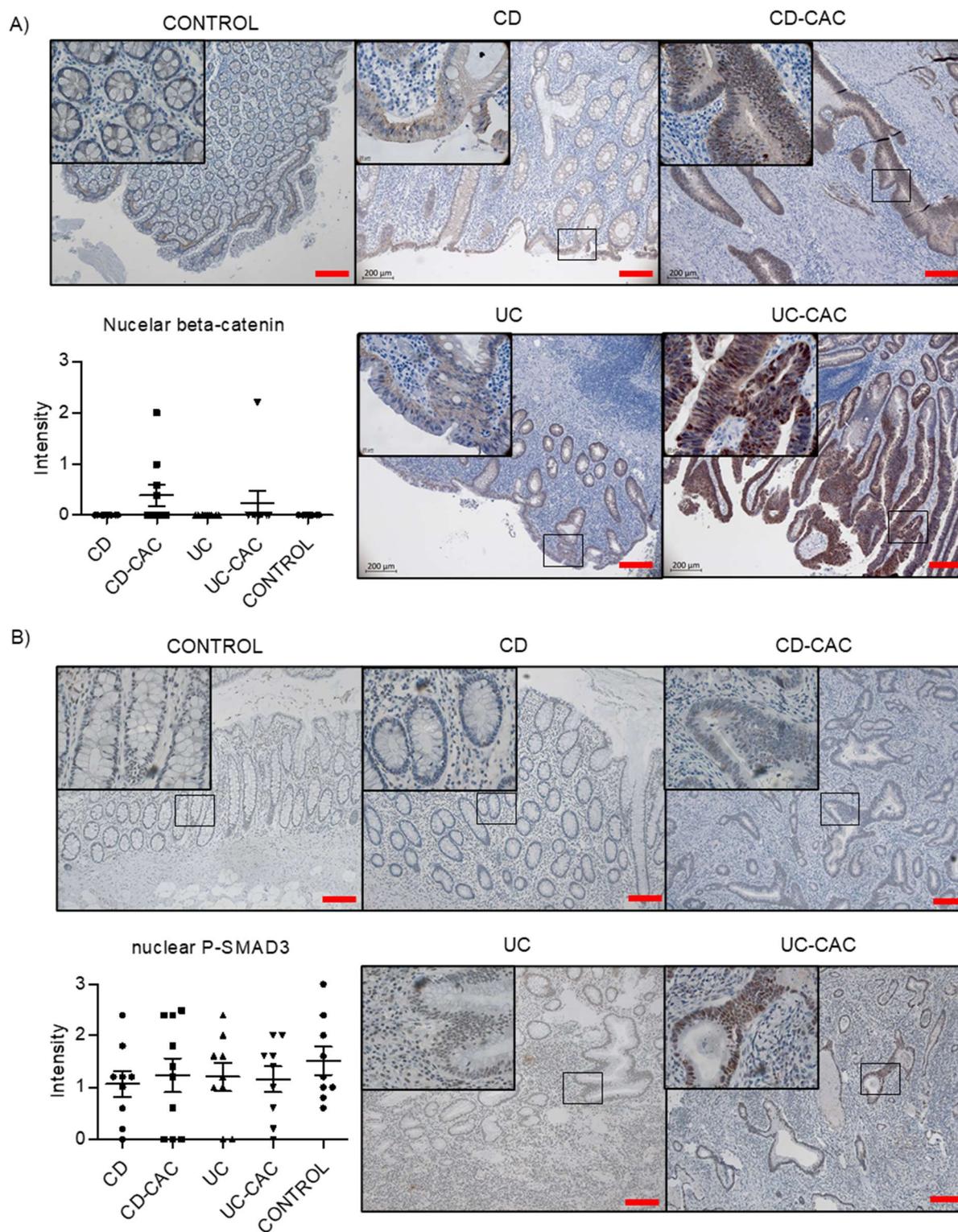


Figure 4.11. Immunohistochemistry of putative osteopontin targets. The signaling molecules  $\beta$ -catenin (A) and P-SMAD3 (B) were stained in the patients' slides by immunohistochemistry and a blinded analysis of staining intensity was performed. 5 different pictures in 10x magnification were taken using a confocal microscope LSM 780 and were then scored according to staining intensity from 0 to 3 by a blinded evaluator. Results of the blinded analysis are depicted in the dot plot chart and statistical analysis was performed using non-parametric Mann-Whitney test. Scale bars represent 200  $\mu$ m.

### **OPN as an EMT-inducing molecule**

One of the most discussed functions of OPN in tumorigenesis is the ability to induce epithelial-to-mesenchymal transition (EMT). In fact, two EMT-related genes: the transcription factor ZEB1 and the extracellular matrix protein fibronectin (FN1) were present in the Nanostring panel, were among the most upregulated genes (Tables 4.3 and 4.4) and shown to be differentially expressed in CDAC (Figure 4.12 A and B). To further investigate EMT in the patients, paraffin slides were stained for EpCAM, a known epithelial marker which should be decreased in EMT. The blinded analysis of the slides reported a tendency for lower EpCAM in the CAC tissue, supporting the hypothesis that EMT is present in the CAC patients (Figure 4.12C).

### **Protein analysis of intestinal cell lines exposed to osteopontin.**

With the intention of validating our data in a different study model, the effects of OPN on intestinal epithelial cells were analyzed using the human intestinal cell lines HT 29/B6 and T84. The activation of signaling pathways by OPN was analyzed by protein phosphorylation of signaling pathway molecules in both cell lines and showed that OPN did not induce STAT3 phosphorylation in neither of the tested cell lines (Figure 4.13). Similarly, for AKT, no changes in the phosphorylated compartment were triggered by OPN. In contrast, ERK1/2 were phosphorylated after exposure to osteopontin in both cell lines (Figure 4.13B), however, no statistical significance was found (Supplementary figures 3 and 4).

## Gene expression analysis to study celiac disease and colitis-associated cancer

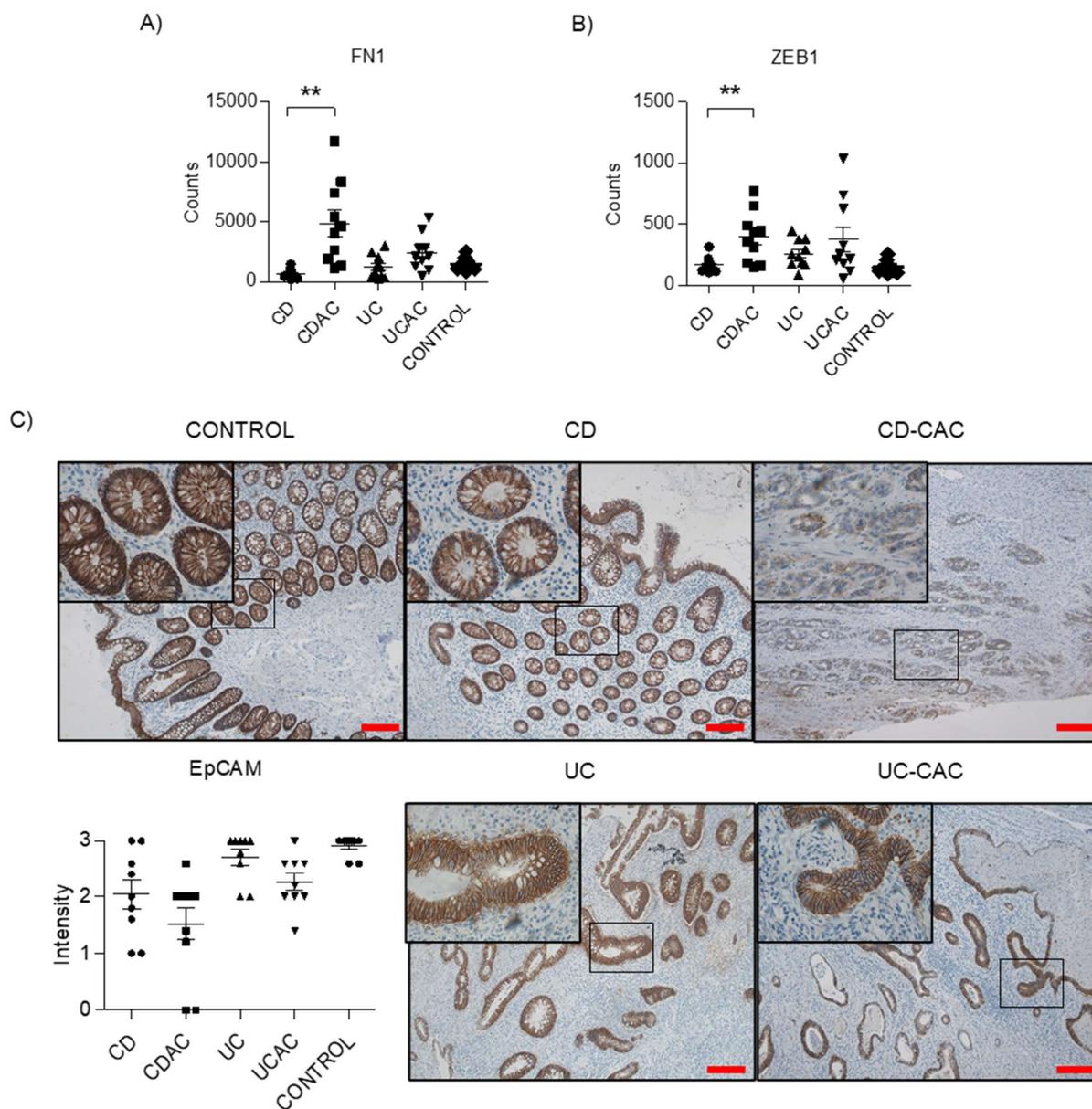


Figure 4.12. Epithelial to mesenchymal transition in colitis-associated cancer. The EMT markers FN1 (A) and ZEB1 (B) were evaluated in the Nanostring experiment. The epithelial marker EpCAM was analyzed by immunohistochemistry in patients' slides using 5 different 10x magnification pictures per patient slide and a blinded analysis of staining intensity was performed (C). Scale bars represent 200  $\mu$ m.

## Gene expression analysis to study celiac disease and colitis-associated cancer

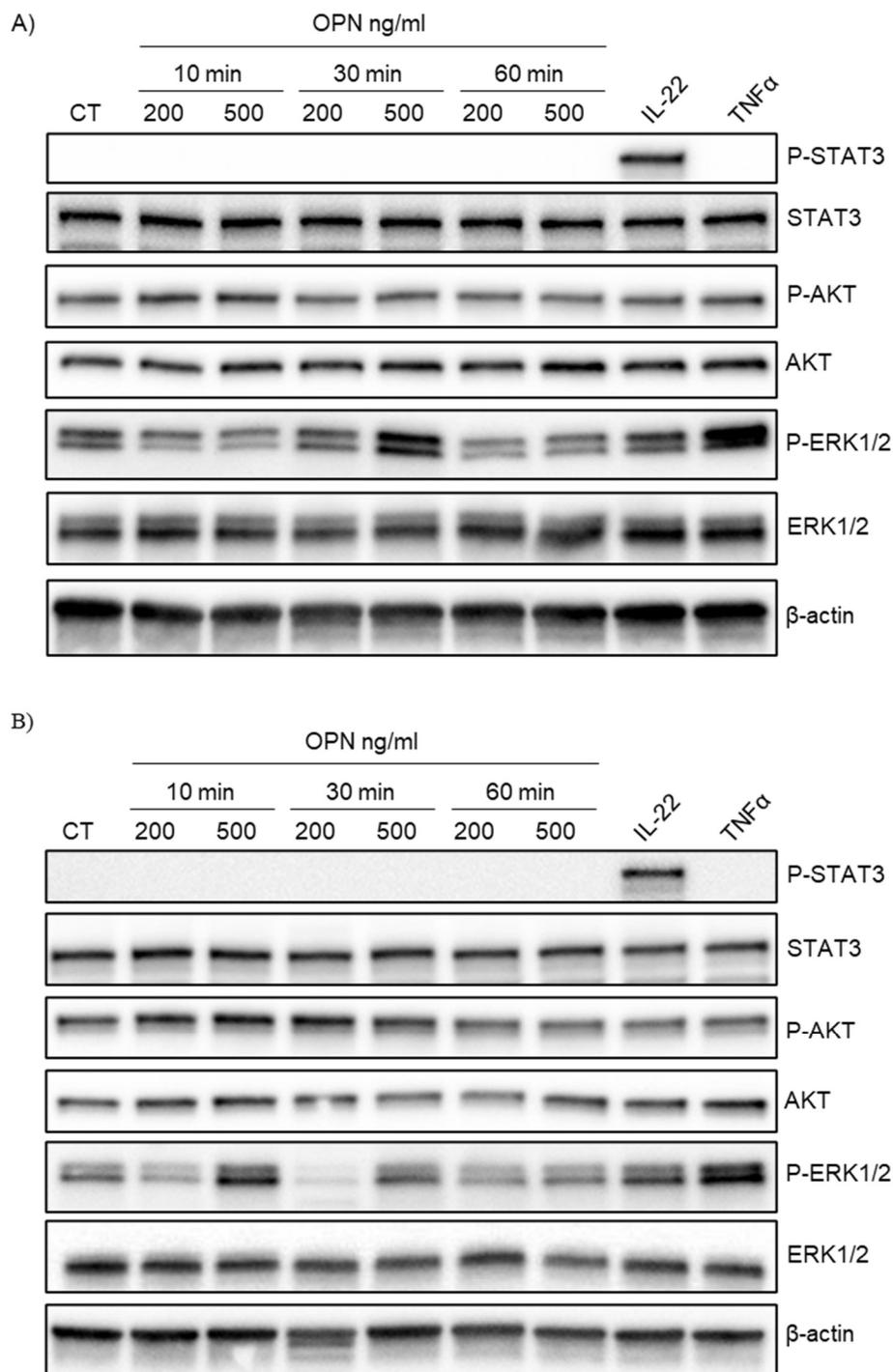


Figure 4.13. Western blot protein analysis of intestinal cell lines exposed to osteopontin. Intestinal cell lines T84 (A) and HT29 (B) were incubated with different concentrations of osteopontin for different periods of time before protein extraction and western blot analysis for total and phosphorylated STAT3, AKT and ERK1/2 and  $\beta$ -actin. Image representative of 3 independent experiments.

### Analysis of the phosphorylation of P56 after OPN treatment

Since NF $\kappa$ B was one of the signaling molecules that could be activated by OPN according to prediction (Figure 4.9), we sought out to investigate whether in intestinal cell lines OPN would induce P-65 phosphorylation and translocation to the nucleus. Caco-2 control clones EC B4 and EC D4 were exposed to OPN for 20 minutes and then were fixed and stained for phospho-P-65. OPN did not induce P-65 phosphorylation and translocation to the nucleus whereas there was a strong nuclear signal when cells were incubated with the positive control TNF $\alpha$ , showing that OPN did not activate P65 in cell line model (Figure 4.14).

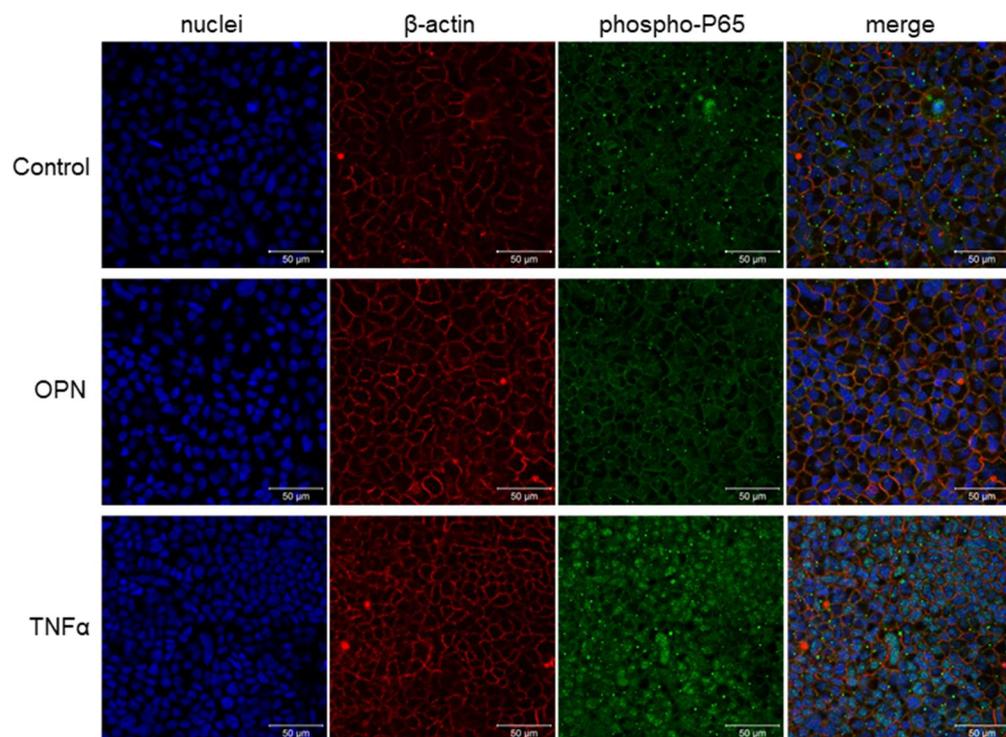


Figure 4.14. NF $\kappa$ B P-65 phosphorylation after osteopontin exposure. Caco-2 control clone EC B4 was exposed to 200 ng/ml of osteopontin for 20 minutes before being fixed and stained in order to assess nuclear translocation of phosphor-P-65 (green staining). Immunofluorescence was performed using antibodies against phosphorylated P-65 (green), phalloidin (red) and DAPI (blue). Images were taken in the confocal microscope LSM 780. Scale bars correspond to 50  $\mu$ m.

### mRNA analysis of cell lines after osteopontin incubation

To evaluate in the cell lines whether OPN could also trigger EMT, quantitative PCR analysis of the known EMT transcription factors SNAIL, SLUG and TWIST1 as well as metalloproteinase 7 - which is a later marker for EMT - was performed after exposure of the intestinal cell lines T84 and HT29/B6 to OPN. In HT29/B6 osteopontin failed to induce transcription of SNAIL and MMP7 (Figure 4.15). There was a slight induction in SLUG after 24h of osteopontin exposure and TWIST1 could not be estimated in this assay, for this cell line. In T84 cells there was no significant induction of any of the genes by osteopontin (Figure 4.16), which confronts the hypothesis that osteopontin promotes EMT in the analyzed intestinal cell lines.

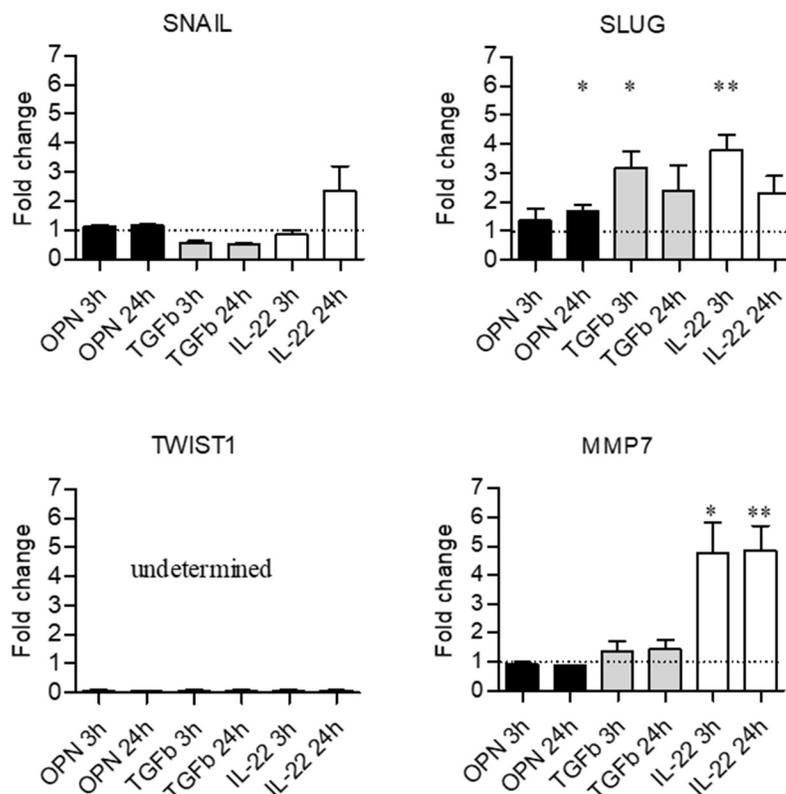


Figure 4.15. Quantitative PCR evaluation of HT29/B6 for EMT markers. HT29/B6 cells were exposed to 200 ng/ml of osteopontin for 3h or 24h and quantitative PCR was performed using TaqMan probes for SNAIL, SLUG, TWIST1 and MMP7. ATCB was used as endogenous control and expression changes were calculated using  $2^{-\Delta\Delta CT}$  method and the dashed line represents the expression of the control condition.

## Gene expression analysis to study celiac disease and colitis-associated cancer

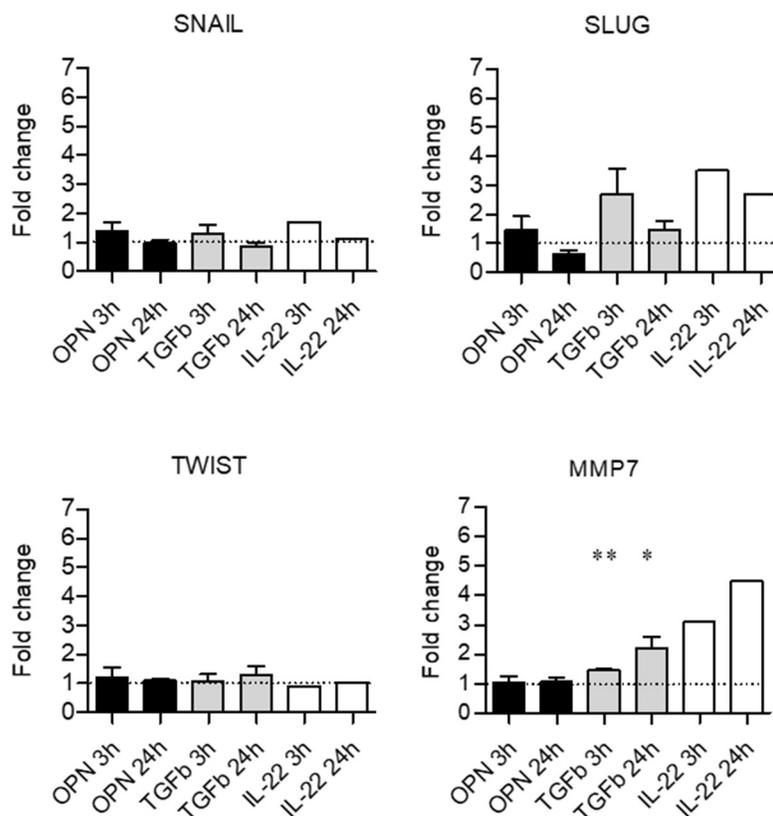


Figure 4.16. Quantitative PCR evaluation of T84 for EMT markers. T84 cells were exposed to osteopontin 200 ng/ml for 3h and 24h and then quantitative PCR was performed using TaqMan probes for SNAIL, SLUG, TWIST1 and MMP7. ACTB was used as endogenous control and expression changes were calculated using  $2^{-\Delta\Delta CT}$  method and the dashed line represents the expression of the control condition.

### RNA-Seq of cell lines

Not finding transcriptional regulation of EMT transcription factors presented a big setback for our primary hypothesis. In an attempt to understand how the two intestinal cell lines were responding to OPN a whole transcriptome sequencing RNA-Seq analysis was performed in the cell lines exposed to osteopontin. The RNA-Seq yielded clustering of samples by time-point and treatment condition as seen in the PCA plots (Figure 4.17 A and B). Greater differences are seen after 24h of treatment and not at 3h, and in HT29/B6 the samples exposed to osteopontin for 24h

## Gene expression analysis to study celiac disease and colitis-associated cancer

do not cluster together, indicating that our biological replicates presented variance among themselves (Figure 4.17A). In T84 we also do not see significant difference between conditions at 3h, but in 24h the controls and treated replicates cluster separate from each other (Figure 4.17B). In the volcano plot for HT29/B6 at 24h we observe that few genes are differentially expressed (Figure 4.17C) and the same is observed for T84 (Figure 4.17D).

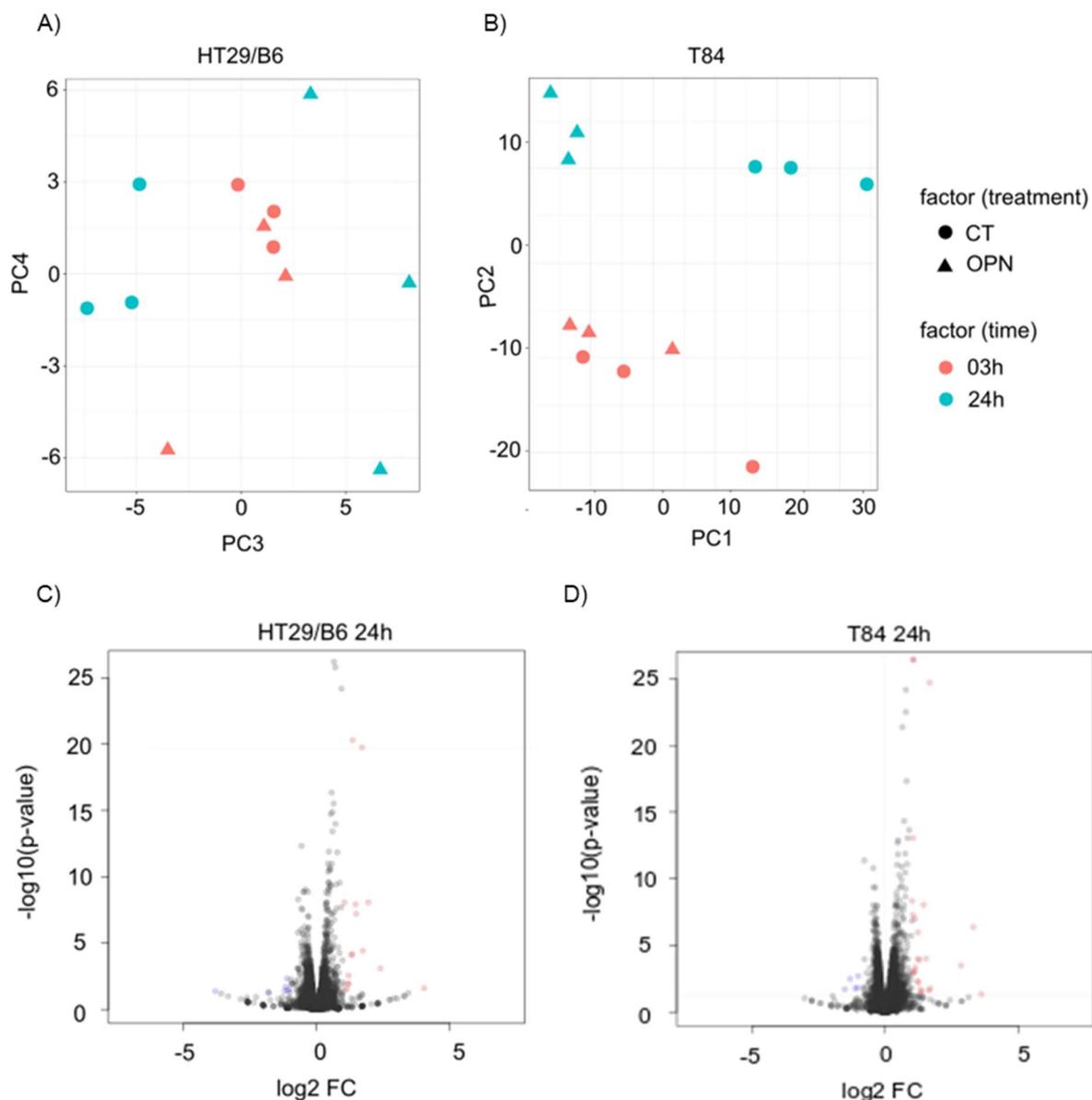


Figure 4.17. RNA-Seq analysis of intestinal cell lines exposed to osteopontin. Intestinal cell lines HT29/B6 and T84 were exposed to osteopontin for 3 and 24 hours and then analyzed via RNA-Seq in three biological

## Gene expression analysis to study celiac disease and colitis-associated cancer

replicates per condition. PCA plots for HT29/B6 (A) and T84 (B) show that greatest variances between untreated and treated cells are seen in the 24 hour-exposure. Volcano plots of the 24h time-point of HT29/B6 (C) and T84 (D) were made show that some genes are significantly regulated by osteopontin in both cell lines.

Among the most upregulated genes in HT29/B6 there are genes related to DNA damage such as STK33 and splicing SRSF12. On the other hand, the most downregulated genes include genes related to chromatin remodeling such as H4C4, BCL11A and H2BC10 (Table 4.5).

For T84, the list of upregulated genes includes genes without described function, but also MN1 which is a transcriptional regulator. The downregulated genes include cytoskeleton-related genes, such as ARHGAP22 and MARK1, and the extracellular matrix gene FBN2 (Table 4.6).

Table 4.5. Genes differentially expressed in HT29/B6

	<b>Gene</b>	<b>Fold change</b>	<b>p-value</b>
<b><i>Upregulated genes</i></b>	STIMATE	4.21	0.0031
	SRSF12	3.71	0.0234
	SH2D4B	3.53	0.0146
	MMP23B	3.48	0.0142
	NPY4R2	3.46	0.0057
	MT1G	3.44	0.0058
	CLDN19	3.40	0.0304
	BTLA	3.39	0.0061
	LINC00885	3.34	0.0341
	STK33	3.33	0.0173
<b><i>Downregulated genes</i></b>	H4C4	-5.00	0.0005
	BARHL1	-4.81	0.0012
	TDRD9	-4.32	0.0027
	RRH	-4.00	0.0073
	BCL11A	-3.58	0.0066
	MIR3177	-3.35	0.0379
	CELP	-3.32	0.0138
	SEMA3D	-3.31	0.0227
	TERB1	-3.21	0.0325
	H2BC10	-3.03	0.0489

Table4.6. Genes differentially expressed in T84

	<b>Gene</b>	<b>Fold change</b>	<b>p-value</b>
<b><i>Upregulated genes</i></b>	LINC01585	4.45	0.0023
	LOC100505501	3.60	0.0077
	SNORD14E	3.49	0.0137
	MN1	3.49	0.0108
	ODF3	3.33	0.0009
	GFPT2	3.31	0.0142
	RN7SL2	3.30	<0.0001
	FAM87A	3.28	0.0220
	ABCA6	3.27	0.0324
	MT1M	3.20	0.0337
<b><i>Downregulated genes</i></b>	FBN2	-3.98	0.0046
	DAW1	-3.88	0.0039
	SPARCL1	-3.45	0.0156
	FAM95C	-3.44	0.0124
	ARHGAP22	-3.29	0.0092
	PCBP3	-3.27	0.0221
	LINC01771	-3.23	0.0161
	EID2B	-3.18	0.0128
	LOC285095	-3.16	0.0427
	MARK1	-3.05	0.0417

### Gene enrichment sets of the RNA-Seq analysis

Enrichment analysis was performed using the tmod database. HT29/B6 presented very few enriched sets, involved in mitochondrial respiration and DNA repair (Table 4.7). ROC curves of the gene sets enriched in HT29/B6 show the genes which were regulated in the set distributed according to p-value and colored according to the direction of regulation. Gene sets LI.M219 and LI.M231 present highly significant upregulated genes (Figure 4.18).

## Gene expression analysis to study celiac disease and colitis-associated cancer

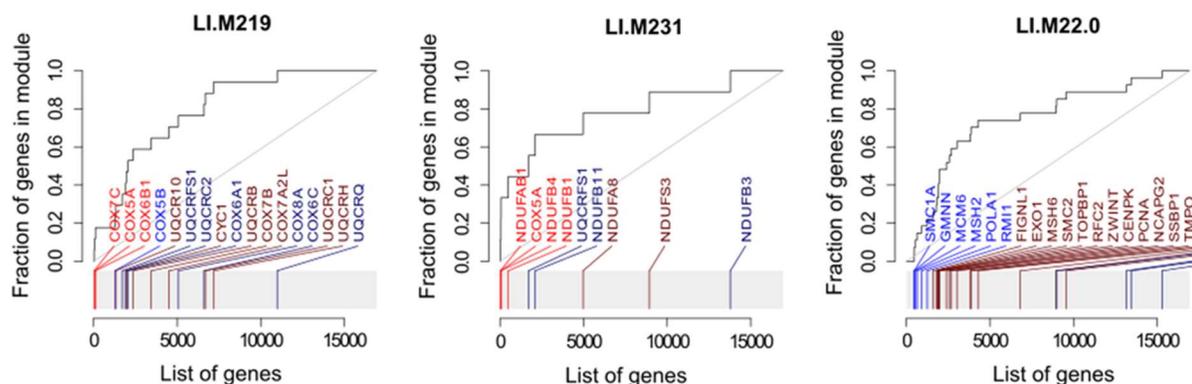


Figure 4.18. ROC curves of the enriched gene sets in HT29/B6 cells. Gene sets with area under the curve  $> 0,7$  and adjusted p-value  $< 0,01$  were represented as ROC curves displaying the fraction of regulated genes according to p-value versus the list of all genes in the set. Bright colors are strongly significant and dark colors mean moderate significance. Red stands for up- and blue, down-regulation

Table 4.7. Gene enrichment analysis of HT29/B6 cells exposed to osteopontin for 24 hours.

ID	Title	N1	AUC	Adj. P-value	Significant genes
LI.M219	Respiratory electron transport chain (mitochondrion)	17	0.80	0.0008	COX5A; COX5B; COX6B1; COX7C
LI.M231	Respiratory electron transport chain (mitochondrion)	9	0.79	0.0007	NDUFAB1; COX5A; NDUFB4; NDUFB1
LI.M22.0	Mismatch repair (I)	27	0.75	0.0079	SMC1A; POLA1; MSH2; GMNN; RMI1; MCM6

N1, number of genes in the set

AUC, Area under the curve

Gene sets were filtered for AUC  $> 0.7$  and adj. p-value  $< 0.01$

In T84 cells many gene sets were enriched involving cellular transport, cell-cycle, phosphatidylinositol signaling, DNA repair, splicing, mitochondrial respiration and protein synthesis (Table 4.8). The ROC curves for the enriched gene sets in T84 show that most of the genes are downregulated, especially for gene sets LI.M147 and LI.M144 and reinforcing that both cell lines respond in a completely different manner to the OPN stimulus (Figure 4.19).

Table 4.8. Gene enrichment analysis of T84 cells exposed to osteopontin for 24 hours.

ID	Title	N1	AUC	Adj. P-value	Significant genes
LI.M147	Intracellular transport	17	0.87	0.0002	SIRT1; EXOC1; VPS4B; NUP107; CLINT1; ZFYVE16; SEC63; ZFAND6; PIK3C2A; PPP1R12A; RAD21; G3BP2; SRP9
LI.M144	Cell cycle. ATP binding	15	0.84	0.0002	RBM7; UBA3; HDAC2; TLK1; RAD21; COPS5; PPP1R12A; VPS4B; CCNC; CUL5; PSMC6
LI.M101	Phosphatidylinositol signaling system	13	0.83	0.0052	AGL; PIK3C2A; DEK; PPP1R12A; PIK3C3; MICU2; SLC35A1
LI.M169	Mitosis (TF motif CCAATNNSNNGCG)	16	0.82	0.0042	SMC1A; TMPO; ORC4; CASP8AP2; CETN3; UPF3B; ORC3; ACTR6
LI.M22.0	Mismatch repair (I)	27	0.81	0.0019	SMC1A; RFC4; MSH2; TMPO; MSH6; RFC2; GMNN; RMI1; PRIM1; CENPK; FIGNL1; MCM6; SSBP1; TOPBP1; SMC2
LI.M250	Spliceosome	12	0.77	0.0052	SNRPE; LSM3; RBMX; SNRPA; SNRPD2
LI.M226	Proteasome	12	0.76	0.0067	PSMD14; PSMA3; PSMC6; PSMA4; PSMC2; POLR2K
LI.M219	Respiratory electron transport chain (mitochondrion)	17	0.74	0.0052	COX5A; COX5B; UQCR10; COX7B; UQCRB; UQCRH; COX7A2L
DC.M4.3	Protein Synthesis	37	0.74	<0.0001	RPL6; ZFAND1; RPL36; RPS3; RPS14; HSF2; EEF1B2; RPL7A; SNRPD2; ELP2; APRT; MPHOSPH10; RPL9; RPA1; RPL5; RPL12; MCCC1; DDX18; RPS20

N1, number of genes in the set

AUC, Area under the curve

Gene sets were filtered for AUC >0.7 and adj. p-value <0.01

## Gene expression analysis to study celiac disease and colitis-associated cancer

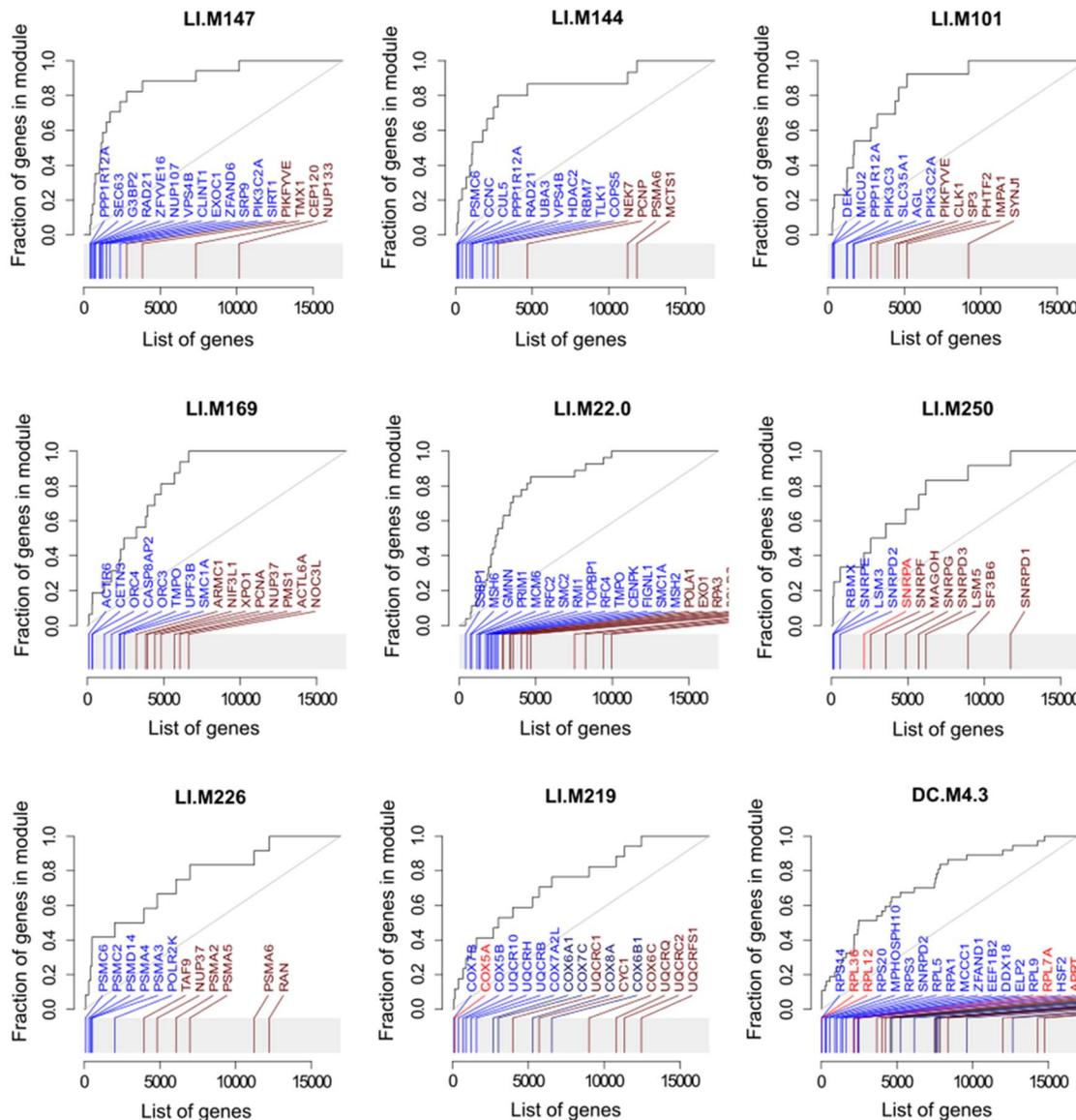


Figure 4.19. ROC curves of the enriched gene sets in T84 cells. Gene sets with area under the curve  $> 0,7$  and adjusted p-value  $< 0,01$  were represented as ROC curves displaying the fraction of regulated genes according to p-value versus the list of all genes in the set. Bright colors are strongly significant and dark colors mean moderate significance. Red stands for up- and blue, down-regulation.

## Gene expression analysis to study celiac disease and colitis-associated cancer

A closer look at the genes described as significant genes in the mitochondrion respiratory chain gene sets, we see that respiratory complexes I, III and IV are represented (Table 4.9). There is a tendency for Complex I upregulation in both cell lines, whereas complex III genes are downregulated especially in T84. Complex VI seems upregulated in HT29/B6, but there is no clear trend for T84 once it shows up- and downregulation of different genes from this complex (Table 4.9).

Table 4.9. Mitochondrial respiratory chain gene sets

Gene name	Full name	Part of respiratory chain complex #	Cell line	Up- or downregulat
COX5A	Cytochrome c oxidase subunit 5A	Complex IV	HT29/B6 and T84	up
COX5B	Cytochrome c oxidase subunit 5B	Complex IV	HT29/B6 and T84	down
COX6B1	Cytochrome c oxidase subunit 6B1	Complex IV	HT29/B6	up
COX7A2L	Cytochrome c oxidase subunit 7A-related	Complex IV	T84	down
COX7B	Cytochrome c oxidase subunit 7B	Complex IV	T84	down
COX7C	Cytochrome c oxidase subunit 7C	Complex IV	HT29/B6 and T84	up
NDUFA B1	Acyl carrier protein, alternative NADH-	Complex I	HT29/B6 and T84	up
NDUFB 1	NADH dehydrogenase [ubiquinone] 1 beta	Complex I	T84	up
NDUFB 4	NADH dehydrogenase [ubiquinone] 1 beta	Complex I	HT29/B6 and T84	up
UQCR1 0	Cytochrome b-c1 complex subunit 9,	Complex III	HT29/B6 and T84	down
UQCRB	Cytochrome b-c1 complex subunit 7,	Complex III	T84	down
UQCRH	Cytochrome b-c1 complex subunit 6,	Complex III	T84	down

## 5 DISCUSSION

### TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE

Barrier defect in celiac disease is a well-known phenomenon and its roots have been studied throughout the years. It is undeniable that the inflammatory process greatly affects the epithelial barrier function through the secretion of cytokines and the induction of apoptosis (143,145,146). However, the evidence for a genetic cause has been long envisioned and recently proven with the identification of susceptibility loci in genes related to cell-cell adhesion (248). Among the indicated genes, LPP, C1orf106 and PTPRK were functionally proven to play a role in cell adhesion (154,160,249).

In the light of those findings, we aimed at determining the role of LPP and C1orf106 genes in cell lines and whether their depletion would affect barrier function of intestinal cell line Caco-2. Both genes were knocked-out via CRISPR-Cas9 editing and the clones were established. In TEER measurement, both genes, but especially C1orf106 showed reduced electrical resistance when compared to controls. This is in accordance with what was found in C1orf106 KO models (159,160). Furthermore, after the thorough evaluation of protein content of claudins, we found that Claudin-3 was significantly upregulated in LPP clones and claudin-1 and -7 showed a tendency for being increased, but it was not significant. Similarly, Claudin-7 and -8 showed a tendency for downregulation in C1orf106 KO clones. Claudin protein content evaluation in LPP or C1orf106 KO models has not been performed by previous studies; it was reported, though, that LPP KO MDCK cells presented normal levels of ZO-1, ZO-2, occludin, catenins, but reduced E-cad (154). In C1orf106 KO cells ZO-1 was displaced from the membrane (159). To conclude this part of our evaluations, a calcium switch assay showed that both KO clones are not able to re-assemble TJ in a way to display a similar TEER as the controls. Impaired TJ re-assembly was observed in LPP clones before (154), but not C1orf106.

## Gene expression analysis to study celiac disease and colitis-associated cancer

In order to achieve a broader impression on the impact of LPP and C1orf106 depletion, we performed an WTS RNA-Seq with the clones in normal conditions and exposed to cytokines (IL-15, IL-22, IFN $\gamma$  and TNF $\alpha$ ). First results show that controls, LPP and C1orf106 KO clones present significant variance to cluster separate from each other in a PCA analysis (Figure 5.1). Indeed, we believe a gene-enrichment analysis will clarify the functions altered by the lack of those two genes.

LPP is a protein with multiple functions, being found both in focal adhesions where it interacts with VASP and  $\alpha$ -actinin; and in the nucleus, where it has the ability to act as a transcription factor (153). LPP interacts with  $\alpha$ -actinin, which is a cross-linking actin protein found in focal adhesions (250) and was reported to be necessary for TGF $\beta$ -induced migration in ErbB2-positive breast cancer cells (251). LPP-deficient present deficient migration (252). Which opposites its role in E-cad-dependent cell adhesions since it is related to EMT induction in cancer cells (154). Considering those recent findings, LPP-KO cells could present a stronger epithelial phenotype than LPP-containing cells, which would be in accordance with our KO clones B5 and B1 presenting increased claudin-3 and not showing significance TEER decrease in normal conditions. On the hand, the fact that LPP is important for E-cad adhesions, could explain the apparent delay in TEER recovery after the calcium switch.

C1orf106 has an indirect role in the negative regulation of E-cad internalization. It was shown to induce degradation of cytohesins-1 and -2, which inhibits their activation of the GTPase ARF6 (159,160), which in turn induces E-cad membrane displacement and degradation (161). Moreover, ARF6 has multiple roles in tumorigenesis such as inducing migration, invasion in proliferation in cancer cells (253). ARF6 inhibition by C1orf106 could be protective in epithelial cells against tumorigenesis.

## Gene expression analysis to study celiac disease and colitis-associated cancer

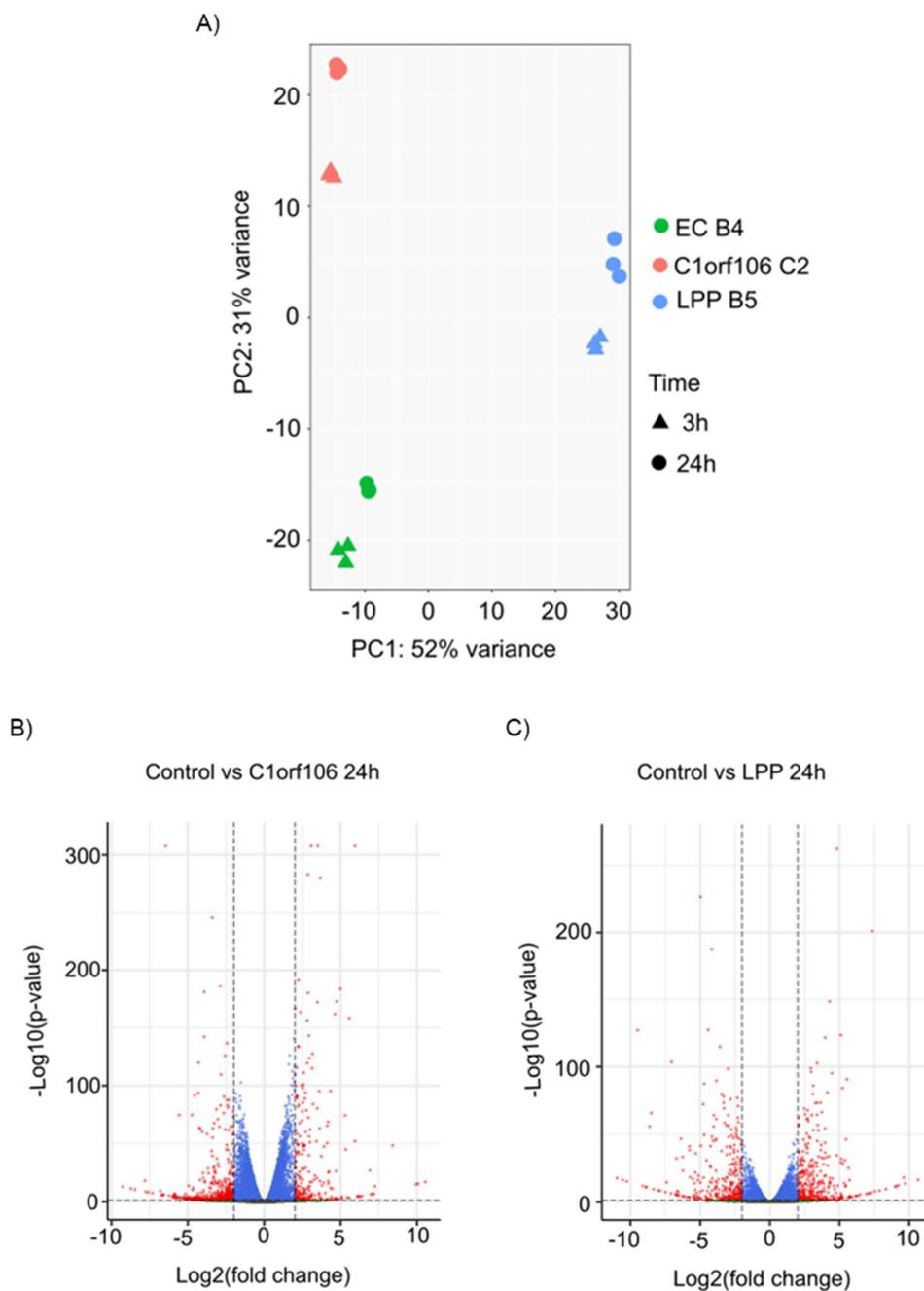


Figure 5.1. RNA-Seq analysis of Caco-2 knock-out clones. Total RNA was extracted from Caco-2 clones EC B4, LPP B5 and C1orf106 C2 and a whole transcriptome sequence RNA-Seq analysis was performed. Variance between expression profile made the clones cluster separate from each other and in both analyzed time-point as shown in the Principal Component Analysis plot (A). Volcanos plots for EC B4 vs C1orf106 C2 (B) and LPP B5 (C) were made and showed that many genes were differentially regulated between clones.

The depletion of C1orf106 led to decrease in TEER, increase in the permeability for lucifer-yellow (160). Moreover, C1orf106 depletion was shown to render mice more susceptible to barrier impairment, shown by diarrhea with increased fecal water content and FITC-dextran leakage, after TNF $\alpha$  injection (159). Considering this information, C1orf106 KO could be especially important in a CeD context, for epithelial cells are exposed to a milieu of inflammatory cytokines during disease activity.

After characterization of the basic barrier function in the caco-2 clones, we sought out to investigate the expression of LPP and C1orf106 in duodenal samples coming from celiac patients.

A significant decrease in expression level of both genes was found in celiac patients when compared to healthy individuals (248). However, in our Western Blot analysis, we observed a rather non-significant scatter in all our patient groups. Since we so far have a limited number of patients enrolled in the study, it could be that with the increase in the number of patients such differences will become clearer. Furthermore, we have measured TEER of those patients in order to correlate with protein and RNA findings (data not shown), however, since in active and RCD patients there can be a significant villus atrophy and crypt hyperplasia, further correction of TEER values with epithelial surface needs to be made. Finally, we also collected frozen biopsies and aim at evaluating expression level of target genes, which we expect to compare to the data acquired from the KO clones and report on specific gene set enrichment or altered pathways.

## **ROLE OF OSTEOPONTIN IN CAC**

For the second part of this thesis, we analyzed colitis-associated cancer, which has been reported to have a progressive decrease in excess risk for IBD patients. This decrease is attributed to better implementation of surveillance strategies and better control of disease activity by the new treatments. On the other hand, the decrease in risk could also result from the aging of the cohorts (190). FFPE material was obtained and a Nanostring was performed. The choice

## Gene expression analysis to study celiac disease and colitis-associated cancer

for the Nanostring rather than other RNA-based screening method was due to it showing more stability in the analysis of fragmented RNA samples than the amplification-based techniques (254). The human immunology panel was selected for CAC having an IBD background and our interest in investigating how the inflammatory process contributes to its development. The most up- and downregulated genes show an idea of general changes in the immune cell compartment and inflammation status. For example: the finding that IBD-related genes A1009 and A1008 (163) being downregulated in CAC.

Being OPN the most upregulated gene in both comparisons, which are completely independent from each other, led us to think it could be involved in a mechanisms of CAC tumorigenesis shared by UC and CD-colitis and instigated us to investigate it further. Indeed, higher OPN expression is in accordance to literature in CRC and various other types of solid cancers (230). We pursued a correlation between expression level and survival rate but, likely due to our small number of patients per group, could not find significance. OPN levels in tissue and peripheral blood have been correlated with survival rates and tumor stage in various types of cancer, including CRC (233). Such an analysis had not been done before, specifically in CAC. We believe that with a larger cohort, we can establish a significance in the survival rate of CAC patients based on their OPN expression levels.

When we examined the histological slides stained for OPN we observed its localization not only in the epithelium, but also in stromal cells. Indeed, OPN is produced and secreted by many immune cells such as, DCs, NKs, T and B cells (230) as well as various epithelial cells, including intestinal epithelial cells (197). In some cancer patients (without significant difference between disease groups) nuclear OPN staining was present, which is reported in the literature as a negative prognostic indicator for survival (255).

Next, to determine possible pathways that could be regulated by OPN, we searched the Qiagen knowledge base to find possible targets, which showed us several molecules and signaling pathways. We then sought out to evaluate candidates, including CD44, beta-catenin and SMAD3.

CD44 showed a tendency of increase in epithelial cells of CDAC and UCAC patients. In carcinomas, it is regarded as a marker for cancer stem cells (256). Moreover, OPN-induced migration of macrophages was seen to be dependent of CD44 expression in mice (223). In hepatocellular carcinoma cells, the OPN promoted cell proliferation through CD44 (257).

For  $\beta$ -catenin, the most important change was not in increased signal, but rather nuclear localization being found only in a few CAC patients. In CRC patient samples, there was a strong correlation between OPN and nuclear  $\beta$ -catenin IHC staining. Moreover, co-expression of these two proteins correlated with lymph node metastasis, tumor invasion and TNM stage (258). In prostate cancer cells, OPN induces  $\beta$ -catenin nuclear translocation through activation of AKT and resulting in expression of MMP7 and CD44 (259).

SMAD3 is necessary for EMT induction in lens epithelium during the development of posterior capsule opacification in the eye (260). OPN was seen co-expressed with phosphorylated SMAD3 in the calcification process blood vessels in cerebral amyloid angiopathy (261).

Since one of the most studied functions of OPN in tumors is the ability to promote EMT, we investigated whether our patients' samples present changes related to the EMT process. Indeed, amidst the Nanostring data we found the transcription factor ZEB1 and the extracellular matrix protein fibronectin (FN1). ZEB1 is one of the EMT core genes and directly represses E-cadherin and induces expression of vimentin (238). Fibronectin expression is increased as a consequence of EMT (236). In our analysis, both genes were upregulated in CDAC patients compared to CD, in UCAC there was a tendency for upregulation, however, without statistical significance. Those

corroborate with the hypothesis that OPN contributes for EMT in the CAC patients. Therefore, we sought out to study it mechanistically in two human intestinal cell lines HT29/B6 and T84.

When those cells were exposed to OPN there was phosphorylation of ERK1/2, but not of AKT or STAT3. In addition, phosphorylation and translocation of NF $\kappa$ B P-65 was not observed. We then evaluated transcriptional regulation of the EMT-core genes SNAI1, SNAI2 and TWIST1 as well as the MMP7; Nonetheless, we could not find significant changes in the expression of those genes. Only SLUG showed a significant up-regulation in HT29 after 24h of OPN incubation, however, it was not sufficient to trigger the EMT process.

OPN promoted proliferation and invasion in intestinal HCT116 cells through activation of PI3K/AKT (262) and also was shown to induce TWIST expression in hepatocellular carcinoma cells through the same pathway, promoting expression of MMP2 and uPA (urokinase-type plasminogen activator). In addition, OPN knock-down decreased the expression of N-cadherin and increased E-cadherin protein content (241). In ovarian cancer cells, OPN induced proliferation, migration, and invasion with expression of vimentin and N-cadherin through both AKT and ERK1/2 pathways (263). In breast cancer cells, OPN induced migration and uPA expression by phosphorylation and nuclear translocation of NF $\kappa$ B P-65 through PI3K/AKT (222). Finally, again in breast cancer cells, OPN induced migration and inhibition of apoptosis, as well as expression of Bcl-2 and Cyclin D1 through activation of JAK2/STAT3 (264).

This was also proven in colorectal cancer cell lines HT29 and COLO205. OPN induces proliferation, migration and invasion of those cell lines, accompanied of increased protein content of  $\beta$ -catenin, SNAIL, MMP2, 3 and 9 while reducing E-cad (233,262). OPN knock-down impaired migration, cell cycle progression, and increased apoptosis rate. Also decreased vimentin expression and increased E-cad (265).

All those data indicate OPN could have activated any of the pathways we investigated, inducing EMT, however, despite a slight activation of ERK1/2, we did not observe the same in our cell lines. We hypothesize that the lack of signaling activation and EMT induction could be due to the use of full-length recombinant OPN, instead of a different isoform, such as OPN-c and OPN-b or the thrombin cleaved OPN.

One other thing that could have influenced our results was the substrate onto which the cells were seeded or the confluency of the culture at the start of the experiments. Those possibilities were investigated once we seeded cells in both cell inserts and plates. Cell inserts on the one hand provide for apical and basolateral stimulation, on the other hand does not allow observation of the culture growth and for that is used at a stage in which the cells should be confluent. We performed OPN exposure of the cells, followed by quantitative PCR analysis of EMT genes, but saw no induction of expression for those genes (data not shown). Our second approach was to seed the cells in plates and start the experiment at around 50% confluency, which are the results presented in this document. Nevertheless, both approaches rendered similar results without significant regulation of the EMT transcription factors, indicating the seeding substrate was not the reason why we did not find EMT in the cell lines.

Finally, we performed an RNA-Seq analysis of the cell lines after exposure to OPN. The analysis showed that longer exposure times were necessary to achieve significant variance in between treated and untreated cells. Even though not many genes were significantly regulated, gene set enrichment showed regulation of respiratory chain, protein synthesis, splicing and transcription, especially in T84 cells. Further analysis of the genes belonging to the respiratory chain sets, showed that there is a tendency for upregulation of Complex I in both cell lines, whereas complex III is downregulated only in T84 and complex IV is upregulated in HT29/B6 and has no specific trend in T84.

In the literature associating OPN to mitochondrion respiration, OPN inhibits the expression on cytochrome c oxidase in murine macrophages (266). Moreover, OPN, via CD44 binding, induces apoptosis of rat cardiomyocytes through the mitochondrial death pathway, with the activation of JNKs and induction of expression of Bax and cytochrome C and ER stress pathway with increased expression of Gadd153 (267). Further investigation revealed that OPN induces ROS production by increasing expression of NOX-4 (NADPH oxidase isoform 4) and decreasing expression of SOD-2 (superoxide dismutase-2). OPN decreased mitochondrial transmembrane potential and induced mitochondrial remodeling with fragmentation of cristae. As a conclusion, the effects of OPN in mitochondrial remodeling and apoptosis were associated to increased expression of BIK (268). Furthermore, in a murine model of heart failure with preserved ejection fraction (HFpEF), OPN deletion improved diastolic function and reduced myocardial fibrosis. The HFpEF mice presented elevated oxidase stress, including significant reduction in the levels of mitochondrial electron transport chain complexes I, II and IV and swollen mitochondria with disorganized cristae. OPN effects coincide with decrease in OGDHL protein. In contrast, OGDHL overexpression improved mitochondrion function of cardiomyocytes (269).

These findings corroborate with the downregulation of mitochondrial respiratory chain complexes III and IV seen especially in T84 cells exposed to OPN, suggesting a new mechanism for OPN in tumorigenesis of CAC.

The fact that HT29/B6 and T84 are cell lines established from colorectal carcinoma patients indicate that those cell lines are not the best model for the study of CAC. Indeed, we believe we would have a more reliable response had we used a better model for that. Mice CAC models include Azoxymethane (AOM)/Dextran sodium sulfate (DSS) model, in which the mice are injected with a carcinogen followed by colitis induction by the ingestion of a heparin-like polysaccharide (DSS) dissolved in water has been largely used and recapitulates the key aspects of CAC tumorigenesis (270). A second model is the AOM/IL-10<sup>-/-</sup> mice. IL-10 KO mice are

inflammation-susceptible and an established model of IBD. The intraperitoneal injection of AOM triggers the tumorigenesis of CAC in those mice (271). Moreover, a recent established model combines Mucin-2 mutation and APC<sup>Min/+</sup>. This model results in an inflammatory background with genetic predisposition to small intestinal polyposis. Mice showed dysplastic lesions from 5 weeks along the entire colon (272). In addition, when OPN<sup>-/-</sup> mice were treated with DSS, there was an aggravation of the acute experimental colitis, whereas OPN depletion was protective in chronic colitis (273).

When it comes to the study of human diseases, mice models have been extremely important over the decades, however, they do not represent a substitution of the human cells and tissues. In this regard, organoids have been shown to be flexible human models that overcome many limitations of immortalized cell lines, such as forming a 3D structure that reproduces tissue architecture and homeostasis and can be derived from virtually any tissue for a long-term propagation. They have been reported to be reproducible and can be used as patient-specific *in vitro* models (274). Thus, the continuation of this project will focus on studying the mechanisms of OPN in CAC progression in organoids established from IBD and CAC patients, which will be more reliable models than immortalized cell lines coming from CRC and which reproduce the diversity of the intestinal microenvironment.

## **6 CONCLUSIONS**

### **TRACKING A PRIMARY BARRIER DYSFUNCTION IN CELIAC DISEASE**

Our results show that the knock-out of LPP and C1orf106 genes result in alterations in TJ and TJ function. C1orf106 clones presented lower TEER values, whereas LPP KO clones presented changes in tight junctional protein content, especially the up-regulation of Claudin-3. After calcium depletion and replacement, KO clones presented lower TEER levels than controls, indicating their barrier function is impaired in comparison to the controls. The protein analysis of LPP and C1orf106 in patients was inconclusive, with no significant difference found between disease groups. Further investigation is needed in order to define the contribution of both proteins in barrier impairment in CeD.

### **THE ROLE OF OPN IN THE PATHOGENESIS OF CAC**

The RNA analysis from patients' samples retrieved many findings, being OPN the most upregulated gene in the analyses between CDAC vs CD and UCAC vs UC. A decrease in the inflammatory process was also observed in the CAC conditions compared to their respective IBD group. Further analysis of OPN revealed a tendency for a poorer survival in high-expressing patients. OPN was present in both epithelium and stroma of CAC patients. Investigation of molecules related to OPN signaling showed that CD44 has a tendency for being increased in CAC patients, as well as  $\beta$ -catenin nuclear translocation. No changes were observed in phospho-SMAD3. There was a tendency for increased EMT in CAC patients as seen by the upregulation of FN1 and ZEB1 in CDAC patients and the immunohistochemical analysis of EpCAM by immunohistochemistry.

On the second part of the project, intestinal cell lines showed phosphorylation of ERK1/2, but not of STAT3 or AKT. Nuclear translocation of P-65 was also not observed after OPN exposure. OPN failed to induce EMT-genes transcriptional regulation in both intestinal cell lines,

being only a slight upregulation in SLUG observed in HT29/B6. Attempting to have a better comprehension of the effects of OPN in the intestinal cell lines, an WTS RNA-Seq experiment was performed. Among the most regulated genes in HT29/B6 there are genes of DNA-damage response and chromatin remodeling. On the other hand, in T84 the most downregulated genes are related to the cytoskeleton and the extracellular matrix. Enrichment analysis showed enrichment of gene sets related to mitochondrial respiration. Of those gene sets, the regulated genes were associated with the complexes I, III and IV, being the complexes III and IV downregulated in both cell lines. Since OPN is known to induce apoptosis and reactive oxygen species production through the mitochondrial death pathway, this could be a new mechanism by which it contributes to the tumorigenesis of CAC.

## 7 APPENDIX – LIST OF PUBLICATIONS

1. Cardoso-Silva D, Delbue D, Itzlinger A, Moerkens R, Withoff S, Branchi F, Schumann M. Intestinal Barrier Function in Gluten-Related Disorders. *Nutrients*. 2019 Oct 1;11(10):2325. doi: 10.3390/nu11102325. PMID: 31581491; PMCID: PMC6835310.
2. Delbue D, Cardoso-Silva D, Branchi F, Itzlinger A, Letizia M, Siegmund B, Schumann M. Celiac Disease Monocytes Induce a Barrier Defect in Intestinal Epithelial Cells. *Int J Mol Sci*. 2019 Nov 9;20(22):5597. doi: 10.3390/ijms20225597. PMID: 31717494; PMCID: PMC6888450.
3. Delbue D., Lebenheim L., Cardoso-Silva D., Dony D., Krug S. M., Richter J., Munoz M., Wolk K., Heldt C., Heimesaat M., Sabat R., Siegmund B., Schumann M. Reprogramming intestinal epithelial cell polarity by interleukin-22. *Frontiers in Medicine - Gastroenterology*. 2021 (accepted).
4. Sehn M.Ş, Cardoso-Silva D.Ş, Manna S., Weiner J., Weixler B., Gröne J., Siegmund B., Elezkurtaj S., Hummel M., Schumann M. Osteopontin in colitis-associated carcinoma (in progress)

## 8 REFERENCES

1. Cardoso-Silva D, Delbue D, Itzlinger A, Moerkens R, Withoff S, Branchi F, et al. Intestinal Barrier Function in Gluten-Related Disorders. *Nutrients*. 2019 Oct 1;11(10).
2. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 2016;14(8):e1002533.
3. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature*. 2011 May 12;473(7346):174–80.
4. Allam-Ndoul B, Castonguay-Paradis S, Veilleux A. Gut Microbiota and Intestinal Trans-Epithelial Permeability. *Int J Mol Sci*. 2020 Jan;21(17):6402.
5. Carlsson AH, Yakymenko O, Olivier I, Håkansson F, Postma E, Keita ÅV, et al. Faecalibacterium prausnitzii supernatant improves intestinal barrier function in mice DSS colitis. *Scand J Gastroenterol*. 2013 Oct 1;48(10):1136–44.
6. Furet J-P, Kong L-C, Tap J, Poitou C, Basdevant A, Bouillot J-L, et al. Differential Adaptation of Human Gut Microbiota to Bariatric Surgery–Induced Weight Loss. *Diabetes*. 2010 Dec;59(12):3049–57.
7. Leslie JL, Huang S, Opp JS, Nagy MS, Kobayashi M, Young VB, et al. Persistence and Toxin Production by *Clostridium difficile* within Human Intestinal Organoids Result in Disruption of Epithelial Paracellular Barrier Function. *Infect Immun*. 2015 Jan;83(1):138–45.
8. In J, Foulke-Abel J, Zachos NC, Hansen A-M, Kaper JB, Bernstein HD, et al. Enterohemorrhagic *Escherichia coli* Reduces Mucus and Intermicrovillar Bridges in Human Stem Cell-Derived Colonoids. *Cell Mol Gastroenterol Hepatol*. 2015 Oct 22;2(1):48-62.e3.
9. Bansil R, Turner BS. The biology of mucus: Composition, synthesis and organization. *Adv Drug Deliv Rev*. 2018 Jan 15;124:3–15.
10. Paone P, Cani PD. Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut* [Internet]. 2020 Sep 10 [cited 2020 Oct 4]; Available from: <https://gut.bmj.com/content/early/2020/09/11/gutjnl-2020-322260>
11. Pelaseyed T, Hansson GC. Membrane mucins of the intestine at a glance. *J Cell Sci*. 2020 Mar 13;133(5).
12. Clamp JR, Creeth JM. Some non-mucin components of mucus and their possible biological roles. *Ciba Found Symp*. 1984;109:121–36.
13. Zhao L, Lu W. Defensins in innate immunity. *Curr Opin Hematol*. 2014 Jan;21(1):37–42.

## Gene expression analysis to study celiac disease and colitis-associated cancer

14. Lehrer RI, Lu W.  $\alpha$ -Defensins in human innate immunity. *Immunol Rev*. 2012 Jan;245(1):84–112.
15. Semple F, Dorin JR.  $\beta$ -Defensins: multifunctional modulators of infection, inflammation and more? *J Innate Immun*. 2012;4(4):337–48.
16. Ragland SA, Criss AK. From bacterial killing to immune modulation: Recent insights into the functions of lysozyme. *PLoS Pathog* [Internet]. 2017 Sep 21 [cited 2020 Oct 8];13(9). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608400/>
17. Clevers H. The Intestinal Crypt, A Prototype Stem Cell Compartment. *Cell*. 2013 Jul 18;154(2):274–84.
18. Bjerknes M, Cheng H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology*. 1999 Jan 1;116(1):7–14.
19. Wang X, Yamamoto Y, Wilson LH, Zhang T, Howitt BE, Farrow MA, et al. Cloning and variation of ground state intestinal stem cells. *Nature*. 2015 Jun 11;522(7555):173–8.
20. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. *Science*. 2010 Jan 29;327(5965):542–5.
21. de Sousa E Melo F, de Sauvage FJ. Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell*. 2019 Jan 3;24(1):54–64.
22. Kim T-H, Saadatpour A, Guo G, Saxena M, Cavazza A, Desai N, et al. Single-Cell Transcript Profiles Reveal Multilineage Priming in Early Progenitors Derived from Lgr5(+) Intestinal Stem Cells. *Cell Rep*. 2016 Aug 23;16(8):2053–60.
23. van Es JH, van Gijn ME, Riccio O, van den Born M, Vooijs M, Begthel H, et al. Notch/ $\gamma$ -secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature*. 2005 Jun;435(7044):959–63.
24. Schonhoff SE, Giel-Moloney M, Leiter AB. Minireview: Development and differentiation of gut endocrine cells. *Endocrinology*. 2004 Jun;145(6):2639–44.
25. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol*. 2011 May;9(5):356–68.
26. van Es JH, Jay P, Gregorieff A, van Gijn ME, Jonkheer S, Hatzis P, et al. Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nat Cell Biol*. 2005 Apr;7(4):381–6.
27. Mei X, Gu M, Li M. Plasticity of Paneth cells and their ability to regulate intestinal stem cells. *Stem Cell Res Ther* [Internet]. 2020 Aug 12 [cited 2020 Oct 19];11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7425583/>

## Gene expression analysis to study celiac disease and colitis-associated cancer

28. Kucharzik T, Lügering N, Rautenberg K, Lügering A, Schmidt MA, Stoll R, et al. Role of M cells in intestinal barrier function. *Ann N Y Acad Sci.* 2000;915:171–83.
29. Ting H-A, von Moltke J. The Immune Function of Tuft Cells at Gut Mucosal Surfaces and Beyond. *J Immunol Baltim Md 1950.* 2019 Mar 1;202(5):1321–9.
30. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol.* 2009;71:241–60.
31. Ferguson A. Intraepithelial lymphocytes of the small intestine. *Gut.* 1977 Nov;18(11):921–37.
32. Mayassi T, Jabri B. Human intraepithelial lymphocytes. *Mucosal Immunol.* 2018 Sep;11(5):1281–9.
33. Bandeira A, Mota-Santos T, Itohara S, Degermann S, Heusser C, Tonegawa S, et al. Localization of gamma/delta T cells to the intestinal epithelium is independent of normal microbial colonization. *J Exp Med.* 1990 Jul 1;172(1):239–44.
34. Umesaki Y, Setoyama H, Matsumoto S, Okada Y. Expansion of alpha beta T-cell receptor-bearing intestinal intraepithelial lymphocytes after microbial colonization in germ-free mice and its independence from thymus. *Immunology.* 1993 May;79(1):32–7.
35. Jabri B, Selby JM, Negulescu H, Lee L, Roberts AI, Beavis A, et al. TCR specificity dictates CD94/NKG2A expression by human CTL. *Immunity.* 2002 Oct;17(4):487–99.
36. Groh V, Steinle A, Bauer S, Spies T. Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. *Science.* 1998 Mar 13;279(5357):1737–40.
37. Guy-Grand D, Cuénod-Jabri B, Malassis-Seris M, Selz F, Vassalli P. Complexity of the mouse gut T cell immune system: identification of two distinct natural killer T cell intraepithelial lineages. *Eur J Immunol.* 1996 Sep;26(9):2248–56.
38. Baum B, Georgiou M. Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling. *J Cell Biol.* 2011 Mar 21;192(6):907–17.
39. Bershadsky A. Magic touch: how does cell-cell adhesion trigger actin assembly? *Trends Cell Biol.* 2004 Nov;14(11):589–93.
40. Whitehead J, Vignjevic D, Fütterer C, Beaurepaire E, Robine S, Farge E. Mechanical factors activate beta-catenin-dependent oncogene expression in APC mouse colon. *HFSP J.* 2008 Oct;2(5):286–94.
41. Tsukita S, Furuse M, Itoh M. Multifunctional strands in tight junctions. *Nat Rev Mol Cell Biol.* 2001 Apr;2(4):285–93.

## Gene expression analysis to study celiac disease and colitis-associated cancer

42. Zihni C, Mills C, Matter K, Balda MS. Tight junctions: from simple barriers to multifunctional molecular gates. *Nat Rev Mol Cell Biol.* 2016 Sep;17(9):564–80.
43. Staehelin LA. Further observations on the fine structure of freeze-cleaved tight junctions. *J Cell Sci.* 1973 Nov;13(3):763–86.
44. Higashi T, Tokuda S, Kitajiri S, Masuda S, Nakamura H, Oda Y, et al. Analysis of the “angulin” proteins LSR, ILDR1 and ILDR2--tricellulin recruitment, epithelial barrier function and implication in deafness pathogenesis. *J Cell Sci.* 2013 Feb 15;126(Pt 4):966–77.
45. Mineta K, Yamamoto Y, Yamazaki Y, Tanaka H, Tada Y, Saito K, et al. Predicted expansion of the claudin multigene family. *FEBS Lett.* 2011 Feb 18;585(4):606–12.
46. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, et al. PDZ domain binding selectivity is optimized across the mouse proteome. *Science.* 2007 Jul 20;317(5836):364–9.
47. Günzel D, Fromm M. Claudins and other tight junction proteins. *Compr Physiol.* 2012 Jul;2(3):1819–52.
48. Van Itallie CM, Rogan S, Yu A, Vidal LS, Holmes J, Anderson JM. Two splice variants of claudin-10 in the kidney create paracellular pores with different ion selectivities. *Am J Physiol Renal Physiol.* 2006 Dec;291(6):F1288-1299.
49. Günzel D, Stuver M, Kausalya PJ, Haisch L, Krug SM, Rosenthal R, et al. Claudin-10 exists in six alternatively spliced isoforms that exhibit distinct localization and function. *J Cell Sci.* 2009 May 15;122(Pt 10):1507–17.
50. Günzel D, Yu ASL. Claudins and the modulation of tight junction permeability. *Physiol Rev.* 2013 Apr;93(2):525–69.
51. Lippoldt A, Liebner S, Andbjør B, Kalbacher H, Wolburg H, Haller H, et al. Organization of choroid plexus epithelial and endothelial cell tight junctions and regulation of claudin-1, -2 and -5 expression by protein kinase C. *Neuroreport.* 2000 May 15;11(7):1427–31.
52. Morita K, Sasaki H, Fujimoto K, Furuse M, Tsukita S. Claudin-11/OSP-based tight junctions of myelin sheaths in brain and Sertoli cells in testis. *J Cell Biol.* 1999 May 3;145(3):579–88.
53. Wolburg H, Wolburg-Buchholz K, Liebner S, Engelhardt B. Claudin-1, claudin-2 and claudin-11 are present in tight junctions of choroid plexus epithelium of the mouse. *Neurosci Lett.* 2001 Jul 13;307(2):77–80.
54. Kitajiri S, Furuse M, Morita K, Saishin-Kiuchi Y, Kido H, Ito J, et al. Expression patterns of claudins, tight junction adhesion molecules, in the inner ear. *Hear Res.* 2004 Jan;187(1–2):25–34.

## Gene expression analysis to study celiac disease and colitis-associated cancer

55. Niimi T, Nagashima K, Ward JM, Minoo P, Zimonjic DB, Popescu NC, et al. claudin-18, a novel downstream target gene for the T/EBP/NKX2.1 homeodomain transcription factor, encodes lung- and stomach-specific isoforms through alternative splicing. *Mol Cell Biol*. 2001 Nov;21(21):7380–90.
56. Kaarteenaho-Wiik R, Soini Y. Claudin-1, -2, -3, -4, -5, and -7 in usual interstitial pneumonia and sarcoidosis. *J Histochem Cytochem Off J Histochem Soc*. 2009 Mar;57(3):187–95.
57. Brandner JM, Kief S, Grund C, Rendl M, Houdek P, Kuhn C, et al. Organization and formation of the tight junction system in human epidermis and cultured keratinocytes. *Eur J Cell Biol*. 2002 May;81(5):253–63.
58. Inai T, Kobayashi J, Shibata Y. Claudin-1 contributes to the epithelial barrier function in MDCK cells. *Eur J Cell Biol*. 1999 Dec;78(12):849–55.
59. Dubé E, Chan PTK, Hermo L, Cyr DG. Gene expression profiling and its relevance to the blood-epididymal barrier in the human epididymis. *Biol Reprod*. 2007 Jun;76(6):1034–44.
60. D'Souza T, Sherman-Baust CA, Poosala S, Mullin JM, Morin PJ. Age-related changes of claudin expression in mouse liver, kidney, and pancreas. *J Gerontol A Biol Sci Med Sci*. 2009 Nov;64(11):1146–53.
61. Rahner C, Mitic LL, Anderson JM. Heterogeneity in expression and subcellular localization of claudins 2, 3, 4, and 5 in the rat liver, pancreas, and gut. *Gastroenterology*. 2001 Feb;120(2):411–22.
62. Yoshida Y, Ban Y, Kinoshita S. Tight junction transmembrane protein claudin subtype expression and distribution in human corneal and conjunctival epithelium. *Invest Ophthalmol Vis Sci*. 2009 May;50(5):2103–8.
63. Németh Z, Szász AM, Tátrai P, Németh J, Gyorffy H, Somorácz A, et al. Claudin-1, -2, -3, -4, -7, -8, and -10 protein expression in biliary tract cancers. *J Histochem Cytochem Off J Histochem Soc*. 2009 Feb;57(2):113–21.
64. Holmes JL, Van Itallie CM, Rasmussen JE, Anderson JM. Claudin profiling in the mouse during postnatal intestinal development and along the gastrointestinal tract reveals complex expression patterns. *Gene Expr Patterns GEP*. 2006 Aug;6(6):581–8.
65. Fujita H, Chiba H, Yokozaki H, Sakai N, Sugimoto K, Wada T, et al. Differential expression and subcellular localization of claudin-7, -8, -12, -13, and -15 along the mouse intestine. *J Histochem Cytochem Off J Histochem Soc*. 2006 Aug;54(8):933–44.
66. Kiuchi-Saishin Y, Gotoh S, Furuse M, Takasuga A, Tano Y, Tsukita S. Differential expression patterns of claudins, tight junction membrane proteins, in mouse nephron segments. *J Am Soc Nephrol JASN*. 2002 Apr;13(4):875–86.
67. Koda R, Zhao L, Yaoita E, Yoshida Y, Tsukita S, Tamura A, et al. Novel expression of claudin-5 in glomerular podocytes. *Cell Tissue Res*. 2011 Mar;343(3):637–48.

## Gene expression analysis to study celiac disease and colitis-associated cancer

68. Zhao L, Yaoita E, Nameta M, Zhang Y, Cuellar LM, Fujinaka H, et al. Claudin-6 localized in tight junctions of rat podocytes. *Am J Physiol Regul Integr Comp Physiol*. 2008 Jun;294(6):R1856-1862.
69. Abuazza G, Becker A, Williams SS, Chakravarty S, Truong H-T, Lin F, et al. Claudins 6, 9, and 13 are developmentally expressed renal tight junction proteins. *Am J Physiol Renal Physiol*. 2006 Dec;291(6):F1132-1141.
70. Enck AH, Berger UV, Yu AS. Claudin-2 is selectively expressed in proximal nephron in mouse kidney. *Am J Physiol Renal Physiol*. 2001 Nov;281(5):F966-974.
71. Li WY, Huey CL, Yu ASL. Expression of claudin-7 and -8 along the mouse nephron. *Am J Physiol Renal Physiol*. 2004 Jun;286(6):F1063-1071.
72. Angelow S, El-Husseini R, Kanzawa SA, Yu ASL. Renal localization and function of the tight junction protein, claudin-19. *Am J Physiol Renal Physiol*. 2007 Jul;293(1):F166-177.
73. Hou J, Renigunta A, Gomes AS, Hou M, Paul DL, Waldegger S, et al. Claudin-16 and claudin-19 interaction is required for their assembly into tight junctions and for renal reabsorption of magnesium. *Proc Natl Acad Sci U S A*. 2009 Sep 8;106(36):15350-5.
74. Konrad M, Schaller A, Seelow D, Pandey AV, Waldegger S, Lesslauer A, et al. Mutations in the tight-junction gene claudin 19 (CLDN19) are associated with renal magnesium wasting, renal failure, and severe ocular involvement. *Am J Hum Genet*. 2006 Nov;79(5):949-57.
75. Simon DB, Lu Y, Choate KA, Velazquez H, Al-Sabban E, Praga M, et al. Paracellin-1, a renal tight junction protein required for paracellular Mg<sup>2+</sup> resorption. *Science*. 1999 Jul 2;285(5424):103-6.
76. Alexandre MD, Lu Q, Chen Y-H. Overexpression of claudin-7 decreases the paracellular Cl<sup>-</sup> conductance and increases the paracellular Na<sup>+</sup> conductance in LLC-PK1 cells. *J Cell Sci*. 2005 Jun 15;118(Pt 12):2683-93.
77. Ben-Yosef T, Belyantseva IA, Saunders TL, Hughes ED, Kawamoto K, Van Itallie CM, et al. Claudin 14 knockout mice, a model for autosomal recessive deafness DFNB29, are deaf due to cochlear hair cell degeneration. *Hum Mol Genet*. 2003 Aug 15;12(16):2049-61.
78. Furuse M, Hata M, Furuse K, Yoshida Y, Haratake A, Sugitani Y, et al. Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J Cell Biol*. 2002 Mar 18;156(6):1099-111.
79. Zheng A, Yuan F, Li Y, Zhu F, Hou P, Li J, et al. Claudin-6 and claudin-9 function as additional coreceptors for hepatitis C virus. *J Virol*. 2007 Nov;81(22):12465-71.
80. Wilcox ER, Burton QL, Naz S, Riazuddin S, Smith TN, Ploplis B, et al. Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell*. 2001 Jan 12;104(1):165-72.

## Gene expression analysis to study celiac disease and colitis-associated cancer

81. Blackman B, Russell T, Nordeen SK, Medina D, Neville MC. Claudin 7 expression and localization in the normal murine mammary gland and murine mammary tumors. *Breast Cancer Res BCR*. 2005;7(2):R248-255.
82. Blanchard AAA, Watson PH, Shiu RPC, Leygue E, Nistor A, Wong P, et al. Differential expression of claudin 1, 3, and 4 during normal mammary gland development in the mouse. *DNA Cell Biol*. 2006 Feb;25(2):79–86.
83. Jakab C, Halász J, Szász AM, Kiss A, Schaff Z, Rusvai M, et al. Expression of claudin-1, -2, -3, -4, -5 and -7 proteins in benign and malignant canine mammary gland epithelial tumours. *J Comp Pathol*. 2008 Nov;139(4):238–45.
84. Markov AG, Kruglova NM, Fomina YA, Fromm M, Amasheh S. Altered expression of tight junction proteins in mammary epithelium after discontinued suckling in mice. *Pflugers Arch*. 2012 Feb;463(2):391–8.
85. Peng S, Adelman RA, Rizzolo LJ. Minimal effects of VEGF and anti-VEGF drugs on the permeability or selectivity of RPE tight junctions. *Invest Ophthalmol Vis Sci*. 2010 Jun;51(6):3216–25.
86. Zhu Y, Brännström M, Janson P-O, Sundfeldt K. Differences in expression patterns of the tight junction proteins, claudin 1, 3, 4 and 5, in human ovarian surface epithelium as compared to epithelia in inclusion cysts and epithelial ovarian tumours. *Int J Cancer*. 2006 Apr 15;118(8):1884–91.
87. Sakai N, Chiba H, Fujita H, Akashi Y, Osanai M, Kojima T, et al. Expression patterns of claudin family of tight-junction proteins in the mouse prostate. *Histochem Cell Biol*. 2007 Apr;127(4):457–62.
88. Greenwell-Wild T, Moutsopoulos NM, Gliozzi M, Kapsogeorgou E, Rangel Z, Munson PJ, et al. Chitinases in the salivary glands and circulation of patients with Sjögren's syndrome: macrophage harbingers of disease severity. *Arthritis Rheum*. 2011 Oct;63(10):3103–15.
89. Maria OM, Kim J-WM, Gerstenhaber JA, Baum BJ, Tran SD. Distribution of tight junction proteins in adult human salivary glands. *J Histochem Cytochem Off J Histochem Soc*. 2008 Dec;56(12):1093–8.
90. Peppi M, Ghabriel MN. Tissue-specific expression of the tight junction proteins claudins and occludin in the rat salivary glands. *J Anat*. 2004 Oct;205(4):257–66.
91. Morrow CMK, Mruk D, Cheng CY, Hess RA. Claudin and occludin expression and function in the seminiferous epithelium. *Philos Trans R Soc Lond B Biol Sci*. 2010 May 27;365(1546):1679–96.
92. Hewitt KJ, Agarwal R, Morin PJ. The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer*. 2006 Jul 12;6:186.
93. Katoh M, Katoh M. CLDN23 gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of CLAUDIN gene family. *Int J Mol Med*. 2003 Jun;11(6):683–9.

## Gene expression analysis to study celiac disease and colitis-associated cancer

94. Michlig S, Damak S, Le Coutre J. Claudin-based permeability barriers in taste buds. *J Comp Neurol*. 2007 Jun 20;502(6):1003–11.
95. Acharya P, Beckel J, Ruiz WG, Wang E, Rojas R, Birder L, et al. Distribution of the tight junction proteins ZO-1, occludin, and claudin-4, -8, and -12 in bladder epithelium. *Am J Physiol Renal Physiol*. 2004 Aug;287(2):F305-318.
96. Furuse M, Hirase T, Itoh M, Nagafuchi A, Yonemura S, Tsukita S, et al. Occludin: a novel integral membrane protein localizing at tight junctions. *J Cell Biol*. 1993 Dec;123(6 Pt 2):1777–88.
97. Li Y, Fanning AS, Anderson JM, Lavie A. Structure of the conserved cytoplasmic C-terminal domain of occludin: identification of the ZO-1 binding surface. *J Mol Biol*. 2005 Sep 9;352(1):151–64.
98. Schulzke JD, Gitter AH, Mankertz J, Spiegel S, Seidler U, Amasheh S, et al. Epithelial transport and barrier function in occludin-deficient mice. *Biochim Biophys Acta*. 2005 May 15;1669(1):34–42.
99. Steed E, Rodrigues NTL, Balda MS, Matter K. Identification of MarvelD3 as a tight junction-associated transmembrane protein of the occludin family. *BMC Cell Biol*. 2009 Dec 22;10:95.
100. Krug SM, Amasheh S, Richter JF, Milatz S, Günzel D, Westphal JK, et al. Tricellulin forms a barrier to macromolecules in tricellular tight junctions without affecting ion permeability. *Mol Biol Cell*. 2009 Aug;20(16):3713–24.
101. Ikenouchi J, Furuse M, Furuse K, Sasaki H, Tsukita S, Tsukita S. Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells. *J Cell Biol*. 2005 Dec 19;171(6):939–45.
102. Otani T, Nguyen TP, Tokuda S, Sugihara K, Sugawara T, Furuse K, et al. Claudins and JAM-A coordinately regulate tight junction formation and epithelial polarity. *J Cell Biol*. 2019 Oct 7;218(10):3372–96.
103. Rodgers LS, Beam MT, Anderson JM, Fanning AS. Epithelial barrier assembly requires coordinated activity of multiple domains of the tight junction protein ZO-1. *J Cell Sci*. 2013 Apr 1;126(Pt 7):1565–75.
104. Ikenouchi J, Umeda K, Tsukita S, Furuse M, Tsukita S. Requirement of ZO-1 for the formation of belt-like adherens junctions during epithelial cell polarization. *J Cell Biol*. 2007 Mar 12;176(6):779–86.
105. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. *Nat Rev Immunol*. 2014 Oct;14(10):667–85.
106. Herbrand H, Bernhardt G, Förster R, Pabst O. Dynamics and function of solitary intestinal lymphoid tissue. *Crit Rev Immunol*. 2008;28(1):1–13.
107. Cornes JS. Number, size, and distribution of Peyer's patches in the human small intestine: Part I The development of Peyer's patches. *Gut*. 1965 Jun;6(3):225–9.

108. Trepel F. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr.* 1974 Jun 1;52(11):511–5.
109. Sathaliyawala T, Kubota M, Yudanin N, Turner D, Camp P, Thome JJC, et al. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity.* 2013 Jan 24;38(1):187–97.
110. Maynard CL, Harrington LE, Janowski KM, Oliver JR, Zindl CL, Rudensky AY, et al. Regulatory T cells expressing interleukin 10 develop from Foxp3+ and Foxp3- precursor cells in the absence of interleukin 10. *Nat Immunol.* 2007 Sep;8(9):931–41.
111. Brandtzaeg P, Johansen F-E. Mucosal B cells: phenotypic characteristics, transcriptional regulation, and homing properties. *Immunol Rev.* 2005 Aug;206:32–63.
112. Yu X, Vargas J, Green PHR, Bhagat G. Innate Lymphoid Cells and Celiac Disease: Current Perspective. *Cell Mol Gastroenterol Hepatol.* 2020 Dec 10;11(3):803–14.
113. Nussbaum JC, Van Dyken SJ, von Moltke J, Cheng LE, Mohapatra A, Molofsky AB, et al. Type 2 innate lymphoid cells control eosinophil homeostasis. *Nature.* 2013 Oct 10;502(7470):245–8.
114. Satoh-Takayama N, Lesjean-Pottier S, Sawa S, Vosshenrich CAJ, Eberl G, Di Santo JP. Lymphotoxin- $\beta$  receptor-independent development of intestinal IL-22-producing NKp46+ innate lymphoid cells. *Eur J Immunol.* 2011 Mar;41(3):780–6.
115. Le Bourhis L, Guerri L, Dusseaux M, Martin E, Soudais C, Lantz O. Mucosal-associated invariant T cells: unconventional development and function. *Trends Immunol.* 2011 May;32(5):212–8.
116. Zeissig S, Blumberg RS. Commensal microbiota and NKT cells in the control of inflammatory diseases at mucosal surfaces. *Curr Opin Immunol.* 2013 Dec;25(6):690–6.
117. Cerovic V, Bain CC, Mowat AM, Milling SWF. Intestinal macrophages and dendritic cells: what's the difference? *Trends Immunol.* 2014 Jun;35(6):270–7.
118. Ueda Y, Kayama H, Jeon SG, Kusu T, Isaka Y, Rakugi H, et al. Commensal microbiota induce LPS hyporesponsiveness in colonic macrophages via the production of IL-10. *Int Immunol.* 2010 Dec;22(12):953–62.
119. Hadis U, Wahl B, Schulz O, Hardtke-Wolenski M, Schippers A, Wagner N, et al. Intestinal tolerance requires gut homing and expansion of FoxP3+ regulatory T cells in the lamina propria. *Immunity.* 2011 Feb 25;34(2):237–46.
120. Persson EK, Scott CL, Mowat AM, Agace WW. Dendritic cell subsets in the intestinal lamina propria: ontogeny and function. *Eur J Immunol.* 2013 Dec;43(12):3098–107.

## Gene expression analysis to study celiac disease and colitis-associated cancer

121. Chu VT, Beller A, Rausch S, Strandmark J, Zänker M, Arbach O, et al. Eosinophils promote generation and maintenance of immunoglobulin-A-expressing plasma cells and contribute to gut immune homeostasis. *Immunity*. 2014 Apr 17;40(4):582–93.
122. Yu LC, Perdue MH. Role of mast cells in intestinal mucosal function: studies in models of hypersensitivity and stress. *Immunol Rev*. 2001 Feb;179:61–73.
123. Lebwohl B, Rubio-Tapia A. Epidemiology, Presentation, and Diagnosis of Celiac Disease. *Gastroenterology*. 2021 Jan;160(1):63–75.
124. Therrien A, Kelly CP, Silvester JA. Celiac Disease: Extraintestinal Manifestations and Associated Conditions. *J Clin Gastroenterol*. 2020 Jan;54(1):8–21.
125. Singh P, Arora A, Strand TA, Leffler DA, Catassi C, Green PH, et al. Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2018 Jun;16(6):823-836.e2.
126. King JA, Jeong J, Underwood FE, Quan J, Panaccione N, Windsor JW, et al. Incidence of Celiac Disease Is Increasing Over Time: A Systematic Review and Meta-analysis. *Am J Gastroenterol*. 2020 Apr;115(4):507–25.
127. Liu E, Dong F, Barón AE, Taki I, Norris JM, Frohnert BI, et al. High Incidence of Celiac Disease in a Long-term Study of Adolescents With Susceptibility Genotypes. *Gastroenterology*. 2017 May;152(6):1329-1336.e1.
128. Lebwohl B, Tennyson CA, Holub JL, Lieberman DA, Neugut AI, Green PHR. Gender and Racial Disparities in Duodenal Biopsy to Evaluate For Celiac Disease. *Gastrointest Endosc*. 2012 Oct;76(4):779–85.
129. Brown NK, Guandalini S, Semrad C, Kupfer SS. A Clinician’s Guide to Celiac Disease HLA Genetics. *Am J Gastroenterol*. 2019 Oct;114(10):1587–92.
130. Megiorni F, Mora B, Bonamico M, Barbato M, Nenna R, Maiella G, et al. HLA-DQ and risk gradient for celiac disease. *Hum Immunol*. 2009 Jan;70(1):55–9.
131. Al-Toma A, Volta U, Auricchio R, Castillejo G, Sanders DS, Cellier C, et al. European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders. *United Eur Gastroenterol J*. 2019 Jun;7(5):583–613.
132. Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raia V, Auricchio S, et al. Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet Lond Engl*. 2003 Jul 5;362(9377):30–7.
133. Lindfors K, Ciacci C, Kurppa K, Lundin KEA, Makharia GK, Mearin ML, et al. Coeliac disease. *Nat Rev Dis Primer*. 2019 Dec;5(1):3.

## Gene expression analysis to study celiac disease and colitis-associated cancer

134. Bodd M, Ráki M, Bergseng E, Jahnsen J, Lundin KEA, Sollid LM. Direct cloning and tetramer staining to measure the frequency of intestinal gluten-reactive T cells in celiac disease. *Eur J Immunol*. 2013;43(10):2605–12.
135. Stamnaes J, Sollid LM. Celiac disease: Autoimmunity in response to food antigen. *Semin Immunol*. 2015 Sep;27(5):343–52.
136. Hujoel IA, Murray JA. Refractory Celiac Disease. *Curr Gastroenterol Rep*. 2020 Mar 17;22(4):18.
137. Malamut G, El Machhour R, Montcuquet N, Martin-Lannerée S, Dusanter-Fourt I, Verkarre V, et al. IL-15 triggers an antiapoptotic pathway in human intraepithelial lymphocytes that is a potential new target in celiac disease–associated inflammation and lymphomagenesis. *J Clin Invest*. 2010 Jun 1;120(6):2131–43.
138. Cobden I, Dickinson RJ, Rothwell J, Axon AT. Intestinal permeability assessed by excretion ratios of two molecules: results in coeliac disease. *Br Med J*. 1978 Oct 14;2(6144):1060–1060.
139. Menzies IS, Laker MF, Pounder R, Bull J, Heyer S, Wheeler PG, et al. Abnormal intestinal permeability to sugars in villous atrophy. *Lancet Lond Engl*. 1979 Nov 24;2(8152):1107–9.
140. Pearson AD, Eastham EJ, Laker MF, Craft AW, Nelson R. Intestinal permeability in children with Crohn’s disease and coeliac disease. *Br Med J Clin Res Ed*. 1982 Jul 3;285(6334):20–1.
141. Schulzke JD, Bentzel CJ, Schulzke I, Riecken EO, Fromm M. Epithelial tight junction structure in the jejunum of children with acute and treated celiac sprue. *Pediatr Res*. 1998 Apr;43(4 Pt 1):435–41.
142. Schulzke JD, Schulzke I, Fromm M, Riecken EO. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. *Gut*. 1995 Dec;37(6):777–82.
143. Schumann M, Günzel D, Buergel N, Richter JF, Troeger H, May C, et al. Cell polarity-determining proteins Par-3 and PP-1 are involved in epithelial tight junction defects in coeliac disease. *Gut*. 2012 Feb;61(2):220–8.
144. Schumann M, Siegmund B, Schulzke JD, Fromm M. Celiac Disease: Role of the Epithelial Barrier. *Cell Mol Gastroenterol Hepatol*. 2017 Mar;3(2):150–62.
145. Ciccocioppo R, Finamore A, Ara C, Di Sabatino A, Mengheri E, Corazza GR. Altered expression, localization, and phosphorylation of epithelial junctional proteins in celiac disease. *Am J Clin Pathol*. 2006 Apr;125(4):502–11.
146. Barone MV, Troncone R, Auricchio S. Gliadin peptides as triggers of the proliferative and stress/innate immune response of the celiac small intestinal mucosa. *Int J Mol Sci*. 2014 Nov 7;15(11):20518–37.

## Gene expression analysis to study celiac disease and colitis-associated cancer

147. Lammers KM, Lu R, Brownley J, Lu B, Gerard C, Thomas K, et al. Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3. *Gastroenterology*. 2008 Jul;135(1):194-204.e3.
148. van Elburg RM, Uil JJ, Mulder CJ, Heymans HS. Intestinal permeability in patients with coeliac disease and relatives of patients with coeliac disease. *Gut*. 1993 Mar;34(3):354–7.
149. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet*. 2007 Jul;39(7):827–9.
150. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010 Apr;42(4):295–302.
151. Kumar V, Gutierrez-Achury J, Kanduri K, Almeida R, Hrdlickova B, Zhernakova DV, 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 Jan 15;24(2):397–409.
152. Petit MM, Mols R, Schoenmakers EF, Mandahl N, Van de Ven WJ. LPP, the preferred fusion partner gene of HMGIC in lipomas, is a novel member of the LIM protein gene family. *Genomics*. 1996 Aug 15;36(1):118–29.
153. Petit MMR, Fradelizi J, Golsteyn RM, Ayoubi TAY, Menichi B, Louvard D, et al. LPP, an Actin Cytoskeleton Protein Related to Zyxin, Harbors a Nuclear Export Signal and Transcriptional Activation Capacity. *Mol Biol Cell*. 2000 Jan;11(1):117–29.
154. Van Itallie CM, Tietgens AJ, Aponte A, Fredriksson K, Fanning AS, Gucek M, et al. Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell-cell and cell-substrate adhesion. *J Cell Sci*. 2014 Feb 15;127(Pt 4):885–95.
155. Huang S-Q, Zhang N, Zhou Z-X, Huang C-C, Zeng C-L, Xiao D, et al. Association of LPP and TAGAP Polymorphisms with Celiac Disease Risk: A Meta-Analysis. *Int J Environ Res Public Health*. 2017 Feb 10;14(2).
156. Ngan E, Kiepas A, Brown CM, Siegel PM. Emerging roles for LPP in metastatic cancer progression. *J Cell Commun Signal*. 2018 Mar;12(1):143–56.
157. Medrano LM, Pascual V, Bodas A, López-Palacios N, Salazar I, Espino-Paisán L, et al. Expression patterns common and unique to ulcerative colitis and celiac disease. *Ann Hum Genet*. 2019 Mar;83(2):86–94.
158. Tang J, Zhang C-B, Lyu K-S, Jin Z-M, Guan S-X, You N, et al. Association of polymorphisms in C1orf106, IL1RN, and IL10 with post-induction infliximab trough level in Crohn's disease patients. *Gastroenterol Rep*. 2020 Oct;8(5):367–73.

## Gene expression analysis to study celiac disease and colitis-associated cancer

159. Manzanillo P, Mouchess M, Ota N, Dai B, Ichikawa R, Wuster A, et al. Inflammatory Bowel Disease Susceptibility Gene C1ORF106 Regulates Intestinal Epithelial Permeability. *ImmunoHorizons*. 2018 May 30;2(5):164–71.
160. Mohanan V, Nakata T, Desch AN, Lévesque C, Boroughs A, Guzman G, et al. C1orf106 is a colitis risk gene that regulates stability of epithelial adherens junctions. *Science*. 2018 Mar 9;359(6380):1161–6.
161. Palacios F, Price L, Schweitzer J, Collard JG, D'Souza-Schorey C. An essential role for ARF6-regulated membrane traffic in adherens junction turnover and epithelial cell migration. *EMBO J*. 2001 Sep 3;20(17):4973–86.
162. Yu YR, Rodriguez JR. Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. *Semin Pediatr Surg*. 2017 Dec;26(6):349–55.
163. Canani RB, Terrin G, Rapacciuolo L, Miele E, Siani MC, Puzone C, et al. Faecal calprotectin as reliable non-invasive marker to assess the severity of mucosal inflammation in children with inflammatory bowel disease. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. 2008 Jul;40(7):547–53.
164. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet Lond Engl*. 2018 23;390(10114):2769–78.
165. Feuerstein JD, Moss AC, Farraye FA. Ulcerative Colitis. *Mayo Clin Proc*. 2019 Jul 1;94(7):1357–73.
166. Porter RJ, Kalla R, Ho G-T. Ulcerative colitis: Recent advances in the understanding of disease pathogenesis. *F1000Research* [Internet]. 2020 Apr 24 [cited 2021 Feb 18];9. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7194476/>
167. Beaudoin M, Goyette P, Boucher G, Lo KS, Rivas MA, Stevens C, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet*. 2013;9(9):e1003723.
168. Luo Y, de Lange KM, Jostins L, Moutsianas L, Randall J, Kennedy NA, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet*. 2017 Feb;49(2):186–92.
169. Huang H, Fang M, Jostins L, Umičević Mirkov M, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017 Jul 13;547(7662):173–8.
170. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012 Nov 1;491(7422):119–24.

## Gene expression analysis to study celiac disease and colitis-associated cancer

171. Chen G-B, Lee SH, Brion M-JA, Montgomery GW, Wray NR, Radford-Smith GL, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum Mol Genet.* 2014 Sep 1;23(17):4710–20.
172. Kaplan GG, Ng SC. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. *Gastroenterology.* 2017 Feb;152(2):313-321.e2.
173. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin Proc.* 2006 Nov;81(11):1462–71.
174. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun.* 2017 Dec 5;8(1):1784.
175. Turner JR. Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol.* 2009 Nov;9(11):799–809.
176. Watson CJ, Hoare CJ, Garrod DR, Carlson GL, Warhurst G. Interferon-gamma selectively increases epithelial permeability to large molecules by activating different populations of paracellular pores. *J Cell Sci.* 2005 Nov 15;118(Pt 22):5221–30.
177. Heller F, Fromm A, Gitter AH, Mankertz J, Schulzke J-D. Epithelial apoptosis is a prominent feature of the epithelial barrier disturbance in intestinal inflammation: effect of pro-inflammatory interleukin-13 on epithelial cell function. *Mucosal Immunol.* 2008 Nov;1 Suppl 1:S58-61.
178. Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. *Nat Rev Immunol.* 2003 Jul;3(7):521–33.
179. Teng MWL, Bowman EP, McElwee JJ, Smyth MJ, Casanova J-L, Cooper AM, et al. IL-12 and IL-23 cytokines: from discovery to targeted therapies for immune-mediated inflammatory diseases. *Nat Med.* 2015 Jul;21(7):719–29.
180. Sandborn WJ, Ferrante M, Bhandari BR, Berliba E, Feagan BG, Hibi T, et al. Efficacy and Safety of Mirikizumab in a Randomized Phase 2 Study of Patients With Ulcerative Colitis. *Gastroenterology.* 2020 Feb;158(3):537-549.e10.
181. Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johans J, et al. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med.* 2019 Sep 26;381(13):1201–14.
182. Veauthier B, Hornecker JR. Crohn's Disease: Diagnosis and Management. *Am Fam Physician.* 2018 Dec 1;98(11):661–9.
183. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015 Sep;47(9):979–86.

## Gene expression analysis to study celiac disease and colitis-associated cancer

184. Wang M-H, Picco MF. Crohn's Disease: Genetics Update. *Gastroenterol Clin North Am*. 2017 Sep;46(3):449–61.
185. Soon IS, Molodecky NA, Rabi DM, Ghali WA, Barkema HW, Kaplan GG. The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. *BMC Gastroenterol*. 2012 May 24;12:51.
186. Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol*. 2015 Apr;12(4):205–17.
187. Khanna R, Mosli MH, Feagan BG. Anti-Integrins in Ulcerative Colitis and Crohn's Disease: What Is Their Place? *Dig Dis Basel Switz*. 2016;34(1–2):153–9.
188. Mantzaris GJ. Anti-TNFs: Originators and Biosimilars. *Dig Dis Basel Switz*. 2016;34(1–2):132–9.
189. Cheifetz AS. Management of Active Crohn Disease. *JAMA*. 2013 May 22;309(20):2150–8.
190. Lutgens MWMD, van Oijen MGH, van der Heijden GJMG, Vleggaar FP, Siersema PD, Oldenburg B. Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. *Inflamm Bowel Dis*. 2013 Apr;19(4):789–99.
191. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990 Jun 1;61(5):759–67.
192. Cooks T, Pateras IS, Tarcic O, Solomon H, Schetter AJ, Wilder S, et al. Mutant p53 prolongs NF- $\kappa$ B activation and promotes chronic inflammation and inflammation-associated colorectal cancer. *Cancer Cell*. 2013 May 13;23(5):634–46.
193. Becker C, Fantini MC, Schramm C, Lehr HA, Wirtz S, Nikolaev A, et al. TGF-beta suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. *Immunity*. 2004 Oct;21(4):491–501.
194. Putoczki TL, Thiem S, Loving A, Busuttill RA, Wilson NJ, Ziegler PK, et al. Interleukin-11 is the dominant IL-6 family cytokine during gastrointestinal tumorigenesis and can be targeted therapeutically. *Cancer Cell*. 2013 Aug 12;24(2):257–71.
195. Popivanova BK, Kitamura K, Wu Y, Kondo T, Kagaya T, Kaneko S, et al. Blocking TNF-alpha in mice reduces colorectal carcinogenesis associated with chronic colitis. *J Clin Invest*. 2008 Feb;118(2):560–70.
196. Icer MA, Gezmen-Karadag M. The multiple functions and mechanisms of osteopontin. *Clin Biochem*. 2018 Sep;59:17–24.
197. Kunii Y, Niwa S, Hagiwara Y, Maeda M, Seitoh T, Suzuki T. The immunohistochemical expression profile of osteopontin in normal human tissues using two site-specific antibodies reveals a wide distribution of positive cells and extensive expression in the central and peripheral nervous systems. *Med Mol Morphol*. 2009 Sep;42(3):155–61.

## Gene expression analysis to study celiac disease and colitis-associated cancer

198. Franzén A, Heinegård D. Isolation and characterization of two sialoproteins present only in bone calcified matrix. *Biochem J.* 1985 Dec 15;232(3):715–24.
199. Young MF, Kerr JM, Termine JD, Wewer UM, Wang MG, McBride OW, et al. cDNA cloning, mRNA distribution and heterogeneity, chromosomal location, and RFLP analysis of human osteopontin (OPN). *Genomics.* 1990 Aug;7(4):491–502.
200. Sørensen ES, Rasmussen LK, Møller L, Jensen PH, Højrup P, Petersen TE. Localization of transglutaminase-reactive glutamine residues in bovine osteopontin. *Biochem J.* 1994 Nov 15;304(Pt 1):13–6.
201. Miyauchi A, Alvarez J, Greenfield EM, Teti A, Grano M, Colucci S, et al. Recognition of osteopontin and related peptides by an alpha v beta 3 integrin stimulates immediate cell signals in osteoclasts. *J Biol Chem.* 1991 Oct 25;266(30):20369–74.
202. Hu DD, Lin EC, Kovach NL, Hoyer JR, Smith JW. A biochemical characterization of the binding of osteopontin to integrins alpha v beta 1 and alpha v beta 5. *J Biol Chem.* 1995 Nov 3;270(44):26232–8.
203. Weber GF, Ashkar S, Glimcher MJ, Cantor H. Receptor-Ligand Interaction Between CD44 and Osteopontin (Eta-1). *Science.* 1996 Jan 26;271(5248):509–12.
204. Denda S, Reichardt LF, Müller U. Identification of Osteopontin as a Novel Ligand for the Integrin  $\alpha 8\beta 1$  and Potential Roles for This Integrin–Ligand Interaction in Kidney Morphogenesis. *Mol Biol Cell.* 1998 Jun;9(6):1425–35.
205. Yokosaki Y, Matsuura N, Sasaki T, Murakami I, Schneider H, Higashiyama S, et al. The Integrin  $\alpha 9\beta 1$  Binds to a Novel Recognition Sequence (SVVYGLR) in the Thrombin-cleaved Amino-terminal Fragment of Osteopontin. *J Biol Chem.* 1999 Dec 17;274(51):36328–34.
206. Green PM, Ludbrook SB, Miller DD, Horgan CM, Barry ST. Structural elements of the osteopontin SVVYGLR motif important for the interaction with alpha(4) integrins. *FEBS Lett.* 2001 Aug 10;503(1):75–9.
207. Bayless KJ, Davis GE. Identification of dual alpha 4beta1 integrin binding sites within a 38 amino acid domain in the N-terminal thrombin fragment of human osteopontin. *J Biol Chem.* 2001 Apr 20;276(16):13483–9.
208. Yokosaki Y, Tanaka K, Higashikawa F, Yamashita K, Eboshida A. Distinct structural requirements for binding of the integrins alphavbeta6, alphavbeta3, alphavbeta5, alpha5beta1 and alpha9beta1 to osteopontin. *Matrix Biol J Int Soc Matrix Biol.* 2005 Sep;24(6):418–27.
209. Briones-Orta MA, Avendaño-Vázquez SE, Aparicio-Bautista DI, Coombes JD, Weber GF, Syn W-K. Osteopontin splice variants and polymorphisms in cancer progression and prognosis. *Biochim Biophys Acta Rev Cancer.* 2017 Aug;1868(1):93-108.A.

## Gene expression analysis to study celiac disease and colitis-associated cancer

210. Mirza M, Shaughnessy E, Hurley JK, Vanpatten KA, Pestano GA, He B, et al. Osteopontin-c is a selective marker of breast cancer. *Int J Cancer*. 2008 Feb 15;122(4):889–97.
211. Patani N, Jouhra F, Jiang W, Mokbel K. Osteopontin expression profiles predict pathological and clinical outcome in breast cancer. *Anticancer Res*. 2008 Dec;28(6B):4105–10.
212. Siddiqui AA, Jones E, Andrade D, Shah A, Kowalski TE, Loren DE, et al. Osteopontin splice variant as a potential marker for metastatic disease in pancreatic adenocarcinoma. *J Gastroenterol Hepatol*. 2014 Jun;29(6):1321–7.
213. Tilli TM, Thuler LC, Matos AR, Coutinho-Camillo CM, Soares FA, da Silva EA, et al. Expression analysis of osteopontin mRNA splice variants in prostate cancer and benign prostatic hyperplasia. *Exp Mol Pathol*. 2012 Feb;92(1):13–9.
214. Yan W, Qian C, Zhao P, Zhang J, Shi L, Qian J, et al. Expression pattern of osteopontin splice variants and its functions on cell apoptosis and invasion in glioma cells. *Neuro-Oncol*. 2010 Aug;12(8):765–75.
215. Shinohara ML, Kim H-J, Kim J-H, Garcia VA, Cantor H. Alternative translation of osteopontin generates intracellular and secreted isoforms that mediate distinct biological activities in dendritic cells. *Proc Natl Acad Sci U S A*. 2008 May 20;105(20):7235–9.
216. Fan X, He C, Jing W, Zhou X, Chen R, Cao L, et al. Intracellular Osteopontin inhibits toll-like receptor signaling and impedes liver carcinogenesis. *Cancer Res*. 2015 Jan 1;75(1):86–97.
217. Leavenworth JW, Verbinnen B, Yin J, Huang H, Cantor H. A p85 $\alpha$ -osteopontin axis couples the receptor ICOS to sustained Bcl-6 expression by follicular helper and regulatory T cells. *Nat Immunol*. 2015 Jan;16(1):96–106.
218. Bellahcène A, Castronovo V, Ogbureke KUE, Fisher LW, Fedarko NS. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nat Rev Cancer*. 2008 Mar;8(3):212–26.
219. Teramoto H, Castellone MD, Malek RL, Letwin N, Frank B, Gutkind JS, et al. Autocrine activation of an osteopontin-CD44-Rac pathway enhances invasion and transformation by H-RasV12. *Oncogene*. 2005 Jan 13;24(3):489–501.
220. Agnihotri R, Crawford HC, Haro H, Matrisian LM, Havrda MC, Liaw L. Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin). *J Biol Chem*. 2001 Jul 27;276(30):28261–7.
221. Angelucci A, Festuccia C, Gravina GL, Muzi P, Bonghi L, Vicentini C, et al. Osteopontin enhances the cell proliferation induced by the epidermal growth factor in human prostate cancer cells. *The Prostate*. 2004 May 1;59(2):157–66.

## Gene expression analysis to study celiac disease and colitis-associated cancer

222. Das R, Mahabeleshwar GH, Kundu GC. Osteopontin stimulates cell motility and nuclear factor kappaB-mediated secretion of urokinase type plasminogen activator through phosphatidylinositol 3-kinase/Akt signaling pathways in breast cancer cells. *J Biol Chem*. 2003 Aug 1;278(31):28593–606.
223. Marcondes MCG, Poling M, Watry DD, Hall D, Fox HS. In vivo osteopontin-induced macrophage accumulation is dependent on CD44 expression. *Cell Immunol*. 2008;254(1):56–62.
224. Lund SA, Wilson CL, Raines EW, Tang J, Giachelli CM, Scatena M. Osteopontin mediates macrophage chemotaxis via  $\alpha 4$  and  $\alpha 9$  integrins and survival via the  $\alpha 4$  integrin. *J Cell Biochem*. 2013 May;114(5):1194–202.
225. Kawamura K, Iyonaga K, Ichiyasu H, Nagano J, Suga M, Sasaki Y. Differentiation, maturation, and survival of dendritic cells by osteopontin regulation. *Clin Diagn Lab Immunol*. 2005 Jan;12(1):206–12.
226. Renkl AC, Wussler J, Ahrens T, Thoma K, Kon S, Uede T, et al. Osteopontin functionally activates dendritic cells and induces their differentiation toward a Th1-polarizing phenotype. *Blood*. 2005 Aug 1;106(3):946–55.
227. O'Regan AW, Hayden JM, Berman JS. Osteopontin augments CD3-mediated interferon-gamma and CD40 ligand expression by T cells, which results in IL-12 production from peripheral blood mononuclear cells. *J Leukoc Biol*. 2000 Oct;68(4):495–502.
228. Hirano Y, Aziz M, Yang W-L, Wang Z, Zhou M, Ochani M, et al. Neutralization of osteopontin attenuates neutrophil migration in sepsis-induced acute lung injury. *Crit Care Lond Engl*. 2015 Feb 26;19:53.
229. Chung JW, Kim MS, Piao Z-H, Jeong M, Yoon SR, Shin N, et al. Osteopontin promotes the development of natural killer cells from hematopoietic stem cells. *Stem Cells Dayt Ohio*. 2008 Aug;26(8):2114–23.
230. Zhao H, Chen Q, Alam A, Cui J, Suen KC, Soo AP, et al. The role of osteopontin in the progression of solid organ tumour. *Cell Death Dis*. 2018 Mar 2;9(3):1–15.
231. Zhou Y, Dai DL, Martinka M, Su M, Zhang Y, Campos EI, et al. Osteopontin expression correlates with melanoma invasion. *J Invest Dermatol*. 2005 May;124(5):1044–52.
232. Raja R, Kale S, Thorat D, Soundararajan G, Lohite K, Mane A, et al. Hypoxia-driven osteopontin contributes to breast tumor growth through modulation of HIF1 $\alpha$ -mediated VEGF-dependent angiogenesis. *Oncogene*. 2014 Apr 17;33(16):2053–64.
233. Wei R, Wong JPC, Lyu P, Xi X, Tong O, Zhang S-D, et al. In vitro and clinical data analysis of Osteopontin as a prognostic indicator in colorectal cancer. *J Cell Mol Med*. 2018;22(9):4097–105.
234. Nieto MA, Sargent MG, Wilkinson DG, Cooke J. Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science*. 1994 May 6;264(5160):835–9.

## Gene expression analysis to study celiac disease and colitis-associated cancer

235. Stone RC, Pastar I, Ojeh N, Chen V, Liu S, Garzon KI, et al. Epithelial-mesenchymal transition in tissue repair and fibrosis. *Cell Tissue Res.* 2016 Sep;365(3):495–506.
236. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol.* 2019 Feb;20(2):69–84.
237. Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol.* 2000 Feb;2(2):76–83.
238. Sánchez-Tilló E, Lázaro A, Torrent R, Cuatrecasas M, Vaquero EC, Castells A, et al. ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene.* 2010 Jun 17;29(24):3490–500.
239. Miyoshi A, Kitajima Y, Sumi K, Sato K, Hagiwara A, Koga Y, et al. Snail and SIP1 increase cancer invasion by upregulating MMP family in hepatocellular carcinoma cells. *Br J Cancer.* 2004 Mar 22;90(6):1265–73.
240. Li NY, Weber CE, Mi Z, Wai PY, Cuevas BD, Kuo PC. Osteopontin up-regulates critical epithelial-mesenchymal transition transcription factors to induce an aggressive breast cancer phenotype. *J Am Coll Surg.* 2013 Jul;217(1):17–26; discussion 26.
241. Yu X, Zheng Y, Zhu X, Gao X, Wang C, Sheng Y, et al. Osteopontin promotes hepatocellular carcinoma progression via the PI3K/AKT/Twist signaling pathway. *Oncol Lett.* 2018 Oct;16(4):5299–308.
242. Song G, Cai Q-F, Mao Y-B, Ming Y-L, Bao S-D, Ouyang G-L. Osteopontin promotes ovarian cancer progression and cell survival and increases HIF-1alpha expression through the PI3-K/Akt pathway. *Cancer Sci.* 2008 Oct;99(10):1901–7.
243. Castello LM, Raineri D, Salmi L, Clemente N, Vaschetto R, Quaglia M, et al. Osteopontin at the Crossroads of Inflammation and Tumor Progression. *Mediators Inflamm.* 2017;2017:4049098.
244. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl.* 2013 Jan 1;29(1):15–21.
245. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
246. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
247. Zyla J, Marczyk M, Domaszewska T, Kaufmann SHE, Polanska J, Weiner J. Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinforma Oxf Engl.* 2019 Dec 15;35(24):5146–54.

## Gene expression analysis to study celiac disease and colitis-associated cancer

248. Kumar V, Gutierrez-Achury J, Kanduri K, Almeida R, Hrdlickova B, Zhernakova DV, 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 Jan 15;24(2):397–409.
249. Fearnley GW, Young KA, Edgar JR, Antrobus R, Hay IM, Liang W-C, et al. The homophilic receptor PTPRK selectively dephosphorylates multiple junctional regulators to promote cell-cell adhesion. *eLife.* 2019 Mar 29;8.
250. Li B, Zhuang L, Reinhard M, Trueb B. The lipoma preferred partner LPP interacts with alpha-actinin. *J Cell Sci.* 2003 Apr 1;116(Pt 7):1359–66.
251. Ngan E, Northey JJ, Brown CM, Ursini-Siegel J, Siegel PM. A complex containing LPP and  $\alpha$ -actinin mediates TGF $\beta$ -induced migration and invasion of ErbB2-expressing breast cancer cells. *J Cell Sci.* 2013 May 1;126(Pt 9):1981–91.
252. Janssens V, Zwaenepoel K, Rossé C, Petit MMR, Goris J, Parker PJ. PP2A binds to the LIM domains of lipoma-preferred partner through its PR130/B" subunit to regulate cell adhesion and migration. *J Cell Sci.* 2016 Apr 15;129(8):1605–18.
253. Li R, Peng C, Zhang X, Wu Y, Pan S, Xiao Y. Roles of Arf6 in cancer cell invasion, metastasis and proliferation. *Life Sci.* 2017 Aug 1;182:80–4.
254. Veldman-Jones MH, Brant R, Rooney C, Geh C, Emery H, Harbron CG, et al. Evaluating Robustness and Sensitivity of the NanoString Technologies nCounter Platform to Enable Multiplexed Gene Expression Analysis of Clinical Samples. *Cancer Res.* 2015 Jul 1;75(13):2587–93.
255. Zduniak K, Ziolkowski P, Ahlin C, Agrawal A, Agrawal S, Blomqvist C, et al. Nuclear osteopontin-c is a prognostic breast cancer marker. *Br J Cancer.* 2015 Feb 17;112(4):729–38.
256. Morath I, Hartmann TN, Orian-Rousseau V. CD44: More than a mere stem cell marker. *Int J Biochem Cell Biol.* 2016 Dec;81(Pt A):166–73.
257. Phillips RJ, Helbig KJ, Van der Hoek KH, Seth D, Beard MR. Osteopontin increases hepatocellular carcinoma cell growth in a CD44 dependant manner. *World J Gastroenterol.* 2012 Jul 14;18(26):3389–99.
258. Youssef NS, Osman WM. Relationship between osteopontin and  $\beta$ -catenin immunohistochemical expression and prognostic parameters of colorectal carcinoma. *Int J Clin Exp Pathol.* 2015;8(2):1503–14.
259. Robertson BW, Chellaiah MA. Osteopontin induces beta-catenin signaling through activation of Akt in prostate cancer cells. *Exp Cell Res.* 2010 Jan 1;316(1):1–11.
260. Meng F, Li J, Yang X, Yuan X, Tang X. Role of Smad3 signaling in the epithelial-mesenchymal transition of the lens epithelium following injury. *Int J Mol Med.* 2018 Aug;42(2):851–60.

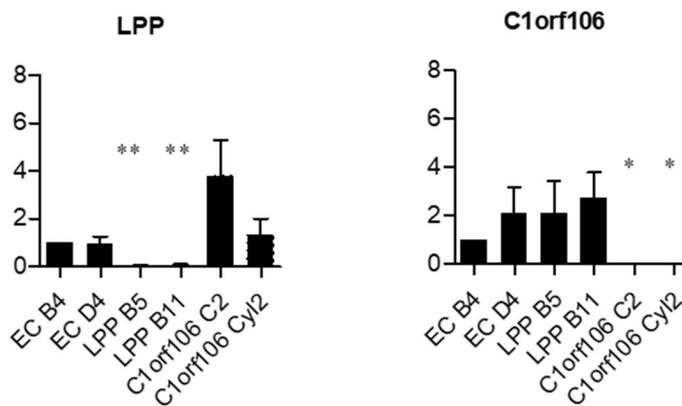
## Gene expression analysis to study celiac disease and colitis-associated cancer

261. Grand Moursel L, van der Graaf LM, Bulk M, van Roon-Mom WMC, van der Weerd L. Osteopontin and phospho-SMAD2/3 are associated with calcification of vessels in D-CAA, an hereditary cerebral amyloid angiopathy. *Brain Pathol Zurich Switz*. 2019 Nov;29(6):793–802.
262. Cheng Y, Wen G, Sun Y, Shen Y, Zeng Y, Du M, et al. Osteopontin Promotes Colorectal Cancer Cell Invasion and the Stem Cell-Like Properties through the PI3K-AKT-GSK/3 $\beta$ - $\beta$ /Catenin Pathway. *Med Sci Monit Int Med J Exp Clin Res*. 2019 Apr 24;25:3014–25.
263. Li Y, Xie Y, Cui D, Ma Y, Sui L, Zhu C, et al. Osteopontin Promotes Invasion, Migration and Epithelial-Mesenchymal Transition of Human Endometrial Carcinoma Cell HEC-1A Through AKT and ERK1/2 Signaling. *Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol*. 2015;37(4):1503–12.
264. Behera R, Kumar V, Lohite K, Karnik S, Kundu GC. Activation of JAK2/STAT3 signaling by osteopontin promotes tumor growth in human breast cancer cells. *Carcinogenesis*. 2010 Feb;31(2):192–200.
265. Xu C, Sun L, Jiang C, Zhou H, Gu L, Liu Y, et al. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed Pharmacother Biomedecine Pharmacother*. 2017 Jul;91:1167–77.
266. Gao C, Guo H, Wei J, Kuo PC. Osteopontin inhibits expression of cytochrome c oxidase in RAW 264.7 murine macrophages. *Biochem Biophys Res Commun*. 2003 Sep 12;309(1):120–5.
267. Dalal S, Zha Q, Daniels CR, Steagall RJ, Joyner WL, Gadeau A-P, et al. Osteopontin stimulates apoptosis in adult cardiac myocytes via the involvement of CD44 receptors, mitochondrial death pathway, and endoplasmic reticulum stress. *Am J Physiol Heart Circ Physiol*. 2014 Apr 15;306(8):H1182-1191.
268. Dalal S, Zha Q, Singh M, Singh K. Osteopontin-stimulated apoptosis in cardiac myocytes involves oxidative stress and mitochondrial death pathway: role of a pro-apoptotic protein BIK. *Mol Cell Biochem*. 2016 Jul;418(1–2):1–11.
269. Yousefi K, Irion CI, Takeuchi LM, Ding W, Lambert G, Eisenberg T, et al. Osteopontin Promotes Left Ventricular Diastolic Dysfunction Through a Mitochondrial Pathway. *J Am Coll Cardiol*. 2019 Jun 4;73(21):2705–18.
270. Parang B, Barrett CW, Williams CS. AOM/DSS Model of Colitis-Associated Cancer. *Methods Mol Biol Clifton NJ*. 2016;1422:297–307.
271. Rothemich A, Arthur JC. The Azoxymethane/Il10  $-/-$  Model of Colitis-Associated Cancer (CAC). *Methods Mol Biol Clifton NJ*. 2019;1960:215–25.
272. De Santis S, Verna G, Serino G, Armentano R, Cavalcanti E, Liso M, et al. Winnie-APCMin/+ Mice: A Spontaneous Model of Colitis-Associated Colorectal Cancer Combining Genetics and Inflammation. *Int J Mol Sci*. 2020 Apr 23;21(8).

## Gene expression analysis to study celiac disease and colitis-associated cancer

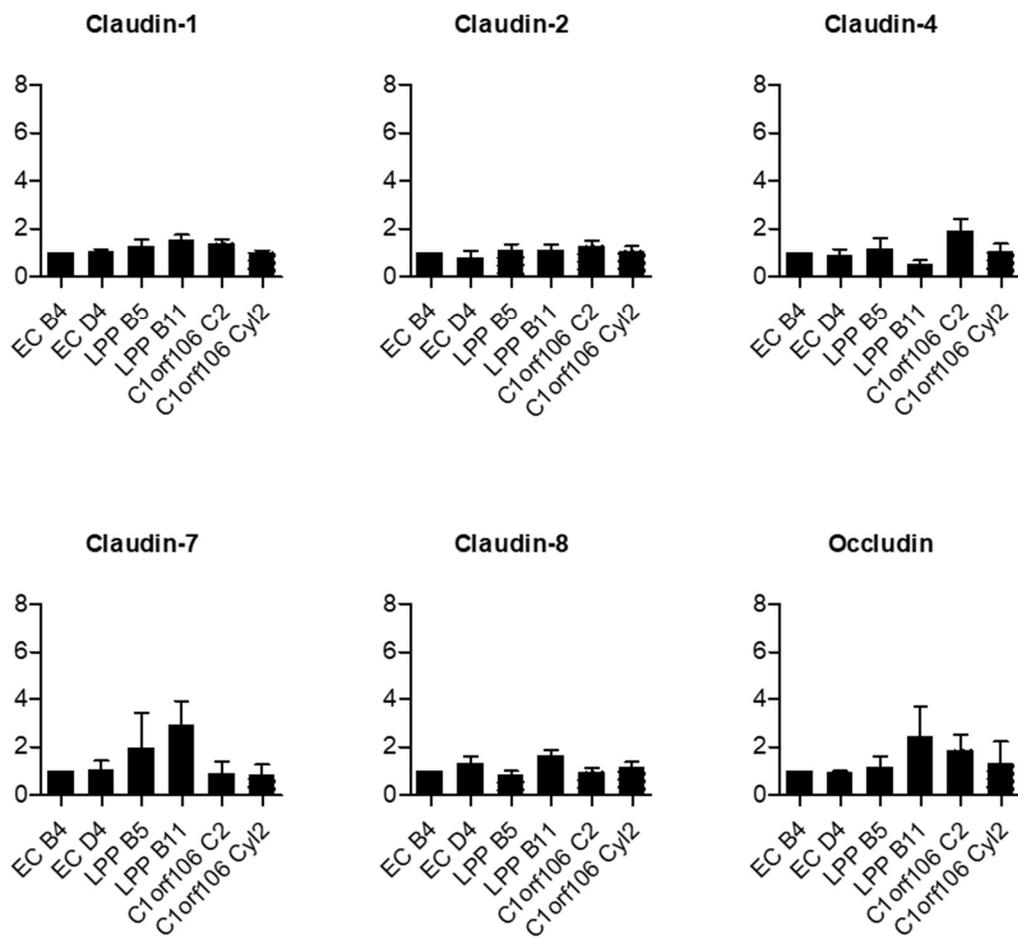
273. Heilmann K, Hoffmann U, Witte E, Loddenkemper C, Sina C, Schreiber S, et al. Osteopontin as two-sided mediator of intestinal inflammation. *J Cell Mol Med*. 2009 Jun;13(6):1162–74.
274. Almeqdadi M, Mana MD, Roper J, Yilmaz ÖH. Gut organoids: mini-tissues in culture to study intestinal physiology and disease. *Am J Physiol Cell Physiol*. 2019 Sep 1;317(3):C405–19.

## 9 SUPPLEMENTARY MATERIAL



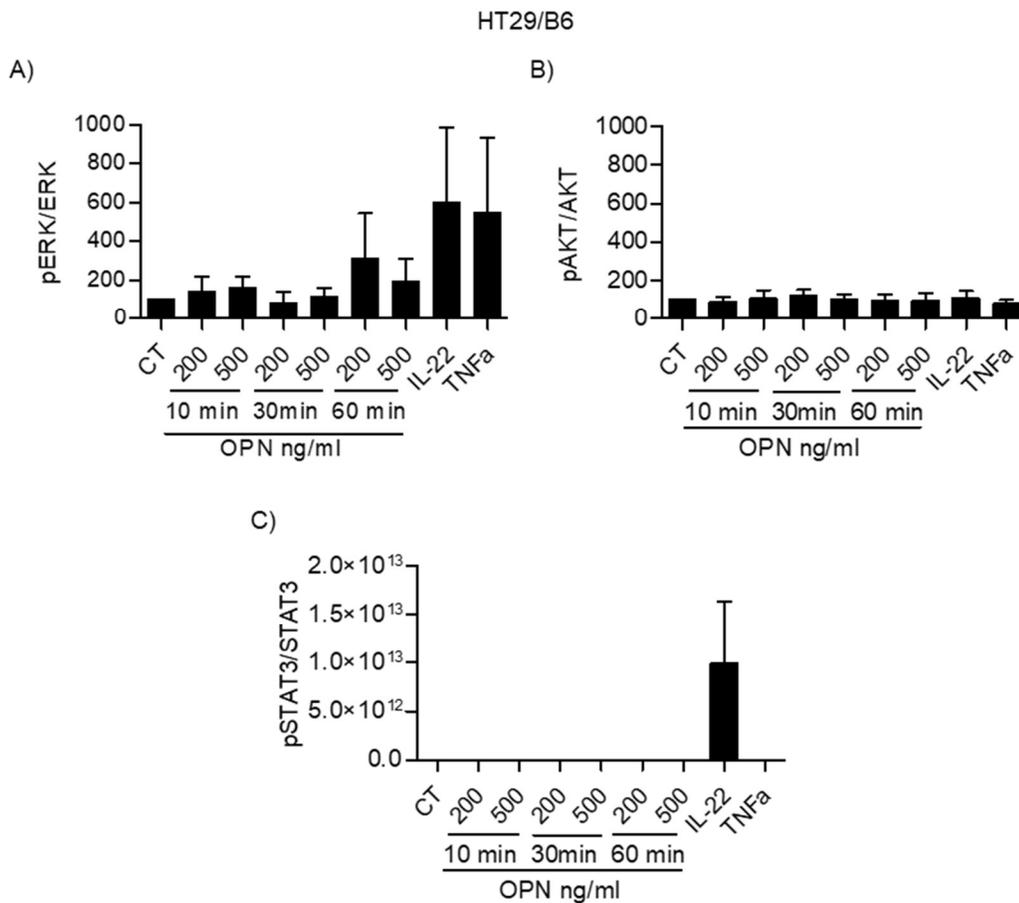
Supplementary figure 1. Densitometric analysis of Caco-2 knock-out clones. Cells from the Caco-2 control clones EC B4 and D4 and knock-out clones for LPP B5 and B11 and for C1orf106, C2 and Cyl2 had their protein content of LPP and C1orf106 analyzed using Western Blotting. Densitometric analysis was performed using actin as loading control and statistics were calculated using unpaired t test by comparing each knock-out clone to the average of the control clones EC B4 and EC D4.

## Supplementary material



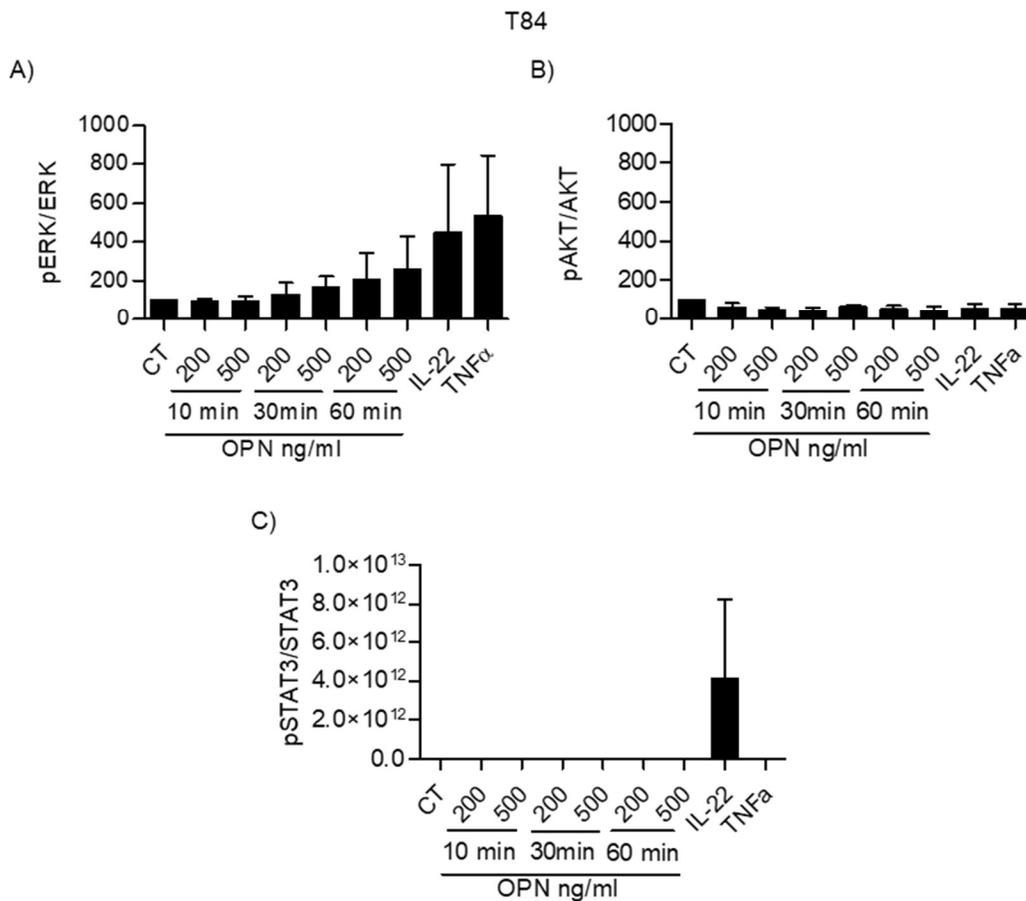
Supplementary figure 2. Densitometric analysis of tight junctional proteins in Caco-2 knock-out clones. Cells from the Caco-2 control clones EC B4 and D4 and knock-out clones for LPP B5 and B11 and for C1orf106, C2 and Cyl2 were analysed by Western Blotting for tight junctional proteins Claudin-1, -2, -4, -7, -8, and occludin. Densitometric analysis was performed using actin as loading control and statistics were calculated using unpaired t test by comparing each knock-out clone to the average of the control clones EC B4 and EC D4.

## Supplementary material



Supplementary figure 3. Densitometric analysis of HT29/B6 cells exposed to osteopontin. HT29/B6 cells were exposed to different concentrations of osteopontin for 10, 30 and 60 minutes and then examined by Western Blotting for phosphorylated and total ERK 1/2 (A) AKT (B) and STAT3 (C). Densitometric analysis was performed using actin as loading control and then a ratio between phosphorylated protein and total protein was calculated. Statistics were calculated using unpaired t test by comparing each experimental condition to the untreated control (CT).

## Supplementary material



Supplementary figure 4. Densitometric analysis of T84 cells exposed to osteopontin. T84 cells were exposed to different concentrations of osteopontin for 10, 30 and 60 minutes and then examined by Western Blotting for phosphorylated and total ERK 1/2 (A) AKT (B) and STAT3 (C). Densitometric analysis was performed using actin as loading control and then a ratio between phosphorylated protein and total protein was calculated. Statistics were calculated using unpaired t test by comparing each experimental condition to the untreated control (CT).

## **DECLARATION OF AUTHORSHIP**

I hereby certify that this thesis first submitted has been composed by me and is based on my own work, unless specified otherwise. No other person's work has been used without acknowledgement in this thesis. All references and literal extracts have been cited, and all sources of information, including graphs and data sets, have been specifically acknowledged.

March 2021, Berlin, Germany

---

Danielle Cardoso da Silva